Fusion of CNN and Transformer Architectures for Proactive Wildfire Detection in Satellite Imagery

Shereen Essam Elbohy, Mona M. Nasr, Farid Ali Mousa
Faculty of Computers and Artificial Intelligence, Helwan University, Egypt[1]
Associate Professor in Faculty of Computers and Artificial Intelligence, Helwan University, Egypt[2]
Associate Professor in Computer Science, Beni Suef University, Egypt[3]

***Abstract*—Wildfires pose a significant threat to ecosystems, human settlements, and air quality, necessitating advanced detection and mitigation strategies. Traditional wildfire detection methods often rely on manual observation and conventional machine learning approaches, which may lack efficiency and accuracy. This study proposes a novel deep learning model based on the ConvNeXt-Small architecture, a hybrid design that fuses the strengths of Convolutional Neural Networks (CNNs) and Transformer-inspired mechanisms, enabling more comprehensive analysis of wildfire patterns in satellite imagery. The model was trained using the Adam optimizer, which provides efficient convergence and adaptive learning. The dataset used consists of real-world satellite images collected from wildfire-affected regions in Canada, covering various geographic and seasonal conditions to reflect real environmental diversity. The results underscore the potential of ConvNeXt-based architecture for real-time, high-precision wildfire detection, offering a powerful tool for early intervention, disaster mitigation, and environmental monitoring efforts.**

Keywords—Wildfire detection; satellite imagery; convolutional neural networks (CNN); transformers; deep learning; hybrid model; proactive monitoring; remote sensing; disaster prevention; computer vision

I. INTRODUCTION

Rapid detection of sudden-onset disasters, such as earthquakes, flash floods, and emergencies like road accidents, is critical for effective response by emergency organizations. However, gathering information during these events is often labor-intensive and expensive, requiring manual data processing and expert analysis. To reduce these efforts, researchers have explored the use of computer vision techniques applied to satellite imagery, synthetic aperture radar, and other remote sensing data [1]. Despite these advances, such methods remain costly to deploy and lack the robustness needed to reliably capture relevant data in time-sensitive situations. Additionally, satellite imagery is often affected by noise from clouds and smoke common during hurricanes and wildfires and provides only an overhead perspective of disaster areas.

Forests play vital roles in our environment and daily lives by providing numerous resources. They are often referred to as the “lungs of the planet”, as they purify the air by producing oxygen (O2) and absorbing carbon dioxide (CO2). Forests also serve as habitats for diverse animal species and act as natural windbreakers that protect agricultural crops. Moreover, they help purify water by filtering out pollutants [2]. Economically, forests contribute significantly by providing jobs and boosting national incomes.

In recent years, many forests and wildlands have been destroyed by wildfires, which are uncontrolled natural disasters posing serious threats to economies worldwide. Each year, millions of acres of land are lost, resulting in significant damage to human lives, vegetation, and forest resources [3]. Wildfires also adversely affect agriculture by drying out the soil and destroying crops near affected areas. These fires can start from natural causes like lightning or human activities such as discarded cigarettes and land clearing. Climate change, rising temperatures, and prolonged droughts have intensified wildfire frequency and severity, especially in fire-prone regions such as California, Australia, and the Mediterranean. The environmental impact includes reduced air quality from pollutant emissions and accelerated global warming. Wildfires disrupt ecosystems by destroying habitats, threatening biodiversity, and increasing soil erosion. However, in some ecosystems, wildfires play a natural role by clearing dead vegetation and promoting new plant growth. Early detection and prevention are essential to minimize wildfire damage [4]. Advances in satellite imagery, artificial intelligence, and deep learning now enable real-time wildfire monitoring and prediction. Natural disasters are inescapable and have serious consequences for economies, ecological systems and human life. They can cause building collapses, spread diseases, and devastate entire nations through events like tsunamis, earthquakes, and forest fires. For instance, seismic forces from earthquakes can lead to the collapse of millions of buildings [5]. Since the 1990s, a variety of machine learning algorithms have been used to anticipate wildfires. A recent study in Italy used the random forest approach for wildfire susceptibility mapping [6]. Floods are the most catastrophic natural calamity, causing severe damage to property, infrastructure, and human life. To map flood susceptibility, an ensemble machine learning strategy using random forest (RF), random subspace (RS), and support vector machine (SVM) approaches was used. [7]. Studies indicate that immediately following a disaster, social media platforms contain valuable information for disaster response, including reports of damage and urgent needs from affected individuals [8], [9]. Much of this information comes in the form of images and videos. Unlike traditional data sources such as satellite imagery, leveraging on-the-ground social media images for relief efforts has remained largely unexplored until recently, as demonstrated in our prior work [10]. This gap is mainly due to technical challenges. Automatically filtering relevant images is difficult because social media streams are noisy, with a large proportion of posts unrelated to humanitarian needs. Additionally, effective filtering requires robust deep learning models trained on vast amounts of labeled data. However, assembling a large-scale, annotated dataset for incident recognition in real-world conditions is both time-consuming and expensive.

The structure of this study is as follows: Section II presents essential baseline knowledge of machine learning. Section III describes the dataset used in detail. Section IV introduces the proposed model. Section V discusses the evaluation results, and finally, Section VI concludes the study.

II. Background

In recent years, the integration of satellite imagery with machine learning (ML) techniques has become increasingly prevalent in various disaster management applications. These include monitoring changes in critical infrastructure, evaluating the impact of natural disasters, and tracking vegetation dynamics [10–11]. Among the ML methods, deep learning, particularly convolutional neural networks (CNNs) has shown exceptional capability in accurately detecting and classifying patterns in satellite data.

A widely adopted approach for forest fire detection using deep learning involves training CNNs on large datasets of satellite images. These datasets are commonly obtained from sources such as NASA's Earth Observing System, ESA's Sentinel program, or commercial providers like Planet Labs [12]. To prepare these datasets, images are labeled to indicate the presence or absence of wildfire activity. Labeling can be done manually by experts or semi-automatically using pre-trained models. Following preprocessing and annotation, the dataset is used to train deep learning models such as CNNs or, in some cases, recurrent neural networks (RNNs), depending on the spatial and temporal complexity of the data.

Once trained, the model is validated on a separate dataset to evaluate performance using metrics like accuracy, precision, recall, and F1-score. A successfully validated model can then be deployed to classify incoming satellite images in real time, enabling early wildfire detection and rapid response.

Deep learning significantly enhances both the speed and precision of wildfire detection, allowing quicker intervention and reducing potential damage [13]. Automated systems also reduce dependence on manual surveillance, which is often slower, more expensive, and less scalable. For instance, a study in [13] proposed a system combining the Inception-v3 CNN architecture with Local Binary Pattern (LBP) feature extraction, achieving high reliability in classifying wildfires from satellite imagery. Despite their accuracy, high-performance CNNs often come with high computational demands, making them unsuitable for mobile or embedded devices. This challenge has led to the exploration of lightweight CNN models like MobileNetV2 [28], which are optimized for devices with limited processing power. MobileNetV2 offers significant reductions in memory and computation requirements while maintaining competitive accuracy. For example, Wu et al. [1] integrated MobileNetV2 into their wildfire detection framework, applying data augmentation techniques such as Gaussian blur and additive Gaussian noise to simulate harsh visual conditions. This improved model robustness and allowed effective transfer learning, ensuring reliable performance even on low-resource platforms. There is a growing interest in designing neural networks that can operate efficiently in constrained environments, such as small satellites or unmanned aerial vehicles (UAVs). These platforms are essential in remote sensing scenarios, where local image classification can minimize the amount of data transmitted to ground stations, reducing both energy usage and communication delays. Rather than sending high-resolution imagery, systems can transmit only the classification outcome (e.g., "Fire" or "No Fire"), easing the load on communication networks like NASA's Deep Space Network (DSN) [29]. This research, therefore, focuses on optimizing neural network hyperparameters for wildfire detection, addressing challenges such as limited annotated data and resource-constrained deployment environments. The resulting CNN model is designed for real-time use on lightweight platforms like smartphones, small satellites, or weather balloons, facilitating accessible and cost-effective wildfire monitoring. Previous studies further support the effectiveness of CNNs in disaster-related applications. For instance, Boonsuk et al. [11] employed a two-layer CNN on the extended Cohn-Kanade dataset, comparing classifiers including Linear SVM, Linear Discriminant Analysis, and Softmax, and achieved over 90% accuracy with minimal variation. In challenging terrains such as mountainous regions where natural disasters like landslides or snowfall disrupt infrastructure deploying human responders can be difficult. In such cases, drones equipped with autopilot systems have become vital for capturing real-time imagery. Zhou et al. [13] introduced a CNN-based framework that employs inter-frame difference methods for noise removal and classifies disasters based on extracted image features. Various researchers have tackled different aspects of natural disaster detection and response through innovative technological solutions. For example, Sulistijono et al. [14] proposed a victim-detection system using aerial imagery transmitted to ground stations, enabling quick identification of victims following an earthquake. Padmawar et al. [15] developed a deep learning model combining Convolutional Neural Networks (CNNs) with Modified Particle Swarm Optimization (MPSO) to predict flood occurrences and locate individuals at risk ahead of time. In the domain of wildfire detection, Chen et al. [16] introduced a UAV-based detection system that utilized histogram stabilization and image smoothing before applying CNN-based classification. To enhance smoke recognition, they combined Local Binary Pattern (LBP) features with Support Vector Machines (SVMs), significantly improving early fire detection performance. Real-time monitoring has also proven vital for effective disaster management. Gonzalez et al. [17] presented the SFEwAN-SD model, a CNN architecture built on AlexNet with deconvolution layers, designed for real-time fire tracking using UAV imagery. Building on this, Samudre et al. [18] enhanced CNN efficiency through pipelining on FPGA platforms, which not only reduced energy consumption but also accelerated processing speeds. Resolution limitations in satellite imagery have posed challenges in disaster analysis. To address this, Lee et al. [19] adopted CNN models like VGG-13 and GoogleNet to work with high-resolution UAV images. Their comprehensive platform included forecasting tools, a web-based dashboard, and real-time alert systems, achieving high accuracy in wildfire early warning applications. Beyond aerial and satellite imagery, social media has become a valuable source for disaster assessment. Nguyen et al. [20] used event-specific CNN features to analyze social media images during major natural disasters, assessing structural damage. Similarly, Direkoglu et al. [21] applied optical flow techniques and CNNs to differentiate between normal and panic-driven human behavior during disasters, using the UMN and PETS2009 datasets. Smoke density classification remains a complex task in wildfire analysis. Yuan et al. [22] addressed this by introducing W-Net, a wave-shaped CNN architecture with encoder-decoder layers tailored for precise smoke density detection. Their method leveraged virtual datasets and layer feedback mechanisms to boost both sensitivity and accuracy. Flood detection has also seen progress through CNN-based systems. Layek et al. [23] developed a flood identification model enhanced with color filtering techniques. Their system consistently outperformed other methods when tested across multiple benchmark datasets.

Table I presents a comparative summary of the datasets, methodologies, and outcomes from various state-of-the-art studies. Building on these insights, the proposed ConvNeXt-Small model in this research is designed to achieve accurate, rapid, and resource-efficient wildfire detection, making it suitable for real-world deployment in resource-constrained environments. Consequently, to effectively extract features within a given timeframe, it is essential to integrate stationary methods like temporal windowing.

TABLE I. LITERATURE REVIEW

Authors	Methodology	Dataset Used	Strengths	Limitations	Results
Hartawan et al. [9]	Enhanced MLP with CNN on Raspberry Pi for victim detection using streaming cameras.	Streaming camera data	Real-time detection, suitable for low-power embedded systems.	Limited visibility, not suitable for satellite images.	Helped evacuation teams effectively identify victims.
Amit et al.[10]	CNN on resized satellite images for flood and landslide detection.	Resized satellite images	Detects multiple disaster types from a broader perspective.	It depends on the resolution and quality of satellite images.	Accurate flood and landslide detection via CNN.
Wu H et al.[1]	MobileNetV2 with data augmentation (Gaussian blur, noise).	Wildfire satellite imagery with augmentation	Efficient, lightweight, accurate; ideal for mobile/embedded use.	May not generalize to unseen disaster conditions.	High accuracy, fast classification on limited hardware.
Zhou et al.[13]	Interframe difference technique + CNN for aerial disaster classification.	Raw aerial images	Enhances image clarity; improves classification with motion information.	Requires multiple frames and clean aerial data.	Accurately identified disaster characteristics.
Sulistijono et al.[14]	CNN-based victim detection using aerial imagery and ground station.	Aerial images from simulation	Fast victim location via real-time detection framework.	Simulated environment; lacks real-world variability.	Successfully tested framework for locating victims.
Padmawar et al.[15]	CNN + Modified Particle Swarm Optimization (MPSO) for flood forecasting.	Flood satellite and imagery data	Hybrid ML improves robustness and early prediction.	High model complexity; expensive in real-time systems.	Predicted flood conditions accurately and timely.
Chen et al. [16]	LBP + SVM + CNN on UAV imagery after histogram equalization and filtering.	Forest fire UAV imagery	Improved raw image quality; enhanced detection accuracy.	Heavily reliant on preprocessing and proper lighting conditions.	Achieved accurate UAV-based fire detection.
Gonzalez et al. [17]	CNN + AlexNet + SFEwAN-SD for smoke density detection from UAV images.	UAV-based smoke density imagery	An effective real-time fire monitoring system with precise smoke localization.	Dependent on labeled training data and real-time transmission quality.	High performance in detecting smoke density in real-time.
Lee et al. [19]	Modified VGG-13 & GoogleNet using high-res UAV imagery for wildfire detection.	High-resolution UAV wildfire images	High spatial resolution enhances model precision.	High computing power required on UAV platforms.	Enabled accurate early-stage wildfire detection.
Layek et al.[23]	CNN with fusion of social media and satellite imagery using color filtering.	Social media images + satellite data	Diverse dataset boosts accuracy and adaptability across event types.	Needs extensive preprocessing to normalize diverse formats.	Outperformed other detection systems across several datasets.

III. DATASETS

The Forest Fire Mapping System in southern Quebec offers a detailed representation of both historical and recent wildfire occurrences, primarily within areas located south of the attributable forest limit. This mapping effort is vital for deepening our understanding of regional fire regimes and contributes significantly to the creation of targeted forest management strategies following wildfire incidents. Moreover, it addresses a broad spectrum of academic, research, and operational objectives, including evaluating the effects of climate change, modeling ecological recovery, and studying long-term forest ecosystem dynamics.

Fig. 1. Dataset location.

TABLE II. FOREST FIRE MAP

Forest Fire Map (Southern Quebec Focus)				
Data Sources (Satellite, Aerials, Surveys, Scar Data, Archives)				
Mapping Categories				
Detailed Mapping (1976-present)	**Simplified Outlines (1972-present)**	**Fire origins (1972-Present)**	**Historical fires (1800s-1975)**	**Fire Regime Zoning (1890-2020)**
- Total /partial Burn - Burn Patterns - 0.1 ha precision	- General perimeters - GIS/GPS Integration	- Ignition Data -Protection zones	- from old forest/eco maps -tree scars, documents	Based on fire data, terrain, species, ignition sources

The data used to support this mapping system is compiled from various sources, such as satellite imagery, aerial photos, ground and aerial reconnaissance, fire scar analysis, and historical archives. It is categorized into five main groups, as outlined in Table II. Fig. 1 shows the dataset location.

A. *Detailed Fire Mapping (1976–Present)*

This dataset offers high-resolution fire information, differentiating between complete and partial burns, and includes burn pattern classifications when available. It is capable of representing fire-affected areas as small as 0.1 hectares, depending on the data source. While coverage in northern southern Quebec is limited, the dataset remains an essential resource for detailed and precise fire analysis.

B. *Simplified Fire Perimeter Mapping (1972–Present)*

This data layer simplifies the outer boundaries of wildfire events by removing internal fragmentation, resulting in a streamlined representation of fire perimeters. The purpose of this generalization is to produce a more manageable and efficient dataset that is well-suited for integration into Geographic Information Systems (GIS), Global Positioning System (GPS) devices, and other digital mapping and analysis tools. Derived from comprehensive and detailed fire datasets, this layer is tailored to support a broad spectrum of users, including forestry professionals, emergency response teams, researchers, and policymakers. By offering a cleaner, more accessible view of fire extents, it enhances usability across various operational, planning, and analytical applications.

C. *Fire Origin Mapping (1972–Present)*

This dataset documents the ignition points of recorded wildfires, as tracked by protection agencies like SOPFEU. It includes key information such as the ignition date, the cause of the fire (either human activity or lightning), and the designated protection zone. The dataset provides comprehensive coverage across all regions of Quebec, offering valuable insights into fire prevention, analysis, and management efforts.

D. *Historical Fire Mapping (Late 19th Century–1975)*

This component reconstructs historical wildfire events using archival forest inventory maps and ecoforest maps. Fire occurrence dates are verified through fire-scar tree analysis and supporting archival records. Its regional scope includes Saguenay–Lac-Saint-Jean, Bas-Saint-Laurent, Gaspésie–Îles-de-la-Madeleine, Abitibi–Témiscamingue, Mauricie–Centre-du-Québec, and Lanaudière–Laurentides.

E. *Fire Regime Zoning (1890–2020)*

This map segments southern Quebec into 13 distinct zones, each representing a unique fire regime. The delineation of these zones is based on factors such as historical burned area data, physiographic characteristics, the fire dependency of dominant tree species, and sources of ignition—both natural and human-induced. The classification extends beyond managed forest areas to include portions of unmanaged northern territories. These zones play a critical role in wildfire risk assessment and are instrumental in forecasting future fire behavior influenced by climate change, fire suppression strategies, and evolving fuel distribution patterns.

The dataset utilized in this work is known as Wildfire. The dataset employed in this study is referred to as the Wildfire Prediction Dataset (Satellite Photos). It consists of high-resolution satellite images, each measuring 350×350 pixels, obtained via the MapBox API. These images are generated based on precise geographic coordinates, latitude, and longitude corresponding to verified wildfire incidents that have affected areas greater than 0.01 acres. This targeted spatial sampling ensures systematic coverage of significant wildfire events. The dataset is divided into two main classes: wildfire images, totaling 22,710, and non-wildfire images, providing a balanced representation suitable for training a robust classification model. A sample from the dataset is illustrated in Fig. 2. This balanced class distribution enhances the model's ability to distinguish between wildfire and non-wildfire scenarios accurately.

Fig. 2. Sample from dataset for wildfire and non-wildfire.

To construct the dataset, satellite imagery was systematically retrieved using the coordinates of documented wildfire locations. This conversion of traditional tabular incident data into visual, spatially referenced images makes the dataset highly compatible with deep learning architectures. These images serve as rich inputs for training models to predict wildfire risk based on spatial patterns observed in previous events. By learning from visual cues in the satellite images, the resulting predictive model can assess the likelihood of wildfire occurrence in specific regions. This capability is essential for supporting early intervention strategies, resource allocation, and wildfire prevention planning.

To promote generalization and mitigate overfitting, the dataset is split into three subsets: 70% for training, 15% for validation, and 15% for testing. The training set is used to optimize model parameters, the validation set assists in tuning hyperparameters and monitoring for overfitting, and the test set is utilized to evaluate the model's performance on previously unseen data.

IV. The Proposed Model

This section describes the methodology used for wildfire detection, which utilizes a hybrid deep learning approach combining ConvNeXt-Small as a feature extractor with a custom-designed classification head for binary classification. This model harnesses the efficiency of convolutional networks while integrating transformer-inspired enhancements, making it well-suited for processing high-resolution satellite images, as shown in Fig. 3.

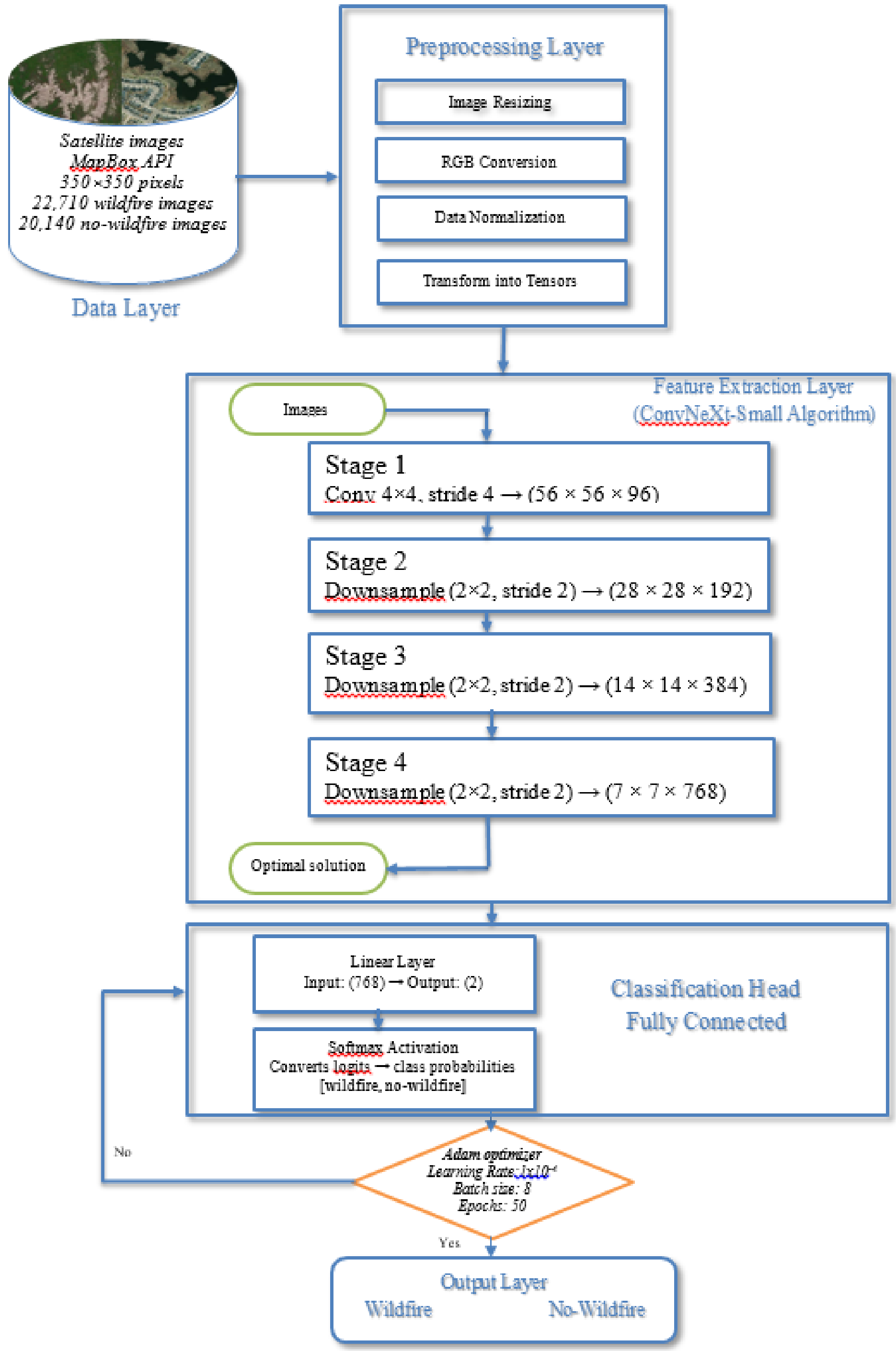

Fig. 3. Proposed model.

A. Backbone: ConvNeXt-Small as Feature Extractor

In this study, a deep learning model based on ConvNeXt-Small is developed, a cutting-edge convolutional neural network that incorporates concepts inspired by vision transformers while preserving the efficiency of convolutional operations. This model was used to classify satellite images into two categories: wildfire and non-wildfire. By combining the structured spatial biases of CNNs with the representational flexibility of vision transformers [24–25], this hybrid architecture significantly improves the analysis of high-resolution satellite imagery, making it especially effective for wildfire detection and monitoring.

ConvNeXt was selected over traditional CNNs such as ResNet [26] and pure transformer-based models like Vision Transformers (ViT) due to its balanced advantages. It achieves superior performance on vision tasks compared to ResNet, while offering efficient training. Unlike ViTs, which generally require very large datasets to perform well [27], ConvNeXt demands less data, making it better suited for moderately sized datasets. Its modular and adaptable architecture also facilitates seamless integration with custom classifier heads tailored for specific applications. Furthermore, ConvNeXt's scalability and speed make it an excellent choice for real-time wildfire alert and detection systems. ConvNeXt-Small forms the backbone of our architecture, enhancing traditional CNNs like ResNet by incorporating elements inspired by vision transformers, thereby boosting its representational capabilities. To reduce computational complexity while effectively capturing both spatial and channel-wise features, depthwise separable convolutions are utilized. Instead of Batch Normalization, Layer Normalization is applied, providing greater stability during training, especially with large datasets. The GELU activation function is employed to improve gradient flow and introduce better non-linearity, resulting in richer feature representations. Furthermore, the use of larger kernel sizes enables improved context aggregation, emulating the global receptive field characteristic of transformers [28]. Pretrained ConvNeXt-Small model are used in PyTorch's torchvision.models, with its original classification layer removed. This allows the model to leverage pretrained visual features learned from ImageNet such as edges, textures, and terrain patterns that are transferable to satellite imagery. The final output feature map from ConvNeXt-Small is a tensor of shape [batch_size, 768], which is then fed into the subsequent classification head.

B. Classification Head: Fully Connected Layer

After feature extraction, a lightweight fully connected neural network serves as the classification head, transforming the high-dimensional feature vectors into wildfire categories. This classification head includes a linear layer that compresses the 768-dimensional feature vector down to two output neurons, representing the two classes: wildfire and no-wildfire. A Softmax activation function is then applied to the output logits to convert them into a probability distribution, providing clear and interpretable prediction scores as shown in Fig. 4.

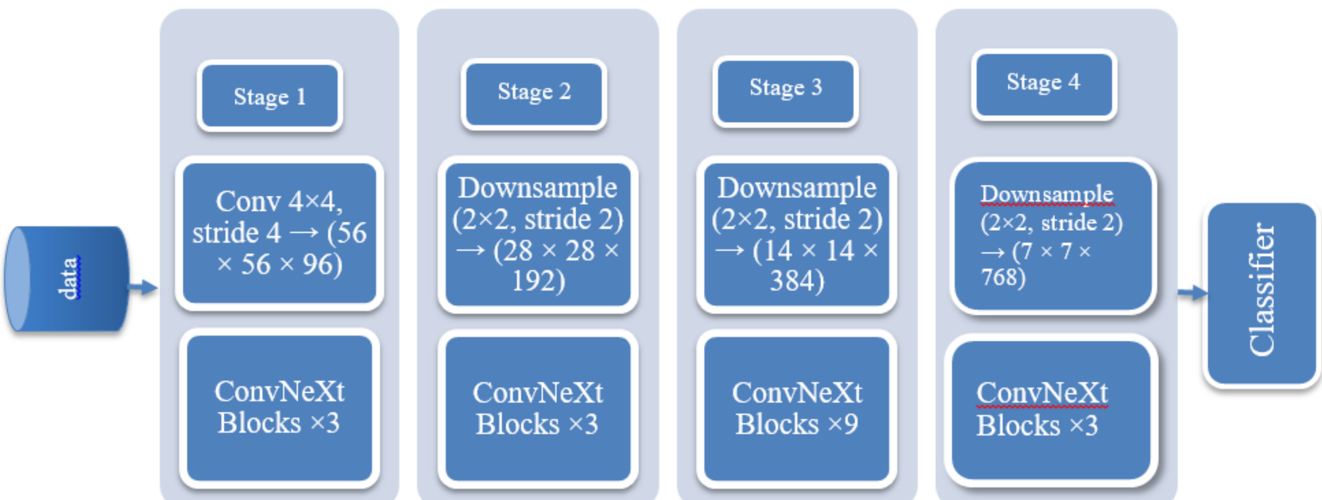

Fig. 4. ConvNeXt-small architecture.

Mathematically, the classification process can be represented as follows:

Eq. (1) is Softmax activation function:

$$\hat{Y} = \text{Softmax}\ (W.\ F + b) \tag{1}$$

where, "F" represents the feature vector output from ConvNeXt, while "W" denote the trainable weight matrix and "b" bias term, respectively. This equation ensures that each input image is assigned a probability distribution over the two classes, where the higher probability value indicates the most likely classification. Algorithm 1 shows the ConvNeXt-Small Algorithm for wildfire detection.

Algorithm 1: ConvNeXt-Small Algorithm for Wildfire Detection

Input: Preprocessed image (224x224)

 # 1. Feature Extraction
x = Patchify(image)
 # Conv layer: 4x4 kernel, stride=4 → 96 channels
x = ConvNeXtBlocks(x, depths= [3,3,9,3])
 # Stacked blocks across 4 stages
x = GlobalAveragePooling(x) # $F \in \mathbb{R}^{768}$

```
# 2. Classifier Head
logits = Linear (F, 768 → 2)
ŷ = Softmax(logits)
 # Probabilities for [wildfire, no-wildfire]
# 3. Loss Function (Cross-Entropy)
L_CE = -Σ (yᵢ * log(ŷᵢ)) for i in {1, 2}
# 4. Training Loop (Adam Optimizer)
for epoch in range(50):
  for batch in training_data:
    ŷ = ConvNeXt(batch_images)
    loss = CrossEntropy(ŷ, batch_labels)
    Backprop + Adam update
  if early_stopping_triggered:
    break
#Output: Predicted class ŷ, Loss L_CE
```

1) The classifier is specifically designed to minimize classification errors while achieving high accuracy in differentiating between wildfire and non-wildfire cases. During training, only the classification head is fine-tuned, while the ConvNeXt-Small backbone remains frozen. This approach allows the model to effectively leverage the knowledge acquired from large-scale ImageNet pretraining and apply it to the wildfire classification task, enhancing its generalization ability.

2) The classification head is optimized using the cross-entropy loss function, which quantifies the difference between the predicted and true labels. The Softmax activation function ensures that the predicted class probabilities sum to one, enabling them to be interpreted as confidence scores. Throughout training, backpropagation updates the weights and biases of the classification head to minimize the classification loss.

C. Loss Function: Cross-Entropy

To optimize model predictions, we use the cross-entropy loss function, which measures the divergence between predicted probabilities and true class labels. The cross-entropy loss is given by:

Eq. (2) Loss Function: Cross-Entropy

$$L_{CE} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \quad (2)$$

where, C is the number of classes (wildfire and no-wildfire), y_i is the true class label, and $\hat{y}_i$ is the predicted probability for the class. This function ensures that the model maximizes the probability of the correct class while reducing the probability of incorrect ones. The cross-entropy loss penalizes incorrect predictions more severely, ensuring that the model learns to focus on correctly classifying challenging samples. This encourages the network to assign high confidence scores to the correct class while suppressing incorrect predictions, leading to a more robust classifier. As training progresses, minimizing this loss function results in a model that effectively distinguishes wildfire from non-wildfire images with improved confidence and reliability.

D. Optimizer and Training Strategy

The Adam optimizer was utilized for training, an adaptive learning rate algorithm that combines the benefits of momentum and RMSprop methods [29]. Adam adjusts learning rates for each parameter dynamically by estimating the first moment (mean) and second moment (variance) of gradients, which promotes efficient and stable convergence. This makes it particularly effective for deep learning tasks involving high-dimensional satellite imagery, as it accelerates training while managing sparse gradients. Key training settings included a learning rate of 1×10^{-4}, carefully chosen to balance between exploration and training stability, and weight decay to regularize the model and reduce overfitting. Training was conducted with a batch size of 8 over 50 epochs. Early stopping was implemented to monitor validation loss and halt training when improvements plateaued, thereby avoiding unnecessary computation and further preventing overfitting [30].

V. Results

The wildfire detection system built on the ConvNeXt-Small architecture follows a structured pipeline, starting with the preprocessing of satellite images from historically affected areas. Each image is resized to 350×350 pixels, normalized, and converted into tensors for PyTorch compatibility. The processed images are then passed through the ConvNeXt-Small model, which extracts high-dimensional feature vectors that capture essential spatial and contextual information. These features are fed into a lightweight classifier head, a linear layer followed by a Softmax activation that outputs probabilities for two classes: wildfire and no-wildfire. During training, the model's predictions are compared with actual labels using the cross-entropy loss function, optimizing performance for accurate binary classification. Once the model is trained, it undergoes a rigorous evaluation phase to assess its predictive performance across multiple metrics. These include test accuracy, precision, recall, and F1-score, all of which help measure the model's classification effectiveness from different perspectives. In this case, the model achieved outstanding results: an overall test accuracy of 99.05%, precision scores of 98% for the no-wildfire class and 100% for the wildfire class, and recall and F1-scores of 99% across both classes as shown in Table III and Fig. 5 to Fig. 7. Furthermore, a confusion matrix is computed to visualize true positive, false positive, true negative, and false negative rates, providing insight into classification errors as shown in Table IV, the model demonstrates excellent discriminative ability, making it highly suitable for reliable wildfire detection in diverse and real-world environmental conditions.

TABLE III. Proposed Model Results

Class	Precision	Recall	F1-Score	Support
No Wildfire	0.98	0.99	0.99	2820
Wildfire	1.00	0.99	0.99	3480
Accuracy			0.99	6300
Macro Avg	0.99	0.99	0.99	6300
Weighted Avg	0.99	0.99	0.99	6300

TABLE IV. Confusion Matrix

	Classification	
	Positive	Negative
Predicted Positive	2805	15
Predicted Negative	45	3435

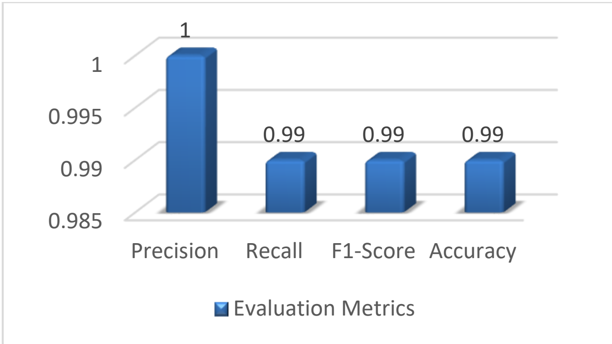

Fig. 5. Classification performance comparison for Wildfire.

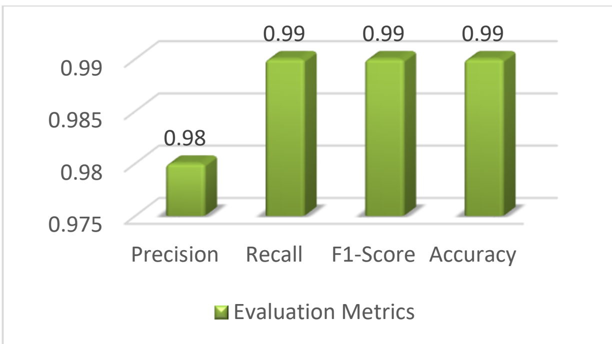

Fig. 6. Classification performance for No Wildfire.

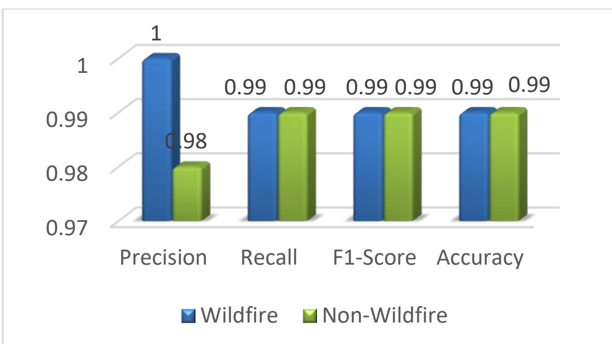

Fig. 7. Classification performance comparison between Wildfire and No Wildfire results.

VI. Conclusion

In this study, this holistic pipeline spanning data preparation, feature extraction, classification, optimization, and evaluation, not only ensures high accuracy but also provides scalability and robustness needed for deployment in operational wildfire monitoring systems. The proposed wildfire detection model, built upon the ConvNeXt-Small architecture, has demonstrated high effectiveness in identifying wildfire-prone areas using satellite imagery. By combining state-of-the-art convolutional design with a lightweight, fully connected classifier head, the model achieved remarkable predictive performance with a test accuracy of 99.05%, precision up to 100% for wildfire cases, and consistently high recall and F1-scores across both classes. These results confirm the model's robustness and reliability, making it a viable solution for early wildfire risk assessment.

The model's strong generalization ability is attributed to an efficient preprocessing pipeline, thoughtful training strategy using cross-entropy loss and the Adam optimizer, and the integration of modern deep learning techniques tailored for moderate-sized datasets. Additionally, the high AUC score of 0.996 indicates excellent discriminative power between wildfire and non-wildfire instances. This system offers a scalable and practical approach for real-time wildfire detection, with potential applications in environmental monitoring, disaster prevention, and resource planning. Future work may focus on expanding the dataset across different seasons and geographical regions, integrating meteorological data, and exploring deployment strategies for on-ground alert systems.

References

[1] Liao, W. H., Chang, Y. S., & Wu, Y. C. (2023, January). Detection of synthesized satellite images using deep neural networks. In 2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM) (pp. 1-5). IEEE.

[2] Bouguettaya, A., Zarzour, H., Taberkit, A. M., & Kechida, A. (2022). A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms. Signal Processing, 190, 108309.

[3] Meimetis, D., Papaioannou, S., Katsoni, P., & Lappas, V. (2024). An Architecture for Early Wildfire Detection and Spread Estimation Using Unmanned Aerial Vehicles, Base Stations, and Space Assets. Drones and Autonomous Vehicles, 1(3), 10006.

[4] Abramov, N., Emelyanova, Y., Fralenko, V., Khachumov, V., Khachumov, M., Shustova, M., & Talalaev, A. (2024). Intelligent methods for forest fire detection using unmanned aerial vehicles. Fire, 7(3), 89.

[5] Mignan, A.; Broccardo, M. Neural network applications in earthquake prediction (1994–2019): Meta-analytic and statistical insights on their limitations. Seism. Res. Lett. 2020, 91, 2330–2342.

[6] Tonini, M.; D'Andrea, M.; Biondi, G.; Degli Esposti, S.; Trucchia, A.; Fiorucci, P. A Machine Learning-Based Approach for Wildfire Susceptibility Mapping. The Case Study of the Liguria Region in Italy. Geosciences 2020, 10, 105.

[7] Islam, A.R.M.T.; Talukdar, S.; Mahato, S.; Kundu, S.; Eibek, K.U.; Pham, Q.B.; Kuriqi, A.; Linh, N.T.T. Flood susceptibility modelling using advanced ensemble machine learning models. Geosci. Front. 2021, 12, 101075.

[8] C. Castillo, Big Crisis Data. Cambridge University Press, 2016.

[9] M. Imran, C. Castillo, F. Diaz, and S. Vieweg, "Processing social media messages in mass emergency: A survey," ACM Computing Surveys, vol. 47, no. 4, p. 67, 2015.

[10] E. Weber, N. Marzo, D. P. Papadopoulos, A. Biswas, A. Lapedriza, F. Ofli, M. Imran, and A. Torralba, "Detecting natural disasters, damage, and incidents in the wild," in European Conference on Computer Vision. Springer, 2020, pp. 331–350.

[11] Boonsuk, R.; Sudprasert, C.; Supratid, S. An Investigation on Facial Emotional Expression Recognition Based on Linear-DecisionBoundaries Classifiers Using Convolutional Neural Network for Feature Extraction. In Proceedings of the 2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE), Pattaya, Thailand, 10–11 October 2019; pp. 1–5.

[12] Browse datasets | NASA Earth Observations (NEO).

[13] Zhou, F.; Huang, J.; Sun, B.; Wen, G.; Tian, Y. Intelligent Identification Method for Natural Disasters along Transmission Lines Based on Inter-Frame Difference and Regional Convolution Neural Network. In Proceedings of the 2019 IEEE International Conference on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), Xiamen, China, 16–18 December 2019; pp. 218–222.

[14] Sulistijono, I.A.; Imansyah, T.; Muhajir, M.; Sutoyo, E.; Anwar, M.K.; Satriyanto, E.; Basuki, A.; Risnumawan, A. Implementation of Victims Detection Framework on Post Disaster Scenario. In Proceedings of the 2018 International Electronics Symposium on Engineering Technology and Applications (IES-ETA), Bali, Indonesia, 29–30 October 2018; pp. 253–259.

[15] Padmawar, P.M.; Shinde, A.S.; Sayyed, T.Z.; Shinde, S.K.; Moholkar, K. Disaster Prediction System using Convolution Neural Network. In Proceedings of the 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 17–19 July 2019; pp. 808–812.

[16] Chen, Y.; Zhang, Y.; Xin, J.; Wang, G.; Mu, L.; Yi, Y.; Liu, H.; Liu, D. UAV Image-based Forest Fire Detection Approach Using Convolutional Neural Network. In Proceedings of the 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China, 18–21 June 2019; pp. 21.

[17] Gonzalez, A.; Zuniga, M.D.; Nikulin, C.; Carvajal, G.; Cardenas, D.G.; Pedraza, M.A.; Fernández, C.; Munoz, R.; Castro, N.;Rosales, B.; et al. Accurate fire detection through fully convolutional network. In Proceedings of the 7th Latin American Conference on Networked and Electronic Media (LACNEM 2017), Valparaiso, Chile, 6–7 November 2017.

[18] Samudre, P.; Shende, P.; Jaiswal, V. Optimizing Performance of Convolutional Neural Network Using Computing Technique. In Proceedings of the 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Pune, India, 29–31 March 2019;pp. 1–4.

[19] Lee, W.; Kim, S.; Lee, Y.-T.; Lee, H.-W.; Choi, M. Deep neural networks for wild fire detection with unmanned aerial vehicle. In Proceedings of the 2017 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 8–11 January 2017;pp. 252–253.

[20] Nguyen, D.T.; Ofli, F.; Imran, M.; Mitra, P. Damage assessment from social media imagery data during disasters. In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Sydney, NSW, Australia, 31 July–3 August 2017; pp. 569–576.

[21] Direkoglu, C. Abnormal Crowd Behavior Detection Using Motion Information Images and Convolutional Neural Networks.IEEE Access 2020, 8, 80408–80416.

[22] Yuan, F.; Zhang, L.; Xia, X.; Huang, Q.; Li, X. A wave-shaped deep neural network for smoke density estimation. IEEE Trans.Image Process. 2019, 29, 2301–2313.

[23] Layek, A.K.; Poddar, S.; Mandal, S. Detection of Flood Images Posted on Online social media for Disaster Response. In Proceedings of the 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), Gangtok,Sikkim, India, 28–28 February 2019; pp. 1–6.18–2123.

[24] Ramos, L., Casas, E., Romero, C., Rivas-Echeverría, F., & Morocho-Cayamcela, M. E. (2024). A study of convnext architectures for enhanced image captioning. IEEE Access, 12, 13711-13728.

[25] Feng, J., Tan, H., Li, W., & Xie, M. (2022, November). Conv2NeXt: Reconsidering Conv NeXt Network Design for Image Recognition. In 2022 International Conference on Computers and Artificial Intelligence Technologies (CAIT) (pp. 53-60). IEEE

[26] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O. & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, 8, 1-74.

[27] Xia, C., Wang, X., Lv, F., Hao, X., & Shi, Y. (2024). Vit-comer: Vision transformer with convolutional multi-scale feature interaction for dense predictions. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5493-5502).

[28] Lyu, K., Li, Z., & Arora, S. (2022). Understanding the generalization benefit of normalization layers: Sharpness reduction. Advances in Neural Information Processing Systems, 35, 34689-34708.

[29] Kartowisastro, I. H., & Latupapua, J. (2023). A Comparison of Adaptive Moment Estimation (Adam) and RMSProp Optimisation Techniques for Wildlife Animal Classification Using Convolutional Neural Networks. Revue d'Intelligence Artificielle, 37(4).

[30] Zhang, Z. (2018, June). Improved adam optimizer for deep neural networks. In 2018 IEEE/ACM 26th international symposium on quality of service (IWQoS) (pp. 1-2). Ieee.