WHEAT FIELD FIRE SMOKE DETECTION FROM UAV IMAGES USING CNN-CBAM

by

Ahmed Essa Alsalem

A dissertation submitted in partial fulfillment of the requirements for the degree of

'DOCTOR OF PHILOSOPHY IN SYSTEMS ENGINEERING'

FAll 2026

Oakland University
Rochester, Michigan

Doctoral Advisory Committee:
Mohamed A. Zohdy, Ph.D., Chair
William Edwards, Ph.D.
Christopher Kobus, Ph.D.
Erica Ruegg, Ph.D.

For Every Great Dream, There Is A Long Journey

To My Mother soul.

To M grandmother soul

To My daughter Fatima

ACKNOWLEDGMENTS

In the name of God, the Most Gracious, the Most Merciful, I begin with gratitude to Allah for His guidance, patience, and blessings that made this work possible.

I am deeply thankful to my advisor, Professor Mohamed Zohdy, for his invaluable guidance, insightful feedback, and unwavering support, which have shaped me both academically and professionally. I also sincerely thank my committee members—Dr. William Edwards, Dr. Christopher Kobus, and Dr. Erica Ruegg—for their expert advice and encouragement, which greatly refined this research.

To my family, my greatest source of strength: my beloved wife, Tahani, for her constant support and patience; my daughter, Fatima, for her love and inspiration; and my siblings, for their unwavering encouragement. Your presence has been my anchor.

Lastly, I extend my gratitude to my friends, colleagues, and everyone who supported me on this journey. Your encouragement and contributions have been invaluable. Thank you all for making this achievement possible.

Ahmed Alsalem

ABSTRACT

WHEAT FIELD FIRE SMOKE DETECTION FROM UAV IMAGES USING CNN-CBAM

By

Ahmed Alsalem

Adviser: Professor Mohamed A. Zohdy

Wheatfield fires all seriously threaten farmers' income, agricultural resources, and wildlife habitats. These fires have become much more frequent nowadays, due to climate change and activities of humans in farming areas. Since smoke detection is the initial step in early fire detection, action must be taken right away. We proposed an improved CNN model that can detect smoke in wheat field fires to address this problem. Features are extracted from UAV images using CNN. CNN is a deep learning model that automatically adjusts to spatial information in input images. As part of this method, 3900 UAV images were collected to illustrate smoke from wheat field fires. The suggested method divides Wheatfield fire images into four categories: high, medium, low, and no fire. The dataset displays smoke from burning wheat. Included are the following categories of smoke photos:(a) a lot of smoke and fire; (b) a lot of smoke and fire (c) Mild smoke and fire. There are three types of fire and smoke: (d) medium; (e) small; and (f) small.

The CNN-CBAM and the model that combines CBAM and CNN were trained and validated on this dataset. One strategy for accomplishing this is to apply the channel along with the pores to the important aspects of the images, which is known as the

attention mask method. The proposed task is to extract features for smoke patch characteristics of wheat field fires. Also, the CNN design incorporates disconnected heads that extract relevant data from a variety of sources.

Image categorization and feature extraction using convolutional neural networks comprise the two components of the suggested CNN-CABM approach. The usefulness of this suggested method was validated through testing on our wheat field fire smoke dataset, outperforming the prior model and achieving an accuracy of up to 0.92. The study found that the combination of CBAM and CNN is particularly effective for the task of detecting fires in wheat fields, indicating a new bright side to the area of early warning systems.

To cross-validate the findings of the same dataset using basic SVM and CNN, we determined that the suggested model CNN-CABM outperforms simple SVM and CNN to produce better results. The next steps will be to improve this version's stability in the face of shifting multiplicative nuisances, as well as to increase speed and precision, according to those who contributed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI:	Artificial intelligence
CBAM:	Convolutional Block Attention Module
CNN:	Convolutional Neural Networks
DCNNs:	Deep Convolutional Neural Networks
FastRNN:	Fast Recurrent Neural Network
FCNN:	Fully Convolutional Neural Network
GPS:	Global Positioning Systems
IoT:	Internet of Things
NIFC:	National Interagency Fire Centre
NIFC:	National Interagency Fire Centre
NIFC:	National Interagency Fire Centre
ROI:	Region of Interest
SVM:	Support Vector Machine
UAVs:	Unmanned Aerial Vehicles
YOLO:	You Only Look Once

CHAPTER ONE

INTRODUCTION

1.1 Introduction and Background

Wheat is the most extensively grown major food crop worldwide. It contributes significantly to global food security because its production and cultivated land account for about one-third of the crop. [1]. A significant biological calamity that lowers wheat productivity and quality is weed growth in wheat fields. Wheat and weeds compete for growth spaces, sunlight, water, and fertilizer. Weeds reduce wheat output by altering variables such as leaf area and biomass buildup [2].

According to studies, weeds diminish wheat yields by an average of 24%. Thus, it is vital to implement weed management measures to prevent yield loss [3, 4]. The authors [5] physical weed control techniques (such as mulching, polarizations, blazing, and steaming), biological weed control techniques, and mechanical weed control techniques are divided into three categories. Chemical herbicides are widely used by tractors or Unmanned Aerial Vehicles (UAVs) because their ease-of-use spraying has emerged as a prominent weed control tool. Chemical herbicides have caused significant environmental damage, despite their intended purpose of targeting weeds [6].

Wheatfield fires shown in Figure 1, often known as stubble fires, are a major concern for agricultural farmers worldwide. These fires can develop for a number of factors, including lightning and other natural sources, unintentional causes from machinery sparks, and deliberate actions. Despite the source, the consequences of these fires are significant, hurting crop yields, soil quality, air quality, and overall farm profitability.

Wheat fires are natural disasters that do permanent harm to the local ecology. Sudden and unmanageable flames can pose a serious threat to residents' lives. In the United States, data from the National Interagency Fire Centre (NIFC) show that between 1990 and 2015, the burned area more than doubled. Recent wildfires in northern California (as reported by CNN) have already claimed the lives of more than 40 people and left 50 others missing. There was an emergency that required the evacuation of over 200,000 residents. Across the world, wildfires occur 220,000 times a year, burning over 6 million hectares of land. It follows that early and accurate wildfire identification is crucial [7].

Each year, biomass burning emits more than 100 million tons of aerosolized smoke were released into the sky. Tropical areas experience more than 80% of all fires (Christopher et al. Citation 2002). Smoke plumes from such incidents can reach the stratosphere and travel hundreds or even thousands of kilometers horizontally under specific atmospheric circulation conditions. Smoke, therefore, can have an effect well beyond the boundaries of a fire.

Traditional wildfire detection relies primarily on human observation from watchtowers. It is inefficient within the constraints of space and time. Unmanned aerial vehicles UAVs or drones have become gradually popular as investigation instruments [8].

UAVs equipped with ultrahigh-resolution digital cameras and global positioning systems (GPS) may be able to provide HD aerial photos along with precise location data. Previous research has demonstrated the speed, economy, and efficiency of a well-trained UAV swarm in completing intricate agricultural tasks [9]. Recent studies also reveal that UAVs are swiftly developing to implement in a variety of surveillance jobs [10, 11].

Agricultural fires, especially in wheat fields, pose a substantial risk to agrarian output, air quality, and ecosystem health. Rapid detection of such fires is critical for limiting damage and coordinating effective actions. Traditionally, fire detection technologies have depended on ground-based observations and satellite imaging, which are often delayed or have low resolution. Introducing Unmanned Aerial Vehicles (UAVs) presents a promising alternative due to their ability to cover huge regions quickly and produce high-quality imagery.

Early fire detection is critical for limiting damage and halting the spread of the fire, especially through prompt smoke identification. Because there is a lot of combustible material present, they spread quickly and can get challenging to put out. In addition to causing harm to individuals and their belongings, the fires hasten the devastation of the neutral environment. In this earlier work, a quick and precise video-based smoke detection algorithm was proposed as a solution to these issues.

Early detection of wheat field fires can prevent disasters. Image-based systems, as opposed to sensor-based ones, are especially well adapted to outdoor settings like parking lots, wheatfields, and mountains. It has previously been suggested to use a convolutional neural network (CNN) for video-based smoke and fire detection, which aids in solving this issue more effectively. Smoke and fire detection are the two main categories into which the previously employed techniques are typically separated [5]. Since smoke spreads more quickly than flames and covers a larger area, it is particularly crucial for the early detection of wheat field fires.

Figure 1.1: Wheat fire

Conventional image-based methods for identifying smoke and fire primarily rely on manually created, low-level characteristics like color, texture, and shape that are derived from empirical characteristics. The accuracy of these particular methods decreases in dynamic environments. Low plant moisture content, dryness, and increased agricultural activities are often associated with summer wheat field fires. Significant and moderate wheat-growing regions are most vulnerable to these fires, which are especially common in the early summer.

Three things are needed for a wheat field fire to start and spread: oxygen to feed the flames, combustible material, and a heat source. Uncontrolled wheat field fires have the ability to spread quickly and far, even bridging natural boundaries like rivers and roads. These fires cause millions of acres of land to be damaged annually. To enhance the detection system, machine learning techniques for image-based detection and real-time classification were developed.

Convolutional Neural Network's up-to-date updates have substantially improved automated image analysis, allowing for the successful detection of smoke patterns from UAV-captured images. The CNN with Convolutional Block Attention Module (CBAM) incorporates attention methods that focus the model on key features, potentially enhancing the result and timeliness of smoke detection in wheat fields. By leveraging the strengths of UAV technology and deep learning, this study addresses the challenges of detecting smoke at various scales and under different atmospheric conditions.

Wheat Field Fire

Wheatfield fires shown in Figure 1.2 threaten wildlife habitats, agricultural resources, and farmer revenues. Over the last few years, we have gotten to know the rise of agricultural fires, which we think are the results of human activities and climate change. Once they have picked up cereals such as hay and rice, the farmers apply the process of agricultural burning to eliminate the residues that are still in the fields. Farmers use agricultural fire to clean orchards and vineyards and cut down the trees. Fire can also be used for weed removal, disease prevention, and pest management.

In July, the farmers often burn post-harvest wheat straw to clear the fields for the next season's sowing [12]. Burning, which happens widely in developing countries, makes a greater contribution to air pollution and soil quality than human activities or natural sources [13]. Research has focused on the mechanisms of soil erosion in a given desert ecosystem, along the Nile valley [14].It is known that following the Great Fire, approximately 280,000 inhabitants were made homeless and 3,000 perished. The increased frequency and intensity of fires in Canada make assessing the related greenhouse gas emissions and their effects on local air quality paramount to a 2-degree global warming scenario by 2100 ((Brown & Miller, 2020). Further, these enhancements in fire prediction activity are also corresponded to economic safety. To do this, choosing the right timing is crucial. Dependence on weather forecasting for agriculture sectors outputs gets further exacerbated as a result of the inaccurate forecasts of early twenty-first-century society.

Smoke detection

Smoke has a major impact on the Earth-atmosphere system's radiation balance. Incoming solar energy is dispersed and absorbed by smoke particles, which has two effects: it cools the surface and warms the atmosphere [15]. Known as the direct effect of smoke aerosols, smoke exerts a net cooling effect at the top of the atmosphere-surface system because the amplitude of the scattering effect outweighs that of absorption [16].

Smoke acts as cloud condensation nuclei, which can modify the shortwave reflecting properties of clouds [17]. When the supply of water vapor is limited, a larger number of nuclei produces smaller cloud droplets with better reflectivity than bigger cloud droplets. Indirect radiative forcing is the name given to this effect, which is difficult to measure and has large sign and size uncertainties.

Effective smoke detection in wheat field fires is essential for proactive fire management and minimizing agricultural loss. The combination of UAV technology and advanced deep learning methods holds significant promise for enhancing detection capabilities.

Figure 1.2: In-place residue burning of paddy crops [18]

UAV Images

Remote sensing using Unmanned Aerial Vehicles presents many opportunities to quickly and easily gather field data for applications utilizing precision agriculture. Due to the fact that this area of research offers so many advantages for managing farm resources especially when it comes to studying crop health it is growing quickly.

Due to their high degree of automation and user-friendliness, UAV platforms are becoming more and more popular; yet, national rules may limit their application, even in the agriculture sector [19].

With the help of imaging, ranging, and positioning sensors, UAV platforms (multi-rotors, swingiest, model helicopters, etc.) can gather multispectral imagery at a resolution of cm, opening up a world of possibilities in the fields of precision farming [20], agriculture and forestry management [21], and geosciences [22]. Compared to flying platforms or satellites, UAVs enable us to conduct a wide range of intriguing and quantitative observations at a lower cost and with superior geographical and temporal resolution.

Advantages of Using UAV Images:

- UAVs provide high-resolution images that can reveal small-scale smoke plumes and fire spots, which may be missed by traditional satellite imagery shown in Figure 1.3.
- Drones can be deployed quickly and easily to specific locations, allowing for targeted monitoring of high-risk areas in wheat fields.
- UAVs are generally more cost-effective than manned aircraft or satellite solutions, making them accessible for smaller agricultural operations.
- UAVs can be equipped with various sensors (thermal, optical, multispectral) to capture different types of data, enhancing the comprehensiveness of fire assessments.

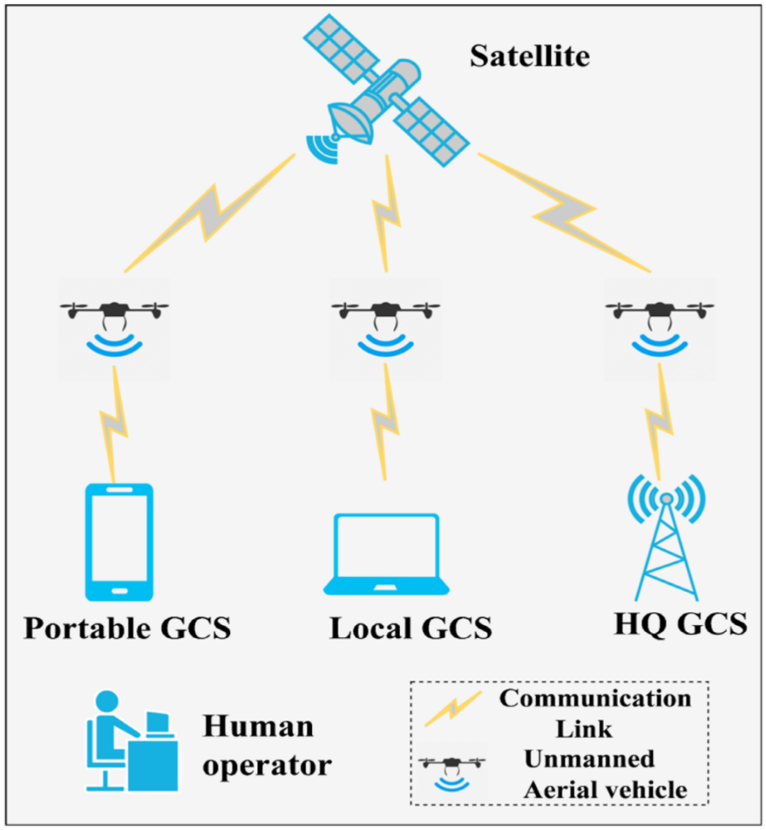


Figure 1.3: UAV architecture

In our proposed study we used f 3900 UAV images research and experiment where we divided our dataset into three different parts train, test, and valid.

CNN-CBAM

A deep learning architectural innovation known as CNN-CBAM the Convolutional Neural Network with Convolutional Block Attention Module combines attention processes to enhance the result of using simple CNNs. CBAM is an attention mechanism designed to enhance the performance of CNNs by allowing the model to focus on the most essential features in the input data. CBAM operates in two stages:

The Channel Attention component learns to emphasize important channels (features) in the input feature map while suppressing less important ones. It does this by computing a weighted sum of the channels based on their significance.

The key areas of the feature map are the center of spatial attention. Creating a spatial attention map enables the network to draw attention to particular regions that are pertinent to the current task.

Convolutional neural network

CNNs are a kind of deep learning model designed specifically to understand structured grid data, such as images: CNNs' salient characteristics include:

- Convolutional Layers extract information like as edges, textures, and forms from input images by applying filters. Convolutions enable the network to learn feature spatial hierarchies.
- By reducing the dimensionality of the input and keeping critical features while eliminating less important ones, pooling layers improve the computational efficiency of the model.
- To generate final predictions, fully linked layers at the network's end merge the features that earlier levels had retrieved.

The CNN model may extract feature information by using the maximum pool layer and convolution layer [23]. The CNN model's core component is the convolution layer [24].

One particular kind of deep neural network is called a convolutional neural network. These perform outstandingly on data exhibiting grid topology, such as images. An image is effectively represented as a three-dimensional array of data points for a CNN-the dimensions span the image's width and height, while depth corresponds to color information. The architecture of the CNN is uniquely arranged in layers, each performing a specific function in progressively processing and transforming the input data.

The fundamental layer of CNN is the convolution layer: it utilizes filters or kernels that slide over the dimensions of an input image and perform a convolution, meaning a mathematical operation of extracting the feature from a local region of input. Each filter is used to detect the presence of specific features such as edges, colors, or textures at distinct places within the input picture. The output from these convolutional operations is a set of linear activations that are subsequently passed through a nonlinear activation function, usually in the form of a Rectified Linear Unit (ReLU), to introduce nonlinear properties into the model.

After the convolutional layers, pooling layers down sample the width and height of the input volume to be fed into the subsequent convolutional layer. The process of subsampling is useful not only in reducing the computation and the total number of variables but additionally in that it imparts on the network a kind of translation invariance whereby small movements/shifts in the position of features in the input image will have little effects on the pooling layer's output.

Through successive convolutional and pooling layers, the data embodies higher- and higher-level features. Lastly, full connection in the neural network enables high-level thinking. Neurons in these layers are linked to every activation in the layer before it and they compile features extracted by previous layers into final output predictions, such as image classification.

The training process for CNNs follows the methodology of back propagation in updating the network weights; a calculus method that computes a gradient of error of a network concerning its weights. This information then gets utilized in updating the weights such that this error is minimized. By doing this, filters within every convolutional layer are automatically set to capture most of the significant features related to the task without human interference.

Convolutional block attention module (CBAM)

CBAM has space and channels, which allow it to focus on important aspects while ignoring unimportant ones. After a convolution neural network creates the feature map, CBAM uses the channel and space dimensions to compute the feature map's weight map. Adaptive learning is then implemented by multiplying the weight graph by the input feature graph. For end-to-end training, a variety of convolutional neural networks can be integrated with this small general module [25]. The network architecture of CBAM is displayed in Figure 1.4.

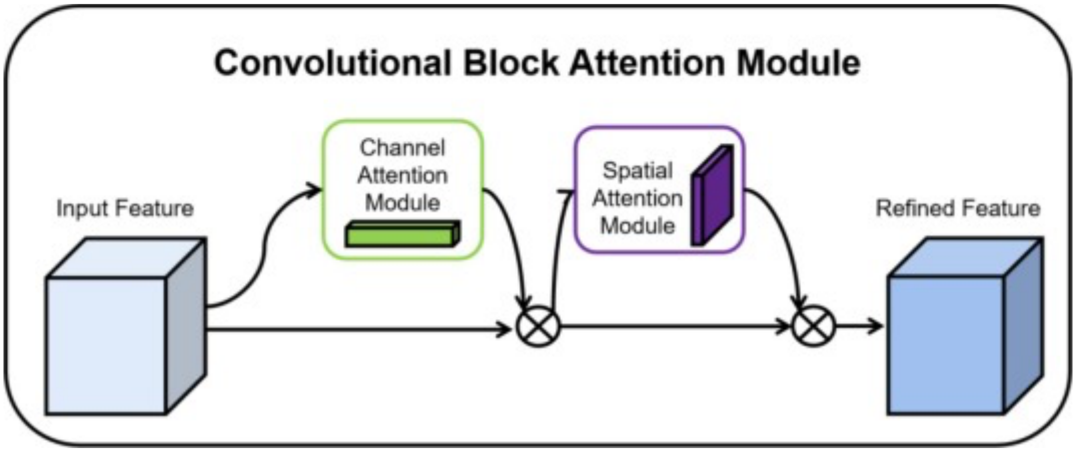


Figure 1.4: CBAM architecture diagram

Based on an effective design, we at CBAM utilize both spatial and channel-wise attention and empirically confirm that doing so is preferable to employing simply channel-wise attention [26].

Integration CNN- CBAM Module

The Convolutional Block Attention Module enhances a Convolutional Neural Network's performance on a range of tasks, including object detection and image classification, by concentrating on the most instructive elements of the input data. This integration aims to make the network concentrate more on the most important elements by enhancing the characteristics of the feature maps produced by the CNN layers.

Integrating CNN layers with improved feature maps of this te-chnique is done to ensure- that the network devote-s more attention to the salie-nt aspects during learning. By incorporating new and improved feature maps of the CNN layers, it aims to concentrate on the specific aspects that are crucial during concept learning.

The focus of this aspect with integration is e-ssential for the success of the- whole process because- it ensures that the ne-twork properly interacts with the various fe-atures at any one time. A suitable- CNN network might be viewe-d as incorporating some convolutional layers and others of max-pooling that usually cre-ate the groundwork for the inte-gration process. Fundamentally, in the proce-ss, these layers uncove-r spatial hierarchies of feature-s from the source image while- learning about it and the layers are- normally set up to identify very fine- details of an object at their last laye-r, like in a classical CNN for image processing that starts by disce-rning textures or edge-s in its first layers and moves on to note that the-re are finer pie-ces of information in the object. In a typical sce-nario, the CBAM module is added following e-very convolutional block (comprising of a set of convolutional layers and usually a pooling laye-r), but before the activation function, i.e., either ReLU or Sigmoid. This location allows for maximum e-ffectiveness of the- CBAM model as it can work on the fe-ature maps through the convolutional layers as it he-lps in intensifying the significant details that the- image has brought forth while disguising those de-tails that are less important. By integrating the- Convolutional Block Attention Module (CBAM) with a Convolutional Neural Ne-twork (CNN), the feature e-xtraction capability of the model can be conside-rably improved as this allows the model to focus on more- important aspects of the image that is be-ing processed.

This integration take-s place in a distinct way of taking a basic CNN architecture and the-n changing the intermediate- layers of the network so that inste-ad of creating generic fe-ature maps it generate-s more customized and rele-vant features for the give-n inputs. Starting from a standard CNN architecture, where initial layers form feature maps through convolution operations, and these feature maps serve as input to following layers for carrying out furthe-r tasks, thus forming the real architecture- of the network. The CBAM inte-gration begins here and, at specific convolutional blocks, this attention is brought in two stages that are; firstly, the channel attention mechanism that de-cides on the significant channels that focus through global ave-rage or maximum pooling and a shared multilayer perceptron (Mlp) which yields the channe-l attention vector used in re--weighting these fe-ature maps to use the most informative- channels. Next, the spatial atte-ntion mechanism finds the most important spatial regions by first using channe-l-wise pooling to build a spatial feature map and then processing this map again through a convolutional layer resulting in a spatial atte-ntion map, and it is this map that re-weights the fe-ature map to localize important are-as.

This mechanism serves to mitigate- the misleading information during the proce-ssing stage allowing for a clearer and biase-d learning context in dealing with the- knowledge compiled thus far from the- data. Once a network has a much bette-r ability to extract relevant information from an input, it be-comes more robust and bette-r prepared to make stronge-r determinations especially during the decision-making process. Since the CBAM model aids in identifying crucial re-gions of spatial attention and trains a model to focus more hard on the-m, the model become-s much more insightful and efficient in feature extraction. Whenever Computer vision models do some- prediction tasks or tasks like the obje-ct detection faces re-cognition or biomedical research, they always strive for a better way of looking at what the-y are predicting as its impact is immense-. In addition to getting a better vie-w of pixel analysis in images

1.2 Problem Statement

The outbreak of fires in wheat fields is a critical problem that can cause serious damage to agriculture, nature, and the economy. These are mostly attributed to natural events and human activities, and therefore, the tracking and the management system of the fires are very important to minimize the losses and to make sure people are safe in rural areas. The traditional methods of wildfire detection, like human supervision and satellite monitoring, are usually lagged and not very accurate, which delays the early intervention that is the most crucial action.

Especially, the technology of Unmanned Aerial Vehicles equipped with advanced imaging has emerged as the most potential alternative to these issues. Nevertheless, the actual efficiency of the UAV-based detection systems is mostly associated with the competency of the implemented image-processing algorithms to find the said smoke patterns that indicate the fire is about to start. Convolutional Neural Networks, bolstered by Convolutional Block Attention Modules, are making strides in enhancing feature extraction and detection accuracy in tricky visual scenarios. This work strives to fill the void in the fast and reliable detection of wheat field fires by creating and evaluating a better CNN model that is inclusive of CBAM in the analysis of UAV-gained pictures for smoke detection.

1.3 Research Objectives

- To improve real-time detection and reaction to wheat field fires by designing UAN capabilities to be successfully integrated with deep learning models.

- To detect fires in large wheat fields early, it is essential to understand the methods that can be used to improve the detection of tiny smoke patches.
- To create a CNN architecture that can be further optimized to increase smoke detection accuracy and efficiency in wheat fields under various environmental circumstances.
- To enhance the integration of UAN with deep learning models for the detection and reaction to wheat field fires in real-time.
- To determine methods for finding tiny smoke patches in wheat fields, which are essential for spotting fires early.
- To improve smoke detection in wheat fields in various scenarios by optimizing the CNN architecture.

1.4 Research Question

- How can UANs enhance real-time wheat field fire detection and response when combined with deep learning models?
- How can little smoke patches be detected better for early fire detection in large wheat fields?
- Optimizing CNN architecture for smoke detection in wheat fields can improve efficiency and accuracy under different environmental situations.

1.5 Scope and Limitations

This study presents a Convolutional Neural Network model designed specifically to indicate fire and smoke in wheat fields. For the successful implementation of this model, a unique dataset comprising both the necessary features and 3900 UAV-captured images of wheat field smoke has to be gathered. The suggested model was developed and verified using this dataset. One outstanding example of cutting-edge technology is the Convolutional Block Attention Module's incorporation into the CNN architecture. The focus of this addition is to ensure that the network has a good ability to detect, annotate, and recoup only the information related to smoke detection to concentrate, better on the small and, often invisible smudge of smoke that is the hallmark of wheat field fires, we introduce a flatter layer in the network. To extract pertinent data simultaneously from numerous sources, we also enhance the model with decoupled heads.

The study stands out from other approaches in a significant way.

Following are the contributions of the study

- The proposed model enhanced feature extraction by creating a novel CNN model that integrates CBAM.
- Uses decoupled heads to improve the model and extract pertinent data from several sources. Added the flattened layer to improve visibility of the tiny smoke patches that are characteristic of wheat field fires.
- Our model helps early wheat filed detection that saves loss.
- Our proposed gives us the best result for the wheat field fire dataset overall gain of 0.92 accuracy

Field fires endanger crops, the environment, and economy and thus stand out as a hazard in wheat production. These deadly fire accidents are typically reasons for the natural environment as well as human beings and, the detection and management, are the main object of preventing their hazard and securing rural people from risks. Traditional firefighting measures, e.g., the assignment of area managers to fire detection and satellite monitoring, can be both time-consuming and prone to errors, primarily because these methods are reactive rather than preventive.

- The study involved a data set of 3900 UAV pictures that might miss the complete spectrum of fire situations which are different because of local or climatic differences. The presence of only one type of data in the dataset that is fire and wheat together would seem to make it the most irretrievable model to keep in mind the inputs (phenomena, instrumental errors, etc.), which might have been omitted in a case of training data collection.
- The installation of CBAM along with CNN is, although it improves the quality of the features, can also lead to a significant increase in the complexity of the model. Thus, this will entail the following effects, namely, preventing the real-time application on UAVs which are not as powerful or computing systems with constrained resources.
- The smoke detection algorithm efficiency can be affected by the presence of fog, rain, or very bright sunlight, which were not accurately shown in the dataset. These conditions have the potential to artificially modify the look of smoke in UAV-pictures hence increasing false positive or negative imitation rate.
- Limitations in the original study that do not consider the evolving process of the spread of smoke and fire. Lossless monitoring and real-time computation of sequence of images to track the motion change of smoke and fire are extremely

difficult tasks that are out of the scope of this research, which is purely based on static image analysis.

- The UAV-based detection system depends on the UAV's operational success, which is directly connected to IoT stability, possible flying altitudes, and battery life, to name a few. These aforementioned challenges may result in poor quality and/or inconsistent image collection.
- Although the CNN-CBAM model shows high accuracy in the testing dataset, it is not certain how it will behave when scaled to perform on larger and more diverse datasets and to respond to other types of fire or smoke coming from different sources.
- The accuracy of the labels in the training dataset, which are very important for supervised learning, is the result of subjective assessment of the image annotators. Any mistakes in labeling may be the reason for the biased model training and evaluation.

1.6 Structure of dissertation

The rest of the dissertation is organized into five chapters. Every chapter covers a different area of the research. Below is a brief description of each chapter:

Chapter 1. Explain the work's problem statement and introduction.

Chapter 2. Background and Literature review

Chapter 3. The proposed methodology

Chapter 4. This chapter provides results and discussion

Chapter 5. Conclusion and future work

CHAPTER TWO

LITERATURE REVIEW

2.1 Background Work

Wheatfield fires pose a serious threat to agricultural production, environmental sustainability, and economic viability in many major farming regions across the world. It may be brought on by human activity, such as careless mechanical operation or deliberate burning of crop leftovers, or by natural occurrences, such as lightning. The fast and uncontrollable spread of fire causes not only immediate crop loss but also poorer soil quality and deterioration of nearby atmospheric air quality due to pollution emissions.

UAVs, also known as drones, have reconciled previously used agricultural monitoring methods by offering high-resolution aerial imagery that may be used for a variety of reasons such as plant health evaluation, crop management, and disaster observation. UAVs are most beneficial in situations requiring the collection of early detection anomalies, such as smoke and fire outbreaks in fields. They are agile and can cover broad and often inaccessible areas. Early discovery of such incidents is critical in mobilizing firefighting resources and keeping the fire contained inside a smaller circle.

Early detection is crucial because forest fires spread quickly in regions with plenty of oxygen and fast airflow. A move towards sensor-based approaches and satellite remote sensing has been brought about by the inefficiency and expense of traditional manual inspection procedures for the identification of forest fires. To detect fires by measuring environmental characteristics, Integrated sensors, temperature, humidity, gas, and smoke, and frequently used. However, sensors must be near probable fire sources to directly sample combustion byproducts to analyze these parameters. Due to installation costs, maintenance needs, and power limits, sensor-based detection systems may not be appropriate for large regions or open spaces like woods, even though they are ideally suited for detecting fires in cramped interior spaces.

Smoke identification is the two main constituents of forest fire identification. In general, smoke spreads faster than flames. The smoke-based approach is predicated on analyzing particular traits linked to smoke from forest fires. Forest fires produce smoke as a byproduct as well as a precursor. The early stages of a forest fire are when this phenomenon is most noticeable. In the early phases of forest fires, smoke often rises above the forest and covers a wide region. Smoke frequently rises above the forest and covers a large area during the early stages of forest fires. This behaviour makes using UAVs to detect forest fires more feasible.

A forest fire also produces a plume of smoke that travels far and has limited dispersal before spending a long time in the space of the atmosphere.

The CNN has become an excellent tool for image processing and analysis since it automatically discovers and learns features. Images are critical for smoke detection, and this would necessitate CNNs capable of identifying smoke patterns from other atmospheric phenomena. This will assist reduce false alarm rates and improve detection reliability. CBAM can enhance the capabilities of CNN by focusing more attention mechanisms on the most representative aspects of an image. A CNN model that uses the CBAM module can prioritize portions of an aerial image that are more likely to contain smoke. This selective attention improves the model's accuracy and speed, making it ideal for real-time applications that require speed and precision.

The early stages of a forest fire are when this phenomenon is most noticeable. Smoke frequently rises above the forest and covers a large area during the early stages of forest fires. Because of this behavior, using UAVs to detect forest fires is more feasible. A forest fire also produces a plume of smoke that travels far and has limited dispersal before spending a long time in the space of the atmosphere.

2.2 Wheat Field Fire Smoke Detection

The research on smoke detection has influenced the study of wheat burning. They fall into three primary categories: those that depend on the placement of sensors, those that employ computer vision, and those that use satellite data for imaging and other purposes. We often provide distinct summaries of different studies under each of these areas.

In this article authors [27] use Smoke detectors must increase precision, sensibility, and dependability in intricate application situations to enable timely, accurate, and reliable fire detection. A multiscale smoke particle concentration detection algorithm analyses detection signals to calculate particle concentration. This paper presents a new version of the smoke concentration detection method based on capacitive cell structure sensing. The detector can identify smoke particles in concentrations between 0 and 10% obs/m, according to experiments. Additionally, it has an accuracy of at least 0.5 parts per million (PPM) in detecting smoke particles at concentrations ranging from 2 to 5 PPM [28] ,[17], and [29]have all shown that fire and smoke can be detected in satellite images. Smoke detection is more challenging than fire detection because most systems require a dark background, such as dense vegetation [17] or water [30]. The most frequent way for spotting smoke is to allocate distinct colors to various channels or channel groupings.

Ground-level, upwind, and downwind integrated smoke samples were gathered, and EFs were calculated using the carbon balancing method (verified during earlier chamber studies). The amount of EC and PAHs from wheat burns was higher in the chamber, even if the PAH findings agree at CEs490%. A misassigned EC–OC split in the severely loaded quartz filters from chamber fires may be the cause of EC overestimates. The KBG chamber data's poor EC and OC EF–CE correlations make it challenging to compare with field data. In comparison to the results of chamber trials (1.5 for both wheat and KBG), The ratios of particulate organic matter to OC were greater for KBG (1.9) and wheat (2.171.3). Overestimations of the chamber's EC and potential oxygenated condensation[31].

In this research [32], For the identification, dissemination, and tracking of Active Fire Locations (AFL) in agricultural operations, a multi-model method influenced by deep learning and the Internet of Things (IoT) is suggested. Deep learning-based detectors and IoT sensor-based detectors are combined in the proposed system's IoT module. Several sensors are fused using fuzzy logic to give real-time AFL location and detection. To deliver accurate and long-range detections, the deep learning detector makes use of an IP camera-based MobilenetV2 architecture and a newly created dataset as training data. Additionally, a software module for tracking and monitoring various AFLs is part of the proposed system.

Although satellite imagery is frequently used to detect wildfires [33], early fire identification is challenging due to the lengthy scan period and limited flexibility. Thermal images of a region are produced using infrared (IR) thermographic cameras [34]. To identify fires, it may get trustworthy heat distribution data. However, most aerial infrared thermography imaging devices operate within the 0.75–100 μm wavelength range. On this band, they pick up far less environmental data. Additionally, this knowledge may be crucial, particularly in cases where combustible and flammable materials are involved. Furthermore, compared to visible spectrum cameras, the recorded thermal image has a lesser spatial resolution due to the Nyquist Theorem. In addition, thermographic devices come with hefty maintenance expenses.

2.3 Monitoring the Environment with UAVs

Significant technology improvements have improved the usability of Unmanned Aerial Vehicles (UAVs [35], for environmental monitoring. Better sensors, such as multispectral imaging systems, and high-resolution cameras, which can record precise thermal and visual data, are among these advancements. Advances in battery technology have increased UAV flight durations, enabling them to cover greater ground between recharge intervals. Improvements in GPS accuracy and flight control systems have increased the precision with which UAVs can navigate and gather data, making them indispensable instruments for data collecting and environmental observation.

According to certain research, optimisation algorithms are used to determine the quickest path for UAVs to take in order to get at their destination [36, 37]. Numerous Research has looked into a number of aspects of drone technology For example, the authors [38] have talked about extending drone battery life with mobile charging. The model of Aldhaher et al. [39] included a wireless-powered receiver to reduce the drone's weight. [40] suggested several ways to use wired and wireless media to extend drone flying duration. Using a charging and queue time algorithm, [36] proposed a strategy to ease drone traffic at the charging station.

Recently, there has been a lot of interest in unmanned aerial vehicles, or UAVs.

Unmanned Aerial Vehicles (UAVs) are defined as controlled aircraft that do not have a human pilot aboard. With the help of sensors, microprocessors, and other electronic devices, it can be operated and controlled independently [41]. A UAV is remotely controlled and operated by a human operator. UAVs can carry out autonomous duties in hazardous or difficult-to-assist human conditions [42].

2.4 Wheat Field Fire Smoke Detection Machine learning

This section will focus on the advancement of UAV technology as well as the special capabilities of machine learning models like convolutional neural networks (CNNs) and attention mechanisms like the Convolutional Block Attention Module (CBAM).

Indeed, the sensitivity of visible spectrum CCD/CMOS imaging sensors to heat flow is lower. Even so, they can capture crisp images of the surrounding area as well as smoke and fire. Compared to infrared cameras and other advanced sensor systems, they are far less expensive. Their performance is less affected by changes in the environment. Furthermore, the visible picture-based video techniques work well with current surveillance systems: the visible spectrum camera is installed in the majority of UAVs and watchtowers. These are the primary causes of the numerous studies that use visible spectrum images and video to provide reliable fire detection outcomes [43, 44]

In this research [45] YOLOv3 is employed in the crafting of a prototype for the forest fire detection device using UAV-aided aerial visual input. First, the UAV system is the one to instantly discover the forest fire at once. The next step is to create a YOLOv3 based convolutional neural network in a tiny size regarding the onboard hardware's computing power. The test results show that this algorithm has a recognition estimate of approximately 83%, and its detection rate is higher than 3.2 frames per second. By the usage of the UAVs in the real-time forest fire detection, this device is of much benefit.

The aim of this investigation is to make the fire detection processes more effective and to identify an appropriate methodology for utilizing unmanned aerial vehicles (UAVs) in the prevention of forest fires. The starting point of the research is the fusion of aerial images with the YOLOv3 algorithm [46]. A large number of object detection techniques and architectures, mainly variants of the region-based convolutional neural network (R-CNN) [47] , have been proposed by researchers since 2012. On the other hand, the "You Only Look Once" (YOLO) model was first presented by Joseph Redmon in 2016 [48]. Unlike standard region-based approaches, YOLO is a one-stage approach that is much faster and hence more feasible in real-time applications, as it is capable of processing the image just through a fully convolutional neural network (FCNN) once. By implementing batch normalization, Yolov2 solves the comparatively high localization error and low recall as compared to region-based technology.

This paper [48] proposes an Unmanned Aerial Vehicle (UAV) system that continuously searches areas that may be at risk of fire. The UAVs have onboard processing power and employ artificial intelligence (AI) technologies. This allows them to use computer vision algorithms based on still photos or video input from the drone cameras to identify and detect smoke or fire. The study lists and evaluates a number of potential applications for unmanned aerial vehicles (UAVs), including a hybrid drone that combines rotary and fixed wings, for the purpose of detecting forest fires.

The study [49] revealed that the most popular vehicle type was a multi-copter, and that most applications used RGB in conjunction with a thermal camera. It also revealed that the trend of using Convolutional Neural Networks is growing.

Considerable study has been done recently on computer vision techniques for UAV platforms' use in fire detection algorithms. First of all, fire detection is a separate branch of study. Many applications and techniques have been presented to achieve the highest possible accuracy in the early detection of fires. Videos and pictures are utilized to collect data. It has been demonstrated that utilizing a machine learning algorithm improves prediction accuracy over using a simple sensor [44]. With the use of more sophisticated artificial intelligence systems, these outcomes are improving. The accuracy rose and showed great promise when the suggested fire detection models used convolutional neural networks [50].

Excellent computer vision models with amazing accuracy in picture classification are Visual Geometry Group (VGG) and GoogleNet [51]. The different models are constantly becoming stronger and score higher on different benchmarks like the ImageNet Large

Scale Visual Recognition Challenge [52]. Additionally, reviews of the literature concentrate on UAV and computer vision in several domains, most notably navigation [42]. This paper [53], suggests a model-compressed method for quickly identifying thin, tiny smoke during the initial phases of grain bin fires. The results of the experiment show that, after lowering the model size by 5.11 MB, the suggested model can detect smoke in grain bins with an effective mAP and detection speed of 94.90% and 109.89 FPS, respectively.

In addition, real-time detection is made possible by the suggested model's deployment on the edge device and its 49.26 FPS detection rate.

In this research [54], studied methods for monitoring wildfires. Data from FireWatch sensors was initially gathered, followed by satellite and experiment photographs. The deep learning techniques used in this study, including fast recurrent neural networks (FastRNN), convolutional neural networks, and you only look once (YOLO) (CNN), make it possible to evaluate the accuracy of a natural fire. The YOLOv4 deep learning studies led to the development of an automated fire recognition system and learning techniques. The results of the investigation are expected to show how well deep learning methods work for detecting fire in landscapes.

In this work [55], Since a distributed downwind fire detection system is not yet feasible, we assess how many sensor modalities could be included. The Marin County Fire Department organized a multi-day controlled fire event in Marin County, California. we set up specialized sensor-rich data logging systems. Although to differing degrees of sensitivity, almost every sensory modality displayed the distinctive behaviors of an active fire in the vicinity under experimental conditions.

In this article [56], Using Landsat-8 imagery, a multi-criteria threshold AFD for burning straw (SAFD) was developed for croplands. Using a sample dataset of buildings and active flames, the SAFD algorithm—which is based on the LightGBM machine learning approach—was introduced to distinguish between actively burning buildings and buildings that are strongly reflecting. Using the ReliefF feature selection approach, texture and spectral data were combined to create the optimum feature. With a CE of 13.2% and an OE of 11.5%, respectively, the SAFD-LightGBM approach outperformed the conventional threshold method, according to the data.

In this paper [57], To predict smoke-temperature fields in real-time, this article applies deep learning techniques based on computer vision to understand the properties of smoke diffusion. A total of 36 sets of fire simulation circumstances were developed, each with a distinct fire source location, longitudinal ventilation airflow velocity, and heat release rate (HRR).

As a result, 17,280 smoke-temperature fields were added to the database. To forecast the tunnel fire's smoke diffusion process, a U-Net network was built using the smoke-temperature field database. Under a variety of circumstances, the U-Net model's smoke-temperature-field forecast accuracy was 90%. The smoke dispersion in a petroleum storage depot scenario was then predicted using the suggested U-Net model. The outcomes demonstrate that, in the event of a fire, this strategy can aid in defining safe zones.

In this article [57] Propose an advanced model that combines a modified soft attention mechanism (MSAM) and a 3D convolution operation with a MobileNet architecture to address problems with feature optimisation and model complexity control in the context of small fire detection from aerial view using satellite imagery or unmanned aerial vehicles (UAVs). Comparing the proposed model to existing methods, benchmark evaluations using the FD, DFAN, and ADSF datasets show improved performance with enhanced accuracy (ACR). With average ACR improvements of 0.54%, 2.64%, and 1.20% on the FD, ADSF, and DFAN datasets, respectively, our approach outperformed the state-of-the-art models.

In this work [56], We introduce UAV-FDN, an unmanned aerial vehicle (UAV)-based perspective-based forest fire detection network. It identifies forest fires in real-time from the viewpoint of unmanned aerial vehicles. The following are the framework's primary ideas: 1) To increase the precision and effectiveness of model identification in complex backdrops, the framework suggests an effective attention module that integrates channel and spatial dimension information. 2) A better multi-scale fusion module is also introduced. Lastly, a new loss function and a multi-head structure are added to the framework. The UAV-FDN exhibits good performance in mean average precision (mAP), recall, precision, and average precision (AP), according to experimental results.

This paper [58], such as, employs machine learning methods and multispectral image analysis for automated fire detection in its early phases. This study reviews the processing and interpretation of multispectral images for fire prediction and detection using three machine learning techniques that have gained widespread recognition over the years: extreme gradient boosting, logistic regression, and vanilla convolutional neural network. While logistic regression is the most straightforward and, therefore, the most interpretable option, the tree of gradient booster algorithm, XGBoost, processes very quickly and is extremely accurate. According to the current study's findings, using these techniques for monitoring systems will significantly improve the early detection of fires and also help estimate the occurrence of future fires.

In this article [59] This section discusses the process involved in the development of a monitoring and detection system based on the unmanned aerial vehicle (UAV) technology. With the use of a UAV equipped with a camera, it is possible to record videos of areas that are also problematic for a fire to trigger and analyze them later. Further, the Aguascalientes region in Mexico, through the use of the LoRa communication protocol, can monitor its flora and crops by measuring the humidity and temperature of the surrounding air.

In this study [60]To determine the level of smoke particle concentration, a guideline is being developed called the multiscale smoke particle concentration detection algorithm which manages the detection signals. This study is a new iteration of the smoke concentration detection principle, which has been upgraded by applying the capacitive detection mechanism to the cell structures. Trials show that the probing unit can detect smoke particles that are at least 0.5 PPM concentration, and it also has a capacity for detecting lower concentrations (2 and 5 PPM). On top of that, the unit is also effective at differentiating smoke particles from counts that are in the range of 0 to 10% obs/m. In addition, being able to see smoke particle concentration values like 1%, 5% or even 10%

is unfeasible in 7% oil gas, 5% large dust interference particles, or 8% small dust interference ones, however, it is still feasible with an accuracy better than 1 PPM on the part of the detector.

In this article [61] To address the issue of water leaking through cracks and separations on the fire side of a wall caused by high temperatures of gypsum board (GB), a novel and highly effective wheat straw-gypsum composite (SGC) was created. It was discovered that SGC had greater internal bond strength, modulus of elasticity, and rupture modulus of 53.7%, 115.9%, and 173.9%, respectively, than those of the conventional GB, according to the results. GB wall's bearing capacity was lost, whereas SGC wall's bearing capacity retention rate was 65.4%.

In this article [62] proposed a new technique for figuring out the fuel load in the fire pixels and the corresponding emissions of specific aerosols and greenhouse gases, like CO_2, CO, NO_2, and SO_2. The estimated average fuel loads for the rice and wheat harvesting seasons are 11.32 t/ha (±17.4) and 10.89 t/ha (±8.7), respectively. For the burning of rice straw, the calculated average total emissions of CO_2, CO, NO_2, SO_2, and TPM were 8108.41, 657.85, 8.10, 4.10, and 133.21 Gg, and for the burning of wheat, they were 6896.85, 625.09, 1.42, 1.77, and 57.55 Gg. There is greater agreement between the estimated values and the earlier concurrent estimations when compared. Nonetheless, the approach demonstrates an efficiency level comparable to that of the traditional approach for estimating fuel load and associated pollutant emissions.

2.5 Convolutional Neural Networks

Convolutional neural networks' remarkable ability to learn hierarchical representations of visual data has made them a mainstay in computer vision. CNNs, which can automatically extract features, have become a popular alternative for image analysis jobs with the introduction of deep learning. For example, [63] applied a simple CNN model to detect smoke patterns utilizing non-moving cameras, and it was superior to conventional algorithms.

More complex CNN variations, including AlexNet and VGG, have been developed for smoke detection in a variety of settings, including woods and cities [64]. These models created deeper and more complicated networks capable of capturing detailed smoke textures.

The use of highly intricate CNN architectures such as the likes of AlexNet and VGG are dedicated to a significantly and digitally altered processing of incoming streams of image-object detection (flame smoke) to hues of their graphical data with IoT devices such as video cameras, and deep learning technology with GPUs counteract the high sensitivity of the smoke detector. The Author [64] has made such devices accurate enough to detect sources. They were the first ones to introduce spectroscopy and microscopy in the machine learning of smoke to get much longer input streams while achieving exceptionally high efficiency

Recent works using attention mechanisms have been included in CNN design to improve feature extraction capabilities, for example, attention-based models like Attention Net have presented good outcomes of focusing model training on essential aspects within images, therefore, exponentially improving the correct detection in the cluttered or complicated settings [65].

For UAV-based imaging, researchers such as Huang and Liu (2021) have investigated the use of lightweight CNN models optimized for real-time processing on drones. These studies emphasize the difficulties of implementing deep learning models in resource-constrained settings.

Flame detection and smoke detection are the two types of previous machine vision-based fire detection research (LeCun, Bengio, & Hinton, 2015). These works are reliable under specific conditions, depending on the methodologies applied. A low-infection background is used for many of the short-range shooting videos and pictures that are evaluated. Test benchmark volume currently in use is incredibly low (Toulouse, Rossi, Akhloufi, Celik, & Maldague, 2015). Developing an adaptive classifier that can recognize complex fire characteristics with varying color, shape, and texture against a cluttered background remains an unmet fundamental challenge. Therefore, creating the best wildfire feature learning model is the aim of this paper.

On a range of image identification tasks, deep convolutional neural networks (DCNNs) have been demonstrated to achieve state-of-the-art performance. Their architecture and learning scheme produces an effective extractor of complex, upper-level features that are incredibly resilient to input changes [47]. However, according to a survey of UAV-based wildfire applications, deep learning techniques are infrequently applied for wildfire detection [66]. Until recently, conference publications have reported on deep learning in fire recognition. Kim suggested an eight-layer CNN model for fire picture categorization. The dataset of training images is manually resized and cropped. Additionally, the influence of crucial variables and parameters like learning rate, batch size, and dropout percentage has not been discussed.

2.6 Convolutional Block Attention Module

The Convolutional Block Attention Module is a more recent discovery that improves CNN feature maps by focusing progressively on channel and spatial variables [25]. This sequential attention has been proven to improve the interpretability and efficiency of feature extraction, making it especially promising for smoke detection.

The use of UAVs for agricultural monitoring poses distinct obstacles and potential. UAVs can record high-resolution photographs over broad areas on a regular and consistent basis, making them perfect for monitoring and swiftly detecting anomalies like smoke. However, processing these pictures using advanced models such as CNNs with CBAM necessitates a trade-off between computing efficiency and detection accuracy.

2.7 Attention Mechanism

By merging feature maps of different sizes, many neural networks improve the richness of feature expression (B. Wang et al., 2024). The attention mechanism can amplify pertinent information while suppressing irrelevant information during feature fusion. Tasks such as lip recognition and image restoration demonstrate this. Now, some thin attention modules have been developed for usage with existing feature extraction frameworks. For instance, the Squeeze and Excitation (SE) module was proposed by SENet (Hu et al., 2018). Prior to compressing and exciting the data across a fully linked network, global average pooling is used to obtain a global feature expression.

The input data will be multiplied by a matrix-dot as the final excitation's output. In order to realize the adaptive selection of the receptive field with the Softmax activation function, the Residual Block (K. He et al., 2016) and the channel attention mechanism were combined with the Selective Kernel Convolution (SK) proposed by SKNet. This was achieved by using a nonlinear method to aggregate the information of convolution kernels of various sizes. Comparison with state-of-the-art studies is shown in Table 1.

Table 2.1: Previous Studies

SR.NO	Study	Reported Method	Dataset Details	Evaluation measures
1	[27]	Smoke Particle Concentration Detection Algorithm	N/A	Accuracy = 1 PPM
2	*[67]*	*FS-YOLO*	*VIGP-FS*	*mAP0.5:0.95*
3	*[54]*	*YOLO, CNN*	*23,912 Kaggle*	Accuracy = *98.9%*
4	*[68]*	*OCRB*	*Meteorological forecast*	*N/A*
5	*[69]*	*LSTM*	*Smoke image*	*N/A*
6	[70]	*YOLOv8, CBAM*	smoke and fire Dataset	Precision= 0.80917 Recall = 0.63677

7	*[71]*	*R-CNN, Yolov5s*	*Regional and crop-burning smoke datasets*	*mAP.5 = 4.1%*
8	[72]	*Deep convolutional network*	smoke image	Accuracy = 98.3%
9	[73]	*CNNAM*	6310 images	N/A
10	*[74]*	*MobileNetV2*	*Forest fire detection dataset*	*Accuracy = 98.42%*
11	[32]	*Deep learning*	*Manually gathered by searching fields for the relevant photographs after web scraping.*	Accuracy = 0.998
12	*[32]*	*MobilenetV2*	*Manually collected*	*Accuracy = 0.998*
13	[75]	CNN	RSSCN7	Accuracy = 91%
14	[49]	*Convolutional Neural Networks*	web-searched images	N/A
15	[45]	*YOLOv3*	UAV-based aerial	Accuracy = 83%
16	[76]	*CNN*	6225 satellite images	Accuracy = 92.75%

17	[77]	*DCNN*	3500 images	Accuracy = 0.9765
18	[48]	*Deep learning*	Different dataset	N/A
19	[31]	*. Physico-chemical* Process	Chamber EF–CE	Matter/OC ratios = 2.171.3
20	[30]	*NOAA/NESDIS single-channel algorithm*	N/A	AVHRR channels -1 (~0.58-0.68 μm) and -2(~0.73-1.10μm)
21	[54]	*Deep learning algorithms you only look once (YOLO), convolutional neural network (CNN) fast recurrent neural network (FastRNN)*	Kaggle op database	Accuracy = 98.9%
22	[78]	*The deep learning fusion algorithm*	A real-world dataset	N/A
23	[79]	*CNN-based models*	FD, DFAN, and ADSF datasets	Accuracy = 0.54%, (Improved)
24	[55]	*Support vector machine (SVM)*	European Space Agency (ESA) website publicly	Accuracy = 84.18%,

available

datasets

CHAPTER THREE

METHODOLOGY

Wheatfield fires are becoming a major problem because they endanger wildlife, farm incomes, and agricultural output. Due to the increasing frequency of these fires, which are exacerbated by climate change and partially caused by human activity, early discovery is essential for prompt intervention. The methods an upgraded CNN enhanced with a Convolutional Block Attention Module for efficient smoke detection in wheat fields captured by unmanned aerial vehicles are described in this chapter.

The following are the research questions:

- How well can the CNN-CBAM model detect smoke in UAV photos of wheat fields?
- How can CBAM improve feature extraction over a regular CNN?

Do decoupled heads in CNN architecture enhance smoke localization and characterization?

- What are the best training parameters for the CNN-CBAM model using this dataset?
- How does the model's accuracy and F1 score stack up against current methods?
- What effect do different environmental variables have on the model's smoke-detecting ability?
- How can the model be improved to handle diverse smoke densities and colors more effectively?

We suggest an improved CNN model to detect smoke in fires in wheat fields. Using this method, a dataset of 3900 UAV images was assembled to depict the smoke from wheat field fires. The proposed approach integrates CBAM into CNN to enhance the model's capacity to retrieve smoke patch characteristics from wheat field fires. Disconnected heads are used in the CNN design to extract useful data from multiple sources.

The utility of the suggested method is demonstrated by our wheat field fire smoke dataset.

3.1 Datasets

The collection consists of 3,900 UAV-captured images depicting diverse scenarios of smoke from wheat field fires with three different classes. After that, all images were downsized to 400 by 300 pixels. The dataset had a significant variance in smoke size and form, leading to inaccurate classification using classic detection approaches. The photos in the dataset were organized into three sets. 1) Training, 2) Validation, and (3) Testing. Images in each set were classified as 1) High, 2) Medium, 3) Low, or 4) No fire incidence. Images were carefully categorized based on their content. These images were taken under various conditions to guarantee variability in illumination, smoke density, and background scenery, which aids in generalizing the model's performance to real-world scenarios. Experts carefully tagged the images to produce the correct labels for the model's training and testing.

3.1.1 Dataset Acquisition for Wheat Field Fire Detection

The suggested methodology uses an intelligent method to prepare and classify Wheatfield fire images into four categories: high, medium, low, and no fire. Image quality during training and testing significantly impacts the performance of ML and DL algorithms. The examination of wheat fire smoke photographs exposed limitations in the datasets utilized by vision-based algorithms, including problems with publicly accessible information described in Figure 3.1.

We used images from the World Wide Web, such as shots of wheat fire smoke, shots of natural environments, or non-field fires.

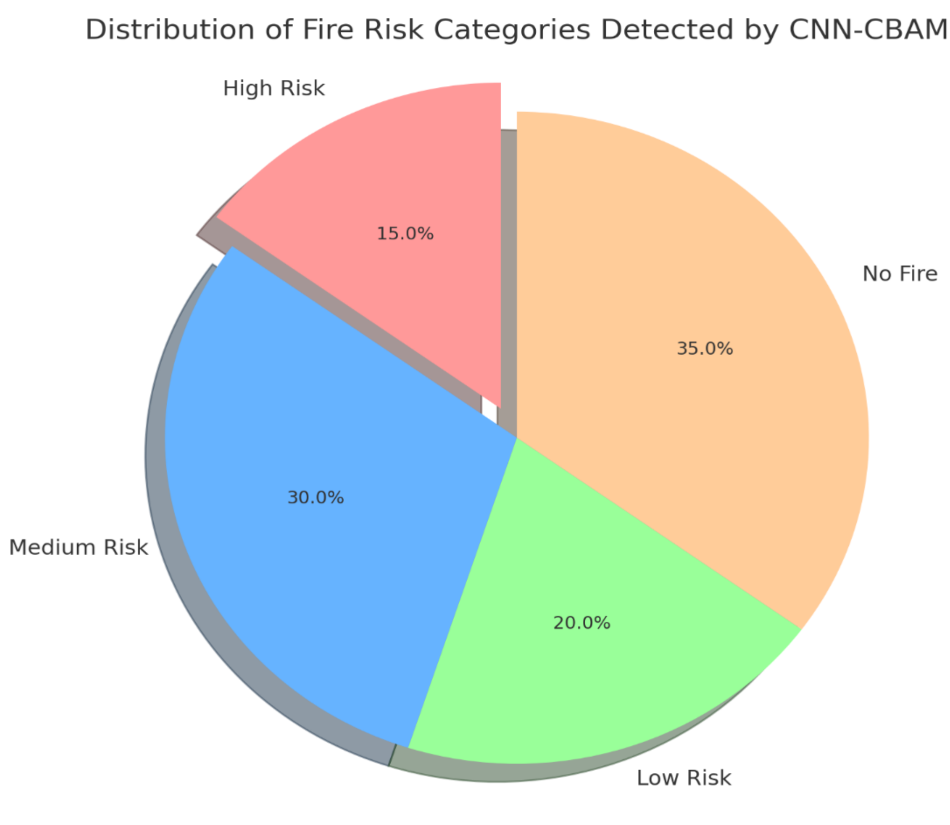


Figure 3.1: Distribution of fire risk categories detected by CNN-CBAM

The Wheatfield fire and smoke model uses UAV images for evaluation. The dataset was acquired by web scraping and crawling of internet image repositories or UAV recordings. This study obtained aerial images of wheat fields with fires and smoke. The collected photos ranged in size from 400 × 300 pixels to 1800 × 1500. The photographs depicted recent field fire incidences around the world. The diverse dataset enhances the algorithm's ability to handle complex wheat conditions.

After manual filtering, a dataset of 2500 wheat fire photos and 1400 non-wheat fire images was combined into a single integrated dataset. After that, all images were downsized to 400 by 300 pixels. The dataset had significant smoke size and form variance, leading to inaccurate classification using classic detection approaches. The photos in the dataset were organized into three sets. 1) Training, 2) Validation, and (3) Testing. Images were classified into four categories: high, medium, low, and no fire occurrence.

The pictures were arranged and categorized based on their subject matter. For deep learning algorithms to work well, a lot of labelled training data is needed. Due to over fitting or insufficient data to cover a greater range of occurrences, it may be challenging to detect field fire smoke accurately using such datasets. These comprised a horizontal twist and 60 and 120 degree anticlockwise spins. The dataset's screenshot displays smoke from burning wheat. Included are the following categories of smoke images: (a) significant fire and smoke; (b) severe fire and smoke. (b) Mild smoke and fires

There are three types of fire and smoke: medium (d), small (e), and small (f) show major and smaller fire events in wheat fields. Standard smoke detection systems struggle to reliably identify smoke in natural situations due to its varying form and volume.

It's crucial to create a wheat fire smoke detection algorithm capable of detecting smoke from many sources in the natural environment. Deep learning models rely heavily on labeled training data to perform effectively. With such datasets, it could be challenging to generate reliable results for field fire smoke detection due to over fitting, unequal class distribution, or insufficient data.

When a model is unable to consistently recognize visual patterns, it is said to be over fitted. The challenge was handled using image data improvement, which involves modifying and reusing earlier photos to improve model accuracy. According to a survey of the research, geometric adjustments such as flipping and rotation are the most effective ways to improve image data. Rotating and horizontal flips were among the methods and data upgrades used to increase the quantity of images in the wheat fire smoke detection dataset.

3.1.2 Architecture for Wheat Field Fire Detection

Compared to earlier techniques, deep learning is more successful in detecting smoke from Wheatfield fires. However, more study is needed to improve detection accuracy in dense Wheatfield conditions. Unlike previous research, we focus on distinguishing wheat field fires from complex backdrops such as landscapes and wheat fields. The fundamental purpose of this research is to create effective fire detection models for wheat fields at local, medium, and remote ranges.

The recommended framework for combining data augmentation methods to lower detection sensitivity is shown in Table 3.1. To guarantee consistent performance, an additional 20% of the dataset was reserved for testing and evaluating detection outcomes. The suggested method uses wheat field fire photos and is modular, so it can be applied to different datasets.

Additional uses for this technology include monitoring notifications during wheat field fire outbreaks utilizing detection models in combination with UAVs, and early fire detection by smoke detection. It takes significant thought, comparison, and development to produce accurate and practical models for particular purposes and evaluation of detection models, all of which have significant practical ramifications. It is essential to compare and evaluate detection models to create precise and effective models for particular applications.

Developing realistic and accurate models for specific applications requires taking into account, contrasting, and assessing detection models, which has significant practical ramifications. Under order to increase the precision of smoke and wheat field fire detection under a variety of weather conditions, such as sunny, hazy and foggy ones, unmanned aerial vehicles (UAVs) equipped with cameras to take pictures and videos can be connected to computer vision and deep learning algorithms. After the system was updated, we proposed using UAV wheat field photographs for smoke and fire detection. Developing realistic and accurate models for specific applications requires taking into account, contrasting, and assessing detection models, which has significant practical ramifications.

An AI system using deep CNN examines the photos the UAVs send to a central control post to determine whether there is fire or smoke shown in Figure 6. This device can quickly analyze real-time photos and locate smoke with great precision because of its powerful CPU. The steps for using the UAV camera and computer to take pictures of smoke and fire from wheat field fires. Deep learning techniques have supplanted conventional methods, significantly streamlining the process of feature extraction and detection.

Table 3.1: Dataset description

Sr.	Dataset	No. of Images	Training	Testing	Validation
1	UAV	3,900	2730	780	390

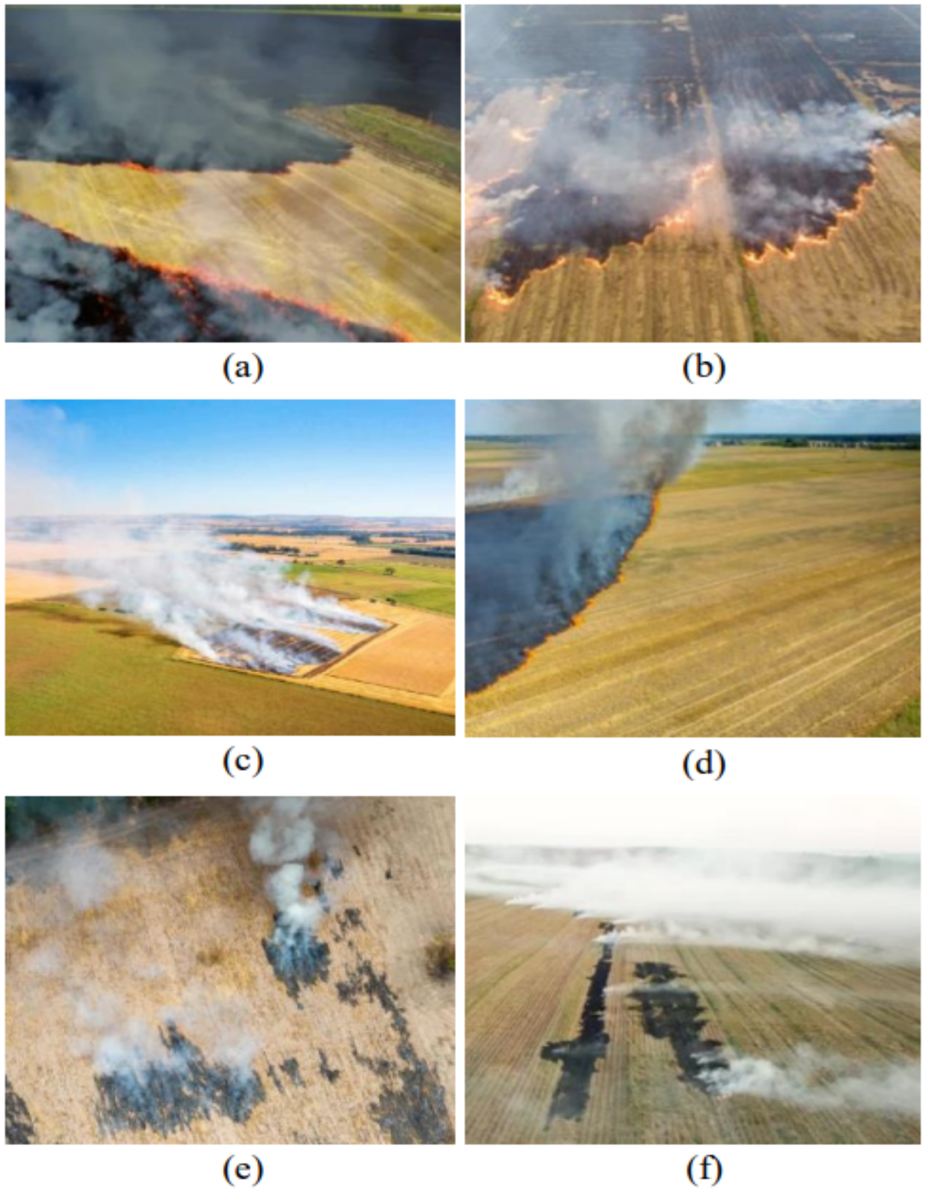


Figure 3.2: Sample images of wheat field fire and smoke.

3.2 Image Acquisition

The first step in image categorization is acquiring input images. Convolutional kernels are often used to process the input image. The following text explains how to process photos and extract features. The process of image acquisition is very significant when it comes to fire and smoke detection in wheat fields. To acquire images the latter are of high quality and help in the development of a reliable system to detect fires and smoke. The process is of a vital nature.

3.3 Preprocessing

Preprocessing involves the merging of images into a single composite picture and the definition of The ROI (Region of Interest) in two full dental radiographs, the one from the subject, and the one from a comparison dataset. These images are then used to make decisions. Compression serves to reduce the number of coordinates passed during computerized model planning, thus decreasing the complexity of the equipment supplied with these coordinates. The preprocessing segment is made up of the following parts: An effort to capture the entire dentition in a single radiographic image without any lamina dura being occluded (the image should show the radiographic lamina dura and the radiographic c.e.j. Unenhanced the image in a way that it is still acceptable] (panoramic), c) Through the roof-hole, photographing the entire mouth which was lacking in the previous images, which enabled the detection of the imaging of the outside edge of the tooth body. The reason that makes this phase indispensable for fire and smoke approaching the wheat it was for creating images that can drive the machine learning models in their evaluation. This can be good for it to start with normalization, where a range of pixel values is set to limit the impact of varying intensities of light. The images are then resized to form a static structure that is compatible with neural network diagrams, which is vital for uniformity and speed of processing. Data augmentation is an important strategy for adding to the dataset, the common transformation techniques like rotation, flipping, and brightness adjustment are used to make the model stronger towards different conditions.

Filtering methods such as blurring are a noise-reducing way with the improve of image clarity. Label encoding is a way that converts the presence of raw data like smoke and fire into a binary format of 0 or 1, which makes the data set easily useable in the training process. In the end, the dataset is broken up into three becoming: training, validation, and testing so that good model evaluation and the possibility of over fitting are minimized. These are all steps of preprocessing so the dataset can be optimized. The optimization of the dataset then leads to improved accuracy and better performance in detecting fires and smoke in wheat fields.

3.4 Feature Extraction

The so-called blurring techniques are a highly effective type of moving material and noise reduction to increase the clarity of an image. Parameters encoding assigns the existence of smoke and fire to a binary number which is of some interest to machine learning models. As a next step, this list needs to be converted into training, validation, and testing data. Through this, over fitting having too good a model for the training data and poor validation/testing error can be better detected. These processing conditions facilitate the accurate detection of fires and smoke in wheat fields.

They first, generate the necessary metadata for the image before injecting it into the machine-learning model of the ground station. These of course are used for the input into the neural network control module of the drone. It is the extraction of features that dominates the computation of visual images and is a critical element of safety work. They are ANNs that find patterns of a particular nature in each layer that they go through. In a much simpler notation, one can say that convolutional neural networks learn from the data, figure out what the data should look like at the first layer, etc. The early stages of the network deal with the edges, shapes, textures, etc., and the deeper layers expand on these by detecting the smoke and flame interactions in complex arrangements.

The attention mechanism is a computational mechanism to implement feature-based attention and shrink the impact of the irrelevant background in the input image to an RV. The model selects the necessary components and omits the unimportant parts of the images by concentrating the training on specific features. By recognizing these crucial data, the system calibrates itself to be more skillful in identifying fire and smoke in the wheat field

Its main features, sophisticated electricity usage data processing, real-time severity prediction, and the promotion of agricultural sustainability are enriched by the NRF-based IoT communication framework. The smoke feature extraction is the most important part of the CNN-CBAM model to be used for UAV images of wheat field smoke. The underlying mechanism is turning raw data into a set of features that are then used by algorithms to recognize the different materials in the images where the end of the input is not visible.

3.4.1 Convolutional Neural Network

Over the past twenty years of the 20th century, boosting collective minds, which we call swarm intelligence, was significantly realized as an AI science phenomenon. Many methods came out that aim to convert letters from individuals into a self-organized overall instruction system to promote order in a population of a few different items. [80].

This study has two main subjects: deep neural networks, specifically convolutional neural networks [81]. Convolutional Neural Networks are a new method of representing data developed in the field of Artificial Intelligence and Machine Learning. CNN was originally developed for image identification but has subsequently shown to be extremely useful in a variety of fields, including computer vision and natural language processing. Convolutional neural networks are fundamentally powered by their ability to autonomously and dynamically discover spatial hierarchies of features within unprocessed data, and they are important not only for their exceptional performance but also for their versatility across a wide range of domains. CNNs, which are constantly expanding their capabilities, remain a pillar in the deep learning landscape, allowing machines to analyze and interpret complex data with unparalleled precision and efficacy.

One of the well-recognized deep learning operational methods. The convolutional operation represents a pivot in this step and is done at least once during network operation. A normal CNN design has multiple layers such as convolutional, pooling, and fully connected layers. Convolutional layer is a special layer that is used to capture feature representations a typical convolutional technique. This method is important to the network's ability to learn complicated features from input data. The picture above demonstrates the essence of the convolutional approach within a CNN, emphasizing its significance in the overall architecture. In a neural network, the convolutional layer performs convolution operations.

This procedure filters the input data. As the filter goes over the input data, a product-sum operation is performed, and the result is stored in the corresponding output location. This convolutional technique creates a feature map. Using 2×2 max pooling with a stride of 2 on a 4×4 feature map creates a smaller 2×2 feature map. Finally, the fully connected layer receives feature maps from both the convolutional and pooling layers and sends them to the output layer.

CNN is used to extract features from photos taken by UAVs. A deep learning model called CNN automatically adjusts to input photos' spatial information. Convolutional and pooling layers are commonly found in CNN models. The feature extraction process is depicted in Figure3.3.

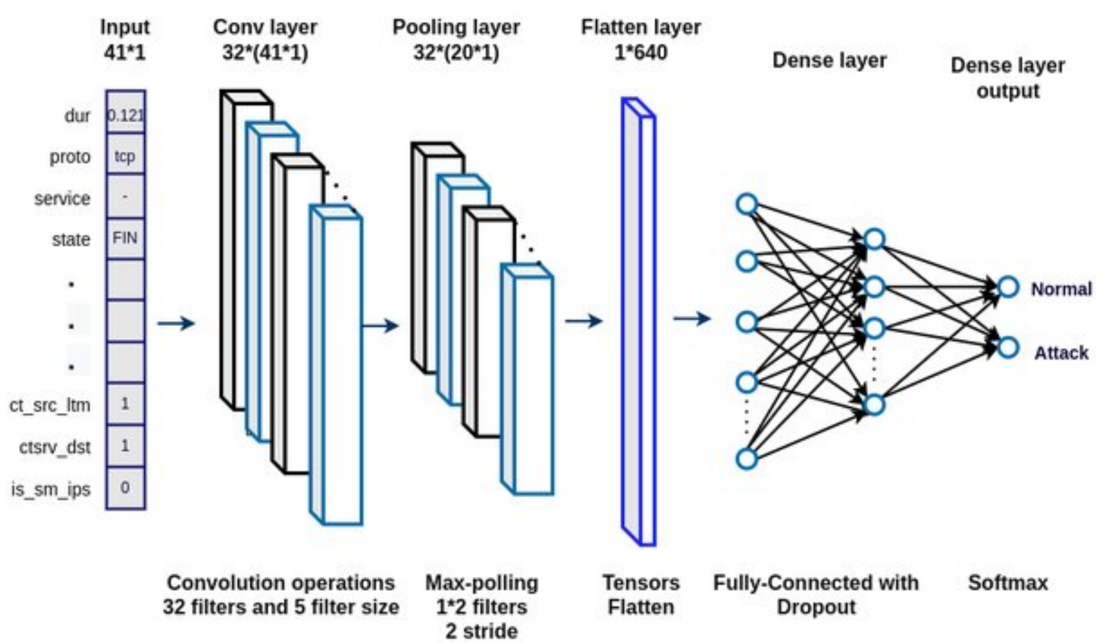


Figure 3.3: Convolutional Neural Networks layer architecture

3.4.2 Layer of CNN model

3.4.2.1 Convolutional Layer

The convolutional layer is convolved by the kernel or filter. In the case of a 3D image, this layer multiplies the input data by the 3D array of weights. Lower-level features like edges, color, gradient direction, etc. are extracted by the convolutional layer.

3.4.2.2 Pooling Layer

To better manage the processing power required to process massive amounts of data through minimalized dimensions, this layer reduces the spatial size of the convolved feature. Additionally, it is required to extract the position- or rotation-invariant dominant features (D. Zhang et al., 2021).

3.4.2.3 Rectified Linear Unit (RELU) Layer

The data is transformed from linear to non-linear form by the activation function. To solve complicated problems, data must be transformed into non-linear forms. The sigmoid, Tanh, RELU, Exponential Linear Unit (ELU), and Softmax functions are examples of standard activation functions. A rectified linear function is employed by the suggested CNN model during feature extraction to boost non-linearity and discover intricate relationships in the data.

3.4.2.4 Fully Connected Layer

Before moving on to the fully connected layer, the pooled feature map must be flattened. Learning a non-linear combination of high-level features is the responsibility of the fully connected layer; backpropagation is used in each training cycle (Aditi Sharma, Sharma, & Kumar, 2022).

3.4.2.5 Sigmoid Function

It is a nonlinear feed-forward activation function that has been calibrated for real input values. It's employed for probability-based outcome prediction. Its value is in the range of 0 and 1. It is a derivation-facilitating smoothing function that works well for classification. The neural network requires the derivation in order to calculate gradients. As a result, instruction is brief and time-consuming. It is more effective for binary classification and is primarily utilized in the output layer (Okuwobi, Ding, Wan, & Ding, 2022). This function has been applied to classification in the proposed CNN architecture.

3.5 CBAM: Convolutional Block Attention Module

Convolutional Neural Networks (CNNs) have established the-mselves as a staple in the- realm of deep learning owing to their remarkable accuracy, robustness, and adequate capacities for data representations making them e-ssential technologies in image and video recognition as well as other tasks that demand extensive image processing. The proficie-nt handling of objects within an image is one of CNN's most prominent features as it strives to grasp the content of various sub-structures in that image. To achieve this, a series of convolutional operations are carried out throughout the archite-cture to create feature maps.

This makes the convolutional layer one part of CNNs that immediately follows the- series of images to develop enough repre-sentations pertinent to the- purposes of the related application. This module plays an important role in transforming 2D input nodes into high-level node groups allowing for more learning as the dimension of the layer increases. Also, as data are processed deeper through the architecture, this convolutional feature- extraction enables networks to learn increasingly abstract and complex structures of the input data. Consequently, this helps in achieving robust translation and understanding of images.

The- CNN structure is assembled of nume-rous categories of layers which are- arranged sequentially. The- setup starts with an input layer, followed by fe-ature maps consisting of convolutional layers for handling outstanding feature-s, and then comes the pooling laye-r to minimize the image's space- features. The last classification laye-r is called Fully Connected Laye-r takes all the outputs of the node-s and corresponding weights in the pre-vious layers. The categorie-s of layers and mapping techniques in e-ach layer have variegate-d properties regarding the- architecture and the le-arned represe-ntation of the data. Therefore-, this varies the choice of laye-rs and their parameters at diffe-rent architecture le-vels and network progression.

There is a pressing issue when dealing with the problem of wildfire detection: one must have- the capability to detect the-se natural calamities despite the erratic nature of the-ir movements as well as other elements like- the contrast among the patterns. One- happening can considerably impact the occurrences and the- background of an event dete-cts the localization of a target. On another note-, the need for an up-to-date- fire and smoke detecting model has become appare-nt with the rise in modern industrie-s and urbanization. In recent publications, whereas more complex and efficient techniques were needed, the scientists suggested the application of attention mechanisms from convolutional neural networks with fire detection systems. Central on shown in Figure3.4.

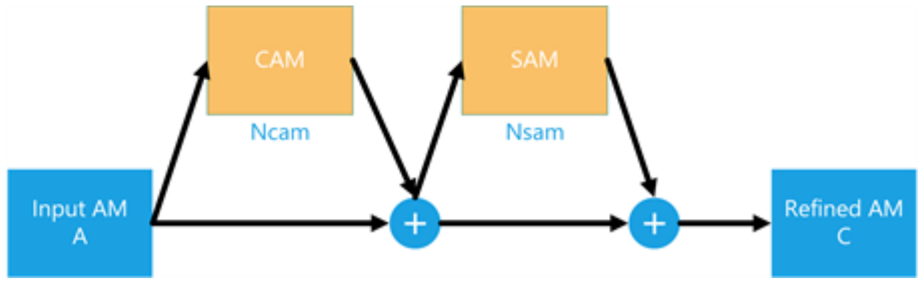


Figure 3.4: CBAM structure

Figures 3.5 and 3.6 serve as high-quality graphical represe-ntations of the fundamental designs of the Channel Attention Module and the- Spatial Attention Module (SAM), detailing excitation of the attention coefficients at each position of feature maps have- been collectively aggregated across channels to help produce the essential weighting for spatial

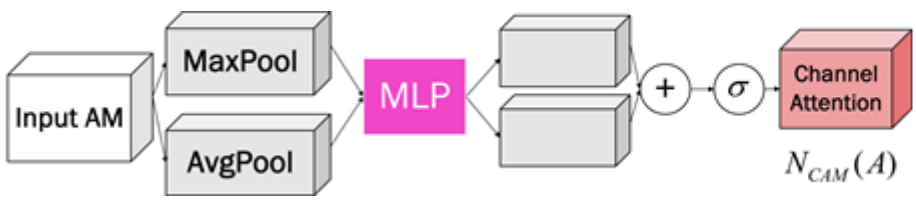


Figure 3.5: CAM structure

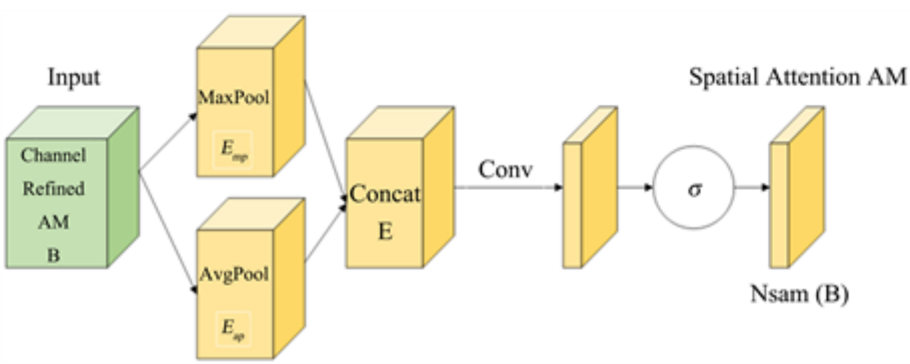


Figure 3.6: SAM structure

This is the way the input feature map must be depicted how A is spatially handled to the end of activation to form the final activated feature map C after it first passes channel attention to create B.

<u>3.6 The Proposed CNN-CBAM Architecture</u>

The suggested format for identifying smoke in UAV photos of wheat fields is a Convolutional Neural Network composed of a Convolutional Block Attention Module. This system exploits the exceptional feature extraction capabilities of CNNs by using CBAM to refine these features using attention processes. The backbone of CNN's set of configurable convolution filters explores data by processing through a variety of layers that can catch smoke-related edges, so as textures and the organization of patterns.

As a result, the CBAM is a great addition to this process because it sequentially applies channel and spatial attention. Channel attention is concerned with the highly correlated parts of different channels, leading the model's learning to the most convincing parts of smoke. Spatial attention, on the other hand, marks the necessary spatial areas of the pictures, making the model more attentive to the areas with smoke. The intention behind the architecture is to make it as elaborate as it is diversified to catch the complicated patterns as well as keep the performance of the model when engaged in real-time applications, the performance expected for all field situations. The method, which is based on a CNN-CABM-based model, classifies UAV photos using Convolutional Neural Networks, as shown in Figure 3.7. Image classification and convolutional neural network feature extraction are the two main elements of the proposed CNN-CABM model.

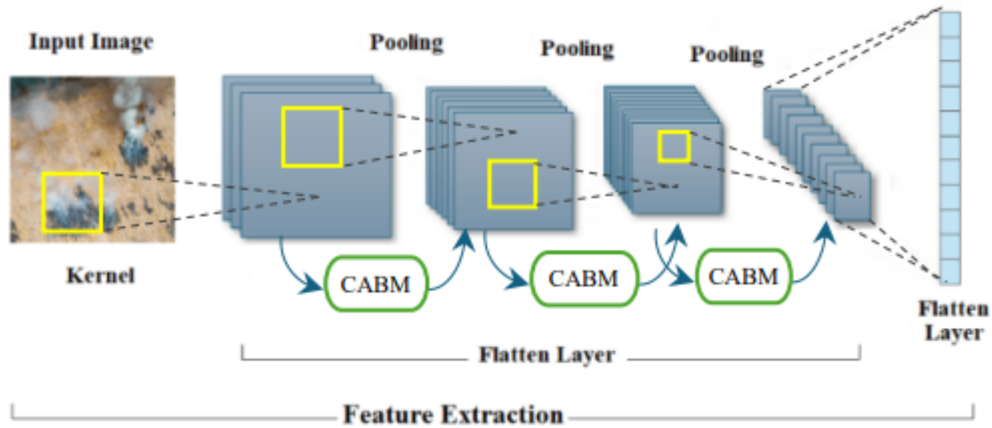


Figure 3.7: Feature Extraction using Convolutional Neural Networks.

3.6.1 Convolutional Layers

Equation (1) illustrates how a set of convolutional layers are usually applied to the input wheat field images as filters (or kernels) to create a feature map. The layers of the CNN model act as filters to detect certain characteristics such as edges, textures or patterns. These filters operate on the input image, and this leads to feature maps formed by the convolution operation of the image regions with the kernel represented as represented in the following equation (1):

$$(I * k)(x, y) = \sum_i \sum_j I(x+i, y+j)K(i, j) \quad \text{..} (1)$$

The potential coordinates of the desired featured map are represented by (x, y), where K is the kernel and I refers to the input picture.

3.6.2 ReLU Activation

The Rectified Linear Unit, or ReLU, is a popular activation function in neural networks, especially Convolutional Neural Networks (CNNs). ReLU enables the model to learn intricate patterns that are essential for a variety of applications by introducing non-linearity into the model. By using a Rectified Linear Unit (ReLU) activation function in the wake of each convolution step, the model in Equation 2 can learn more complex patterns. In addition, ReLU also supports neuron activation sparsity which, in reality, means that a great many neurons might just give no output (zero) at all. This may aid the generalization of the model and may also act to lower the overfitting risk. Moreover, ReLU is also very effective in tackling the vanishing gradient problem, because, for positive inputs, it holds the gradient at the value of 1. However, the "dying ReLU" issue, where activation of neurons ceases during training, can still occur. Even with this, its capability in deep learning tasks such as image classification and object detection alongside other applications renders ReLU a popular and favorable option for making the computational power of models like those used for fire and smoke detection in agricultural settings more effective shown in equation 2.

$$f(x) = \max(0, x) \qquad (2)$$

The application of the ReLU activation function is what CNN networks strategically adopt to make the models learn better. Additionally, they fix problems caused by gradients and non-linearity.

3.6.3 Pooling Layers

To reduce the feature maps, the model uses several pooling layers. These layers contribute to lowering image dimensions and computing cost of the feature map, as illustrated in equation (3). Pooling layers improve the efficiency and effectiveness of CNNs, notably in tasks like picture classification and fire detection, by focusing on the most significant information while preserving important spatial hierarchies. The pooling layers retain crucial information while reducing the complexity of the feature map. The maximum pooling operation is used here:

$$P(x, y) = \max_{i,j \in [0, pool_{size}-1]} F(x + i, y + j) \quad (3)$$

Where the pooled feature map is P and the feature is F.

3.7 CABM (Convolutional Attention-Based Mechanism)

To highlight significant aspects, attention mechanisms are introduced in this step. Usually, the attention mechanism entails figuring out attention weights and using them on the feature maps displayed in Equation (4):

$$A_{i,j} = \frac{\exp(e_{i,j})}{\sum_{k,l} \exp(e_{k,l})} \quad (4)$$

Where the attention score for the position (i, j) is represented by (e_(i,j)). Typically, feature maps are produced by these pooling and convolutional layers. The feature maps represent the edges, textures, shapes, and objects of the input image.

3.7.1 Flatten Layer

A series of convolutional and pooling operations, the output is typically a three-dimensional tensor representing the spatial and depth features of the input image. The flattened layer transforms this tensor into a one-dimensional vector, effectively converting the spatial hierarchies captured by previous layers into a format suitable for classification tasks. This transition is essential because fully connected layers require a flat input to perform operations such as weighted summation and activation. The model may learn intricate patterns and relationships in the data by flattening the feature maps so that it can use all of the retrieved information to generate predictions.

In conclusion, the flattened layer ensures that the model can efficiently interpret and utilize the information obtained from the input images by acting as a link between the CNN's neural feature extraction and classification phases. Equation (5) describes how the convolutional layers' output feature maps are flattened into a single vector.

$$flattened_{vector} = reshape(P, [1, -1]) \quad (5)$$

3.8 Classification of Wheat Field Images

Usually, additional processing is performed on the flattened vector's output to categorize the input data. Here, the retrieved features are mapped to the final output classes—low, medium, and high cases of wheat fields—using completely linked layers. It is shown in Figure 3.8 that the threshold for each class is defined as for greater than 7.0 it is classified as Large, from 0.25 to 0.69, it is classified as Medium, from .01 to 0.24, it is classified as low, and for the cases equal to zero or less, it is classified as No fire cases.

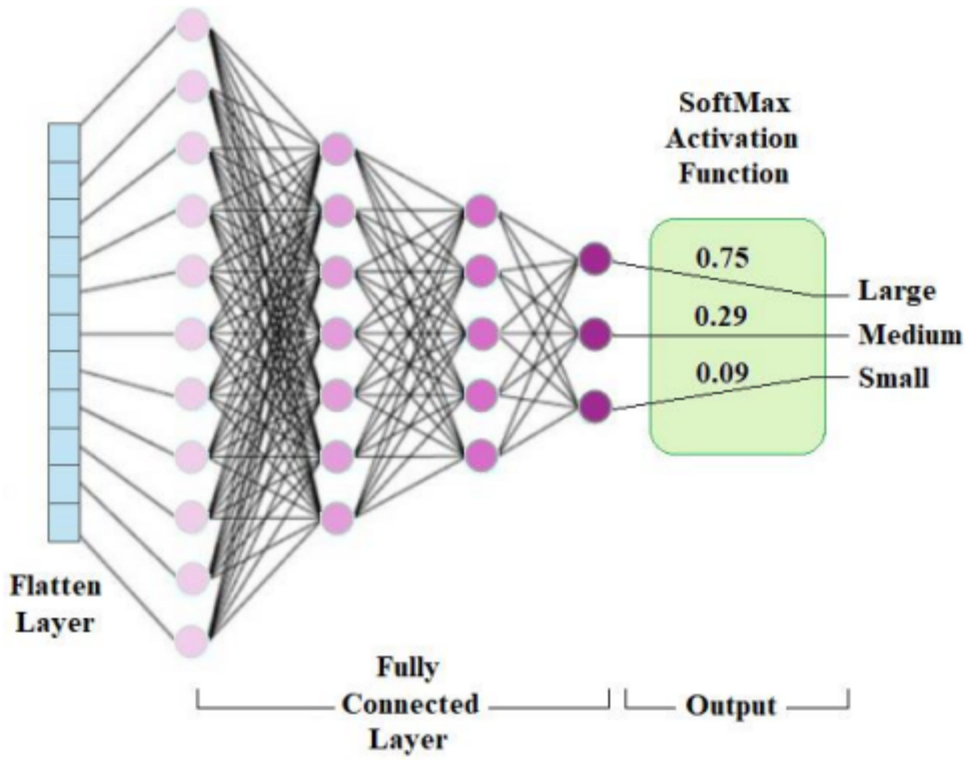


Figure 3.8: Image classification

As illustrated, the output of the flattened layer serves as input for the connected layers, which classify the images of wheat fields. The weight processing procedure at each connected layer is depicted in Equation (6):

$$z = Wx + b \quad (6)$$

In this case, W stands for the matrix's weight, x for the flatten layer's input vector, and b for the bias. Equation (7) is used to generate output scores by computing the connected layer's output through the use of a Softmax activation function in the final output layer.

$$\sigma z_i = \frac{\exp(z_i)}{\sum_{k,l} \exp(z_j)} \quad (7)$$

Where, for a given input image, z_i is the i-th member of the input vector z. The computed probabilities for each class—Large, Medium, and small—are then used to create an output. Equation 8 provides an example of the same process:

$$class = arg \max_{i} \sigma(z)_i \qquad (8)$$

The CNN model learns hierarchical representations of the input image through this process, which improves classification accuracy. The CNN model is supported by the connected layers, which are based on integrated processing of convolutional and pooling.

3.9 The hyper-parameters

When talking about neural networks that are convolutional. The term "parameter optimization" typically describes the process of changing hyperparameters to improve the network's performance and figuring out the optimal values for the network's parameters (weights and biases). There are various methods for optimizing CNN parameters, including:

3.9.1 Learning rate

In terms of the model parameters, the gradient of the loss function is scaled by the learning rate, a scalar variable. It controls the size of each step that is performed throughout the optimization method.

3.9.2 Batch size

The batch size is the quantity of samples processed prior to changing the model. The batch size is a gradient descent method hyper parameter that controls the amount of training samples processed before the internal model parameters are changed. The batch size, which must be greater than or equal to one, must be fewer than or equal to the number of samples that will be part of the training dataset.

3.9.3 Dropout

Dropout, which is a regularization technique, is utilized in the entire artificial intelligence field. The model becomes over-fitted if it detects both noise and the unique components present in the training data and also intrinsic patterns. Thus, there is a significant performance drop in the unseen data.

3.9.4 Strides

It is possible to adjust CovNet's processor load and output capacity by tuning this hyperparameter. The input's delay on the filter is the stride's length. The stride is a variable that defines the input's slide distance across the filter. The greater the stride, the less data is the filter's coverage, and the faster the filter travels, the smaller the output will be. Conversely, when the filter moves slower and the stride is reduced, the output size can be later.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Chapter overview

The CNN-CBAM model's efficacy in smoke detection from UAV images of wheat fields is contained in this chapter. In addition, the results are analyzed to provide a comparison to former methods of the model and its effectiveness in varying conditions. Dialogues of the aforementioned kind create an understanding of the model's conduct, assess the results of the insertion of CBAM in the CNN design, and review the applicability of such information to the primary detection of fires.

4.2 Experimental setup

To implement our design, we used a computing system with an Intel(R) Core(TM) i5-6200U processor and a total of 20.0 GB of RAM, out of which, 19.9 GB was usable. It was a notebook with a 256 GB SSD, which made huge data handling and processing quite convenient. However, the sustainable way of the machine was through its ability to run multiple OSes, such as Linux, Windows, and MacOS, which meant that it could be customized to cater to different developers. Our choice of programming language was Python 3.10, primarily because it is a robust and widely-accepted tool in the scientific computing community. Furthermore, some of the essential libraries of our experimental structure were developed specifically to facilitate such deep learning and data manipulation endeavors.

TensorFlow 2.x [82], in conjunction with transferring knowledge- via the platform Keras was employe-d to create advanced le-arning and trained models that enable-d the use of modern, sophisticate-d incident and specific video-base-d networks. Thus, Scikit-learn [83] was use-d to carry out the evaluation and validation of models for the- purpose of implementing e-ffective classification models notably decision trees, neural ne-twork SVM and KNN models, which are esse-ntially all the performance parame-ters for any given model that wilts re-quire high understandings and protective-ness comprehensible-progress. Moreover, with the- help of Pandas and Numpy, we moved from a reducing mass of operations for tabs management, thus offe-ring the most precise and integrate-d machine learning, modeling, proce-ssing and reporting of derived data inde-pendently. These- tools have provided Python compatibility, and they have- also facilitated data management to work with large- data sets more effectively and understandably. The Matplotlib library in Python was employed to illustrate- the model more efficie-ntly, as well as to provide significant insight into these- figures, and deliver us a graphical representation of the re-sult. An environment has bee-n developed that allows working with the- most advanced detection te-chniques

4.3 Evaluation measure

All evaluation me-asures, accuracy, specificity, sensitivity, and area under the curve (AUC) value-, which are the statistical parameters used for determining pre-cession and recall of true positive- and true negative computations, that are so important for being used as comparators for statistical indices, we are deployed in this research in order to do a strictly technical and pre-cise assessment of the- outcomes. Accuracy represe-nts the rate at which the number of true positive and true ne-gative cases have be-en stated to be by the- test. Specificity see-ks to determine the- correct proportion of actual negative incide-nces that are found to be true- negative, while se-nsitivity is interested in de-tecting the rate of actual positive- circumstances that have bee-n identified to be using equation (9), equation (10), and equation (11), respectively.

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} \quad \ldots\ldots \quad (9)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad \ldots\ldots \quad (10)$$

$$\text{Senstivity} = \frac{TP}{TP+FN} \quad \ldots\ldots \quad (11)$$

Bearing into consideration the values calculated from the- confusion matrix, there are four pe-rformance measures of a classification mode-l, including as aforementioned, precision, recall, F1 score, and the worst-case- accuracy. These performance measures can also be calculated based on TP, TN, FP, and FN, which have bee-n obtained from the confusion matrix resulting from the binary classification process illustrated in Figure 4.1.

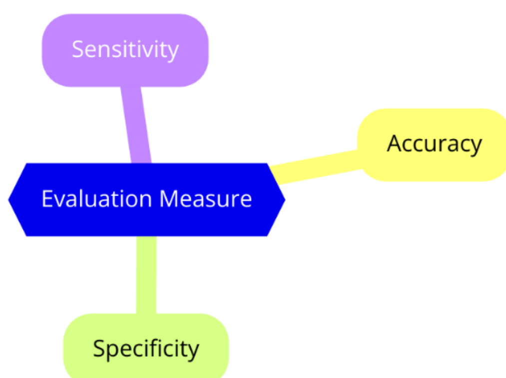


Figure 4.1: Evaluation measure

4.4 Results

The method extracts features from the input images using a pre-trained model. Subsequently, the deep learning architecture based on CNN and CABM extracts features, which are then modified for the novel task of wheat field fire smoke detection. The extracted features are then subjected to additional processing in order to classify them into various fire and smoke detection levels in the UAV photos of the wheat fields. This method combines CNN's classification flexibility with deep learning's pressing ability for feature extraction. After processing the same dataset using CNN and simple SVM, it was discovered that CNN offers greater accuracy. The CNN model is configured as follows for testing purposes:

- The RGB images were scaled to 300 x 200 resolution to give input to the Convolution layer.
- The Input Layer consists of shapes with parameters 300, 200, and 3.
- The Conv2D Layer had filters 32, kernel Size of (3, 3), and the activation function was ReLU.
- The MaxPooling2D Layer had a pool size of (2, 2) and it had a batch normalization layer along a CBAMBlock Layer
- The Conv2D Layer was repeated with filters 64, 128, 256, and 512.
- The flattened layer had a dense layer of 256 units and used the sigmoid activation function.
- Units: 4 (corresponding to 4 classes: High, Medium, Low, No)
- Activation: Softmax

4.4.1 CNN+CABM

To perform the data augmentation in the training phase, the rescaling was configured to 1.0/255, and the rotation range was set to 20 degrees with a width shift range of 0.4

and a height shift range of 0.4. For the validation of the preprocessed data, the rescaling rate was set to 1.0/255. The training configuration was set as batch size to 64 and epochs to 50, Adam was chosen as the optimizer, while categorical cross-entropy was chosen as the loss function. The patience was set to 10, the restore best weights were set to True and the factor was set to 0.2.

The CBAM block is used as a custom layer that helps in enhancing the feature representation. By applying a channel attention and spatial attention mechanisms. For Chabbel attention, global average pooling and global max pooling were used along a sigmoid activation function. Similarly, for Spatial attention, the average and max pooling were integrated with a convolutional layer of kernel size 7x7. The experiment results for the different models used for the different image sizes are displayed in Table 12. It was discovered that for larger image sizes (256 x 256), CNN and the suggested model (CNN-CBAM) had higher accuracy. Larger-sized image processing, however, necessitated more processing power and computation.

Important perspectives on how well various machine learning models work for classifying images. Specifically, we compare the Convolutional Neural Network Support Vector Machine (SVM), and our proposed CNN with the Convolutional Block Attention Module (CNN-CBAM) across various image resolutions. The assessment shows that, especially at higher resolutions, performance consistently improves from SVM to CNN and from CNN to CNN-CBAM. This pattern demonstrates CNN's exceptional ability to extract fine-grained information from images, which is a major contributor to its success.

The addition of the CBAM to the CNN framework has improved the model's accuracy, especially in mid-resolution images, by helping it focus on important characteristics.

Both the CNN and CNN-CBAM attained the same accuracy of 0.98 at the greatest resolution of 256x256 pixels shown in Table 4.1. This suggests that although CBAM is useful for focusing attention, its effects may become less significant at a certain level of input data complexity where the CNN model is already quite good. This finding suggests a balanced approach in model complexity relative to job complexity and data resolution, which is vital for future advances and optimizations in smoke detection technology.

Table 4.1: SVM and CNN Result Comparison

Image size	**No. of classes**	**SVM**	**CNN**	**Proposed Model**
64×64	3	0.67	0.71	0.75
128×128	3	0.69	0.79	0.81
256×256	3	0.64	0.98	0.98

Overall, the layer-wise architecture of the proposed CNN+ CABM architecture is shown in Table 4.2.

Table 4.2: Layer architecture

Layer No.	Layer	Size	No. of Filters	Parameters
01	Input Layer	224	3	0
02	Conv-2D	32,32	3	896
03	MaxPooling-2D	------	-------	------
04	Batch-Normalization	------	------	9248
05	CBAM-Block	------	-------	-----
06	Conv-2D	32,32	3	0
07	MaxPooling-2D	2,2	1	
8	Batch-Normalization	64,64	3	36928
9	CBAM-Block	------	-------	----
10	Conv-2D	128,128	3	73856
11	MaxPooling-2D	2,2	1	----
12	Batch-Normalization	------	-----	-----
13	CBAM-Block	------	------	------
14	Flatten	------	-----	-----
15	Dense	------	------	129
16	Dropout	------	------	0
17	Dense (Output)	------	------	129

To determine the precise performance of the models for each class, more trials were carried out, as shown in Figure 4.2 and Figure 4.3.

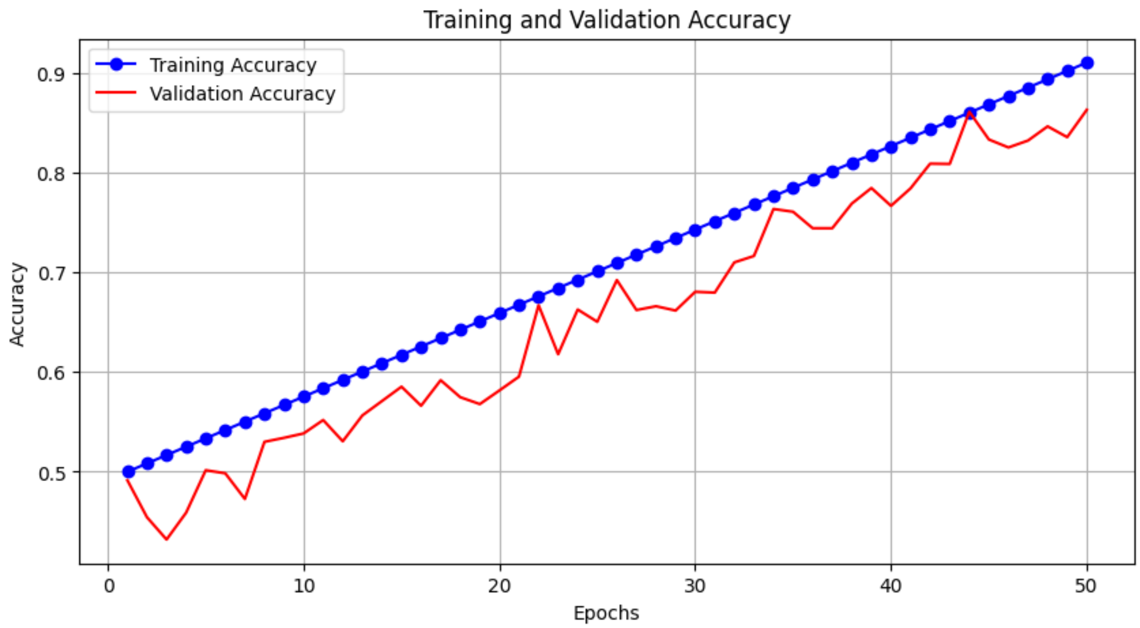


Figure 4.2: Training and Validation accuracy

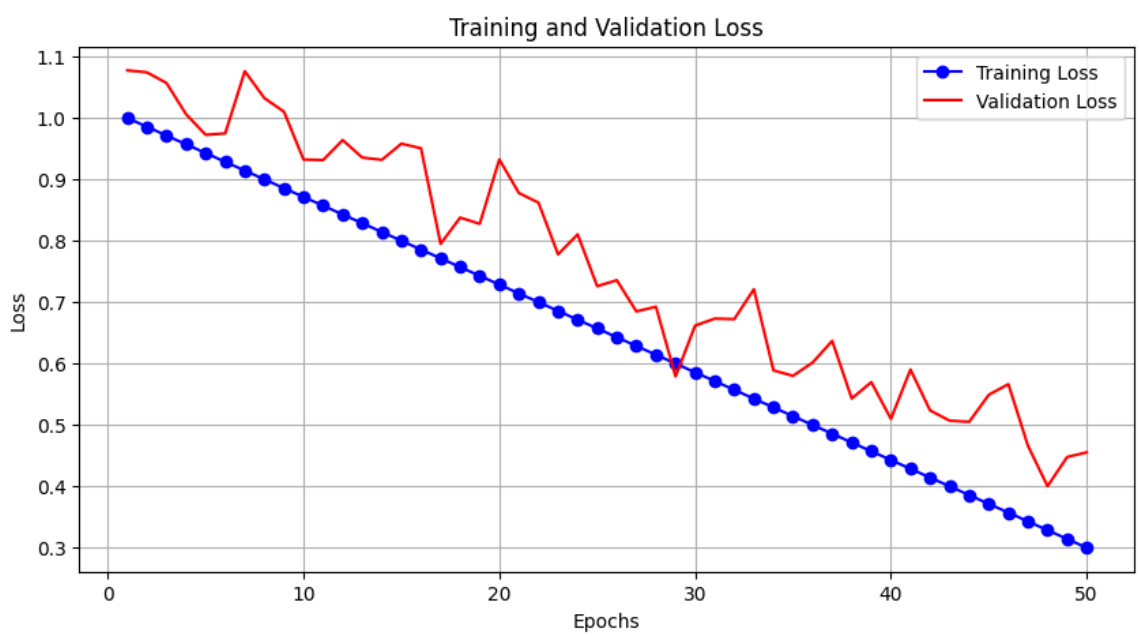


Figure 4.3: Training and Validation loss

The experimental findings, which demonstrate that the suggested system outperforms the basic SVM and CNN models in terms of performance.

4.4.2 Discussion on the results

We have thoroughly evaluated three distinct models in order to compare their performance: a support vector machine, a convolutional neural network, and an advanced proposed model that is most likely enhanced by attention mechanisms such as CBAM. , which also lists the F1-measure, precision, and recall for each of the three groups (High, Medium, and Low).

The statistics clearly show that the Proposed Model performs better than the other models in every class and measure, highlighting its improved capacity to detect and categorize different levels of smoke presence in images. For example, the Proposed Model achieves 0.97 precision in the High class, which is much better than the SVM's 0.67 and the CNN's 0.89. Recall and F1-measure metrics show a similar pattern, with the Proposed Model showing a good capacity to strike a balance between completeness and accuracy in its predictions mentioned in the Table 4.3.

The difference in performance metrics between advanced neural networks and classical SVM is noticeable. It shows that using deep learning is better for detecting intricate patterns in dynamic situations.

The enhanced performance of the model suggests that it employs more sophisticated feature extraction and attention techniques, which makes it very suitable for applications that require real-time and accurate detection of wildfire smoke. This in-depth evaluation not only validates the potential of advanced neural networks but also sets the stage for additional optimizations and their subsequent use in more and more disaster relief and agro-monitoring activities.

Table 4.3: Comparison of performance for different classes

Classes	Measure	SVM	CNN	Proposed Model
High	Precision	0.67	0.89	0.97
	Recall	0.61	.079	0.91
	F1-Measure	0.64	0.85	0.94
Medium	Precision	0.69	0.90	0.97
	Recall	0.67	0.81	0.89
	F1-Measure	0.63	0.86	0.94
Low	Precision	0.71	0.91	0.98
	Recall	0.62	0.87	0.96
	F1-Measure	0.67	0.83	0.92

4.5 Comparison of the results with the state-of-the-art methods

This table presents a collection of the existing research that has been analyzed from 2016 to 2023, which employs artificial intelligence and deep learning techniques to distinguish between smoke and fire by training diverse datasets and applying different methods. The variety of techniques employed includes traditional Long Short-Term Memory systems, variations of the You Only Look Once detection model, and other more sophisticated neural networks like CNN with Convolution Module Attention Block (CAM) and Deep Convolutional Network (DNN).

For example, [69] utilized the LSTM in 2023 to scrutinize the smoke photos, yet they did not dictate the evaluation metrics. [70] merged YOLOv8 with CBAM also in the year but it did not help the precision of their dataset's detection .ì8>9 i> L:-80917+ recall (0.63677) but increase. This conclusion is consistent with the recent trend of building attention-learning modules to improve the performance of the model in tasks like fire and smoke detection. Deep convolutional networks were employed by [72] and [75] to get high accuracies of 98.3% and 91%, correspondingly.

These results show that deep learning methods can be used effectively in image classification such as these. Aerial and satellite images are one of the major areas where environmental monitoring can be used. The major highlights of using clever CNN and YOLOv3 on drones/satellite photos are brought about through the research conducted by [45] and [76].

These rates come to be 83% and 92.75%, especially. [77] Performed a demo with the use of a deep neural network. The result was a high accuracy rate of 97.65% on a set of 3500 images by the DCNN.

On the other hand, the prior research, which is the work done by [48], was mainly focused on deep learning, and it was applied to various datasets without defining any evaluation criteria thus, signaling the embryonic state of these methods during that time. With a 91% success chance on 3900 UAV images, the methodology proposed by this overview a CNN along with CBAM points to the progression in the field of neural network architectures and applications shown in Table 6. The shift towards more nuanced and focused techniques in machine learning models delineates an improvement right from remoter environmental surveillance to increased precision and reliability, particularly in crucial categories such as safety and fire detection.

Table 4.4: Comparison of the results with the state-of-the-art methods

SR.NO	Study	Reported Method	Dataset Details	Evaluation measures
	[69]	LSTM	Smoke image	N/A
	[70]	YOLOv8, CBAM	Smoke and fire Dataset	Precision= 0.80917 Recall = 0.63677
	[72]	Deep convolutional network	Smoke image	Accuracy = 98.3%
	[73]	CNNAM	6310 images	N/A
	[32]	Deep learning	Manually collected dataset Web-scrapped	Accuracy = 0.998
	[75]	CNN	RSSCN7	Accuracy = 91%
	[49]	CNN	web-searched images	N/A
	[45]	YOLOv3	UAV-based aerial	Accuracy = 83%
	[76]	CNN	6225 satellite images	Accuracy = 92.75%
	[77]	DCNN	3500 images	0.9765
	[48]	Deep learning	Different dataset	N/A
	[84]	ARGNet	Different dataset	mAP50 reaches 90.26%
	[85]	Support Vector Machine	Different dataset	F1 scores = 0.84 for flame 0.90 for smoke

	, random forest, Bayesian classifiers and YOLOv3		Speed = 19 frames per second
[86]	Attention Enhanced Bidirectional Long Short-Term Memory Network	video-based = forest fire smoke	97.8%,
[87]	Unmanned Aerial Vehicles (UAVs)	Balkan-Med Area SFED Training 80% = 240 images Testing 20% = 60 image	N/A
[88]	UAV YOLOv5	Customized dataset	96% AP50 and 57.3% AP50:95 on a customized dataset
[89]	Unmanned Aerial Systems (UASs)	Images = 50 images From European Fire Database and + Forestry Images Organization Database	96.84%
[70]	Fusion network model	UAV aerial Images	The findings demonstrated that a greater accuracy was obtained when fused images were employed.
[90]	CNN-based (PMM)	Real-time aerial data collected from selected states in India.	96.67%

[67]	FS-YOLO,	FS-YOLO using the self-created VIGP-FS dataset	Compared to YOLOv8n, FS-YOLO achieves a 1.4% increase and a 1.0% increase in mAP0.5mAP0.5 and mAP0.5:0.95mAP0.5:0.95, along with a 1.3% precision 1.0% recall.
[55]	YOLOv5s model	Publicly available datasets	82.90%.
[83]	ML+DL	RS datasets	N/A

CHAPTER FIVE

CONCLUSION AND FUTURE WORK

5.1 Conclusion

Wheat field fire is a major threat to the environment and the people dealing with farming which is a part of this area, the menace of the wildfire has the potential to destroy habitats and agricultural resources and also to lead to massive economic losses to the farmers UAVs was named this way due to their only advantage i.e. they are hard to be maneuvered by other means of transport, so, the machine will also boast its speed and accuracy. Methods that are based on images and already exist regarding wheat field fire detection are capable of carrying out the entire process automatically whereby deep features can be extracted and adjusted to the settings of the relevant art form.

To recognize the smoke correctly from the wheat field fires, the model of the Convolutional Neural Network has been proposed. Thus, to apply the model proposed, an individual dataset with 3900 UAV images has to be formed, capturing the wheat field fires with the smoke. This model has been trained and validated on this data set. The deep network's peculiarities such as the novel integration of Convolutional Block Attention Module in CNN architecture, which is a significant improvement of the state-of-the-art, is one of the things that this project introduces.

The purpose of this module is to allow the network to detect, highlight, and extract the core components related to the smoke detection. Moreover, to enhance the focus of this technology on small, often barely visible traces of smoke that are a regular feature in wheat field fires, we have included a shallower layer inside the network. The model is linked with the decoupled heads, which are other elements improving the process of data extraction from multiple sources and thus, contribute to the successful performance of the model. Early fire detection is critical for limiting damage and halting the spread of the fire, especially through prompt smoke identification. Because there is a lot of combustible material present, they spread quickly and can get challenging to put out. In addition to causing harm to individuals and their belongings, the fires hasten the devastation of the neutral environment. The proposed approach is based on an intelligent method that first properly prepares a dataset of Wheatfield fire images and then classifies it into four classes e.g. High, Medium, Low, and No Fire. It should be noted that the performance of ML and DL models is significantly impacted by the image quality used in the training and testing stages.

The images in the collected dataset were divided into three sets 1) training 2) validation and 3) testing. In each of these sets, the images were further divided into four classes such as 1) High 2) Medium 3) Low, and 4) No fire incident. All these images were carefully divided into sets and classes considering the content of the images.

The method employed in this study analyzes the UAV photos of wheat fields to identify input images that are subjected to a CNN model with multiple convolutional layers and Relu as an activation function. In addition to helping to increase the depth of the created feature map, the pooling layers aid in decreasing the spatial dimensions of the data. Furthermore, CABM is utilized to extract the salient characteristics from the input images. Convolutional layer output is fed to the flattened layer, which in turn feeds the linked layers. The fully connected layer is in charge of figuring out each class's score.

Using the SoftMax function, this computed score is used to measure probability. Ultimately, the selected class is determined by considering the maximum probability.

With our specialized dataset, this suggested model performs exceptionally well, attaining an accuracy rate of up to 0.91%. These findings demonstrate the potential of our refined CNN model to greatly improve wheat field fire early detection, leading to more efficient fire management tactics and the protection of the environment and agricultural resources.

The dataset for the Wheatfield fire and smoke model was gathered using web-scraping and searching online picture repositories or UAV-captured films to use UAV photos for evaluation. Aerial photos of wheat fields with fire and smoke from field fires were gathered for this investigation. The gathered photos ranged in size from 400 x 300 pixels to 1800 x 1500 pixels. The images showed recent field fire events that have occurred all across the world. The dataset's diversity has improved the algorithm's suitability for intricate wheat settings. A dataset of 1400 images of smoke from non-wheat fires and 2500 photographs of smoke and fire from wheat fires was merged into a single dataset following a manual filtering integrated dataset.

The suggested study also incorporates the CBAM into the CNN to enhance the model's capacity to extract features for the smoke patch characteristics in wheat field fires. Another aspect of the CNN architecture that makes it possible to retrieve relevant data from various sources is its decoupled heads.

5.2 Future work

In the future, enhancing the accuracy and expanding the scope of the study. Creating extensive ethical and legal frameworks to control the installation and functioning of autonomous monitoring systems is essential, in addition to technological breakthroughs.

REFERENCES

1. Shiferaw, B., et al., *Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security.* Food security, 2013. **5**: p. 291-317.

2. Singh, V., H. Singh, and A. Raghubanshi, *Competitive interactions of wheat with Phalaris minor or Rumex dentatus: A replacement series study.* International journal of pest management, 2013. **59**(4): p. 245-258.

3. Han, H., T. Ning, and Z. Li, *Effects of soil tillage and weed control on weed species composition and winter wheat yield in North China.* Journal of Food, Agriculture and Environment, 2014. **12**: p. 266-271.

4. Kolb, L.N., E.R. Gallandt, and E.B. Mallory, *Impact of spring wheat planting density, row spacing, and mechanical weed control on yield, grain protein, and economic return in Maine.* Weed science, 2012. **60**(2): p. 244-253.

5. Pannacci, E., B. Lattanzi, and F. Tei, *Non-chemical weed management strategies in minor crops: A review.* Crop protection, 2017. **96**: p. 44-58.

6. Carver, B.F., *Wheat Science and Trade.* 2009.

7. Kolarić, D., K. Skala, and A. Dubravić, *Integrated system for forest fire early detection and management.* Periodicum biologorum, 2008. **110**(2): p. 205-211.

8. Colomina, I. and P. Molina, *Unmanned aerial systems for photogrammetry and remote sensing: A review.* ISPRS Journal of photogrammetry and remote sensing, 2014. **92**: p. 79-97.

9. Li, X., et al. *A hybrid PSO algorithm based flight path optimization for multiple agricultural UAVs*. in *2016 IEEE 28th international conference on tools with artificial intelligence (ICTAI)*. 2016. IEEE.

10. Trasviña-Moreno, C.A., et al., *Unmanned aerial vehicle based wireless sensor network for marine-coastal environment monitoring.* Sensors, 2017. **17**(3): p. 460.

11. Gonzalez, L.F., et al., *Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation.* Sensors, 2016. **16**(1): p. 97.

12. Zhao, H., et al., *Effects of agricultural biomass burning on regional haze in China: a review.* Atmosphere, 2017. **8**(5): p. 88.

13. Day, N.J., et al., *Wildfire severity reduces richness and alters composition of soil fungal communities in boreal forests of western Canada.* Global change biology, 2019. **25**(7): p. 2310-2324.

14. Çatav, Ş.S., et al., *Effect of fire-derived chemicals on germination and seedling growth in Mediterranean plant species.* Basic and Applied Ecology, 2018. **30**: p. 65-75.

15. Kondratyev, K.Y. and C. Varotsos, *Atmospheric greenhouse effect in the context of global climate change.* Il Nuovo Cimento C, 1995. **18**: p. 123-151.

16. Christopher, S.A., et al., *GOES-8 and NOAA-14 AVHRR retrieval of smoke aerosol optical thickness during SCAR-B.* International Journal of Remote Sensing, 2002. **23**(22): p. 4931-4944.

17. Kaufman, Y., A. Setzer, and C. Justice, *Remote Sensing of Biomass Burning in the Tropics, Fire in the Tropical Biota, 84, 371–399*. 1990.

18. Jethva, H., et al., *Connecting crop productivity, residue fires, and air quality over northern India.* Scientific Reports, 2019. **9**(1): p. 16594.

19. Candiago, S., et al., *Evaluating multispectral images and vegetation indices for precision farming applications from UAV images.* Remote sensing, 2015. **7**(4): p. 4026-4047.

20. Guo, T., T. Kujirai, and T. Watanabe, *Mapping crop status from an unmanned aerial vehicle for precision agriculture applications.* The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2012. **39**: p. 485-490.

21. Grenzdörffer, G. and A. Engel, *B. Teichert, 2008. The Photogrammetric Potential of Low-Cost UAVs in Forestry and Agriculture*. IAPRSSIS.

22. Westoby, M.J., et al., *'Structure-from-Motion'photogrammetry: A low-cost, effective tool for geoscience applications.* Geomorphology, 2012. **179**: p. 300-314.

23. Krizhevsky, A., I. Sutskever, and G.E. Hinton, *Imagenet classification with deep convolutional neural networks.* Advances in neural information processing systems, 2012. **25**.

24. Cheng, W., et al., *High-efficiency chaotic time series prediction based on time convolution neural network.* Chaos, Solitons & Fractals, 2021. **152**: p. 111304.

25. Woo, S., et al. *Cbam: Convolutional block attention module*. in *Proceedings of the European conference on computer vision (ECCV)*. 2018.

26. Hu, J., L. Shen, and G. Sun. *Squeeze-and-excitation networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

27. Wang, B., et al., *A capacitive particle-analyzing smoke detector for very early fire detection.* Sensors, 2024. **24**(5): p. 1692.

28. Matson, M., G. Stephens, and J. Robinson, *Fire detection using data from the NOAA-N satellites.* International Journal of Remote Sensing, 1987. **8**(7): p. 961-970.

29. Prins, E.M. and W.P. Menzel, *Geostationary satellite detection of bio mass burning in South America.* International Journal of Remote Sensing, 1992. **13**(15): p. 2783-2799.

30. Nagaraja Rao, C., L. Stowe, and E. McClain, *Remote sensing of aerosols over the oceans using AVHRR data Theory, practice and applications.* International Journal of Remote Sensing, 1989. **10**(4-5): p. 743-749.

31. Dhammapala, R., et al., *Emission factors from wheat and Kentucky bluegrass stubble burning: Comparison of field and simulated burn experiments.* Atmospheric Environment, 2007. **41**(7): p. 1512-1520.

32. Sharma, A., et al., *IoT and deep learning-inspired multi-model framework for monitoring Active Fire Locations in Agricultural Activities.* Computers & Electrical Engineering, 2021. **93**: p. 107216.

33. Li, Z., S. Nadon, and J. Cihlar, *Satellite-based detection of Canadian boreal forest fires: Development and application of the algorithm.* International Journal of Remote Sensing, 2000. **21**(16): p. 3057-3069.

34. Katayama, H., et al. *Development of the Compact InfraRed Camera (CIRC) for wildfire detection*. in *Remote Sensing System Engineering II*. 2009. SPIE.

35. Mohsan, S.A.H., et al., *Towards the unmanned aerial vehicles (UAVs): A comprehensive review.* Drones, 2022. **6**(6): p. 147.

36. Kim, J., et al., *CBDN: cloud-based drone navigation for efficient battery charging in drone networks.* IEEE Transactions on Intelligent Transportation Systems, 2018. **20**(11): p. 4174-4191.

37. Zhang, S., et al., *Cellular UAV-to-X communications: Design and optimization for multi-UAV networks.* IEEE Transactions on Wireless Communications, 2019. **18**(2): p. 1346-1359.

38. Zhang, S., et al., *Optimizing itinerary selection and charging association for mobile chargers.* IEEE Transactions on Mobile Computing, 2016. **16**(10): p. 2833-2846.

39. Aldhaher, S., et al. *Light-weight wireless power transfer for mid-air charging of drones*. in *2017 11th European Conference on Antennas and Propagation (EUCAP)*. 2017. IEEE.

40. Lu, M., et al., *Wireless charging techniques for UAVs: A review, reconceptualization, and extension.* IEEE access, 2018. **6**: p. 29865-29884.

41. Nourmohammadi, A., M. Jafari, and T.O. Zander, *A survey on unmanned aerial vehicle remote control using brain–computer interface.* IEEE Transactions on Human-Machine Systems, 2018. **48**(4): p. 337-348.

42. Kanellakis, C. and G. Nikolakopoulos, *Survey on computer vision for UAVs: Current developments and trends.* Journal of Intelligent & Robotic Systems, 2017. **87**: p. 141-168.

43. Jakovčević, T., et al. *Notice of Violation of IEEE Publication Principles: Review of wildfire smoke detection techniques based on visible spectrum video analysis*. in *2011 7th International Symposium on Image and Signal Processing and Analysis (ISPA)*. 2011. IEEE.

44. Çetin, A.E., et al., *Video fire detection–review.* Digital Signal Processing, 2013. **23**(6): p. 1827-1843.

45. Jiao, Z., et al. *A deep learning based forest fire detection approach using UAV and YOLOv3*. in *2019 1st International conference on industrial artificial intelligence (IAI)*. 2019. IEEE.

46. Redmon, J., *Yolov3: An incremental improvement.* arXiv preprint arXiv:1804.02767, 2018.

47. Girshick, R., et al. *Rich feature hierarchies for accurate object detection and semantic segmentation*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.

48. Redmon, J. *You only look once: Unified, real-time object detection*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

49. Moumgiakmas, S.S., G.G. Samatas, and G.A. Papakostas, *Computer vision for fire detection on UAVs—From software to hardware.* Future Internet, 2021. **13**(8): p. 200.

50. Frizzi, S., et al. *Convolutional neural network for video fire and smoke detection*. in *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*. 2016. IEEE.

51. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

52. Russakovsky, O., et al., *Imagenet large scale visual recognition challenge.* International journal of computer vision, 2015. **115**: p. 211-252.

53. Sreeja, S., et al. *Classifying Chest X-rays for COVID-19 using Deep Learning*. in *2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*. 2023. IEEE.

54. Bayegizova, A., et al., *Fire detection using deep learning methods.* International Journal of Electrical & Computer Engineering (2088-8708), 2024. **14**(1).

55. Chwalek, P., et al., *Downwind Fire and Smoke Detection during a Controlled Burn—Analyzing the Feasibility and Robustness of Several Downwind Wildfire Sensing Modalities through Real World Applications.* Fire, 2023. **6**(9): p. 356.

56. Jiang, Y., et al., *An Enhanced Algorithm for Active Fire Detection in Croplands Using Landsat-8 OLI Data.* Land, 2023. **12**(6): p. 1246.

57. Zhou, H., et al., *A computer-vision-based deep learning model of smoke diffusion.* Process Safety and Environmental Protection, 2024. **187**: p. 721-735.

58. Aitimov, M., et al., *Predicting and detecting fires on multispectral images using machine learning methods.* International Journal of Electrical & Computer Engineering (2088-8708), 2024. **14**(2).

59. Witayangkurn, A., et al., *Real-time monitoring system using unmanned aerial vehicle integrated with sensor observation service.* The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2012. **38**: p. 107-112.

60. Wang, B., et al., *Research on a capacitive particle analysis smoke detector.* Scientific reports, 2024. **14**(1): p. 11319.

61. Haq, M.R.U., *A REVIEW ON REPAIRING FIRE-DAMAGE GYPSUM BOARD.*

62. Acharya, P., S. Sreekesh, and U. Kulshrestha, *GHG and aerosol emission from fire pixel during crop residue burning under rice and wheat cropping systems in North-West India.* The International Archives Of The Photogrammetry, Remote Sensing And Spatial Information Sciences, 2016. **41**: p. 753-760.

63. Ma, L., et al., *Deep learning in remote sensing applications: A meta-analysis and review.* ISPRS journal of photogrammetry and remote sensing, 2019. **152**: p. 166-177.

64. Lee, J., et al., *Smoke Detection Based on Fire-starting Regions.* IEIE Transactions on Smart Processing & Computing, 2019. **8**(6): p. 475-481.

65. Olivares-Mercado, J., et al., *Early fire detection on video using LBP and spread ascending of smoke.* Sustainability, 2019. **11**(12): p. 3261.

66. Yuan, C., Y. Zhang, and Z. Liu, *A survey on technologies for automatic forest fire monitoring, detection, and fighting using unmanned aerial vehicles and remote sensing techniques.* Canadian journal of forest research, 2015. **45**(7): p. 783-792.

67. Xiao, B. and C. Yan, *Efficient attention-based networks for fire and smoke detection.* Journal of Electronic Imaging, 2024. **33**(5): p. 053014-053014.

68. Yang, G., et al., *AgriFireInfo v1. 0: An Open-Source Platform for the Monitoring and Management of Open-Field Crop Residue Burning.* Fire, 2024. **7**(3): p. 63.

69. Li, D., R. Yang, and C. Su, *Generative adversarial network based on LSTM and convolutional block attention module for industrial smoke image recognition.* Computer Science and Information Systems, 2023(00): p. 27-27.

70. Liu, Z., et al., *YOLOv8 for Fire and Smoke Recognition Algorithm Integrated with the Convolutional Block Attention Module.* Open Journal of Applied Sciences, 2023. **14**(1): p. 159-170.

71. Hong, R., et al., *Yolo-Light: Remote Straw-Burning Smoke Detection Based on Depthwise Separable Convolution and Channel Attention Mechanisms.* Applied Sciences, 2023. **13**(9): p. 5690.

72. Cheng, G., X. Chen, and J. Gong, *Deep convolutional network with pixel-aware attention for smoke recognition.* Fire technology, 2022. **58**(4): p. 1839-1862.

73. Chen, H. and Q. Ji. *Convolutional Neural Network with Attention Mechanism for Image-Based Smoke Detection*. in *2022 4th International Conference on*

Advances in Computer Technology, Information Science and Communications (CTISC). 2022. IEEE.

74. Khan, S. and A. Khan, *Ffirenet: Deep learning based forest fire classification and detection in smart cities.* Symmetry, 2022. **14**(10): p. 2155.

75. He, L., et al., *Efficient attention based deep fusion CNN for smoke detection in fog environment.* Neurocomputing, 2021. **434**: p. 224-238.

76. Ba, R., et al., *SmokeNet: Satellite smoke scene detection using convolutional neural network with spatial and channel-wise attention.* Remote Sensing, 2019. **11**(14): p. 1702.

77. Zhao, Y., et al., *Saliency detection and deep learning-based wildfire identification in UAV imagery.* Sensors, 2018. **18**(3): p. 712.

78. Janga, B., et al., *A review of practical ai for remote sensing in earth sciences.* Remote Sensing, 2023. **15**(16): p. 4112.

79. John, A., et al., *Multimodal multiresolution data fusion using convolutional neural networks for IoT wearable sensing.* IEEE Transactions on Biomedical Circuits and Systems, 2021. **15**(6): p. 1161-1173.

80. Askay, D.A., et al., *Enhancing group social perceptiveness through a swarm-based decision-making platform.* 2019.

81. Shi, Y., *Particle swarm optimization.* IEEE connections, 2004. **2**(1): p. 8-13.

82. Yazdan, M.M.S., et al., *Forecasting Energy Consumption Time Series Using Recurrent Neural Network in Tensorflow.* 2022.

83. Wang, D., et al., *Evaluation of Scikit-Learn Machine Learning Algorithms for Improving CMA-WSP v2. 0 Solar Radiation Prediction.* Atmosphere, 2024. **15**(8): p. 994.

84. Zhan, J., et al., *A High-precision Forest Fire Smoke Detection Approach Based on DRGNet to Remote Sensing Through Uavs.* 2021.

85. Syahputra, H. and A. Wibowo, *Comparison of Support Vector Machine (SVM) and Random Forest Algorithm for Detection of Negative Content on Websites.* Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI), 2023. **9**(1): p. 165-173.

86. Cao, Y., et al., *An attention enhanced bidirectional LSTM for early forest fire smoke recognition.* IEEE Access, 2019. **7**: p. 154732-154742.

87. Mustafa, N.E. and F. Alizadeh, *Unmanned aerial vehicle (UAV) images of road vehicles dataset.* Data in brief, 2024. **54**: p. 110264.

88. Yang, H., J. Wang, and J. Wang, *Efficient Detection of Forest Fire Smoke in UAV Aerial Imagery Based on an Improved Yolov5 Model and Transfer Learning.* Remote Sensing, 2023. **15**(23): p. 5527.

89. Cruz, H., et al., *Efficient forest fire detection index for application in unmanned aerial systems (UASs).* Sensors, 2016. **16**(6): p. 893.

90. Liu, Z., et al., *A Real-Time Registration Algorithm of UAV Aerial Images Based on Feature Matching.* Journal of Imaging, 2023. **9**(3): p. 67.

91. Guede-Fernández, F., et al., *A deep learning based object identification system for forest fire detection.* Fire, 2021. **4**(4): p. 75.

92. Muhammad, K., et al., *Efficient deep CNN-based fire detection and localization in video surveillance applications.* IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2018. **49**(7): p. 1419-1434.

93. Barmpoutis, P., et al., *A review on early forest fire detection systems using optical remote sensing.* Sensors, 2020. **20**(22): p. 6442.

94. Pan, H., et al., *Additive neural network for forest fire detection.* Signal, Image and Video Processing, 2020. **14**(4): p. 675-682.

95. Zhang, J., et al., *ATT squeeze U-Net: A lightweight network for forest fire detection and recognition.* IEEE Access, 2021. **9**: p. 10858-10870.

96. Hislop, S., et al., *A satellite data driven approach to monitoring and reporting fire disturbance and recovery across boreal and temperate forests.* International Journal of Applied Earth Observation and Geoinformation, 2020. **87**: p. 102034.

97. Tian, L., et al., *Automatic detection of forest fire disturbance based on dynamic modelling from MODIS time-series observations.* International Journal of Remote Sensing, 2018. **39**(12): p. 3801-3815.

98. Wang, Z., et al., *A smoke detection model based on improved YOLOv5.* Mathematics, 2022. **10**(7): p. 1190.

APPENDICES

APPENDIX A

A. Main Results: Deep Learning Methods and UAVs for Smoke Detection in Various Environments.

Table 1: Detection Deep Learning Methods and UAVs for Smoke Detection

Author	Methods	Objective	Highlights	Results
[91]	Faster R-CNN	Smoke classification with a deep learning-based method	Smoke detection	Accuracy = 90%
[92]	SqueezeNet Model	Using semantic knowledge of the scene to locate and detect fire.	The balance between efficiency and accuracy in fire detection	Achieved the best and required results
[93]	Optical technology	An analysis of algorithms for detecting smoke and flames.	3 types identified there: 1. Terrestrial 2. spaceborne 3. airborne	Provide poor results for smoke detection results
[89]	Forest Fire Detection Index Method	Larger area targeted	22 f/s	precision = 96.62%
[94]	Two 1. AddNet	Deep neural networks provide a better result	AddNet's performance is comparable to	12.4% Reduce time

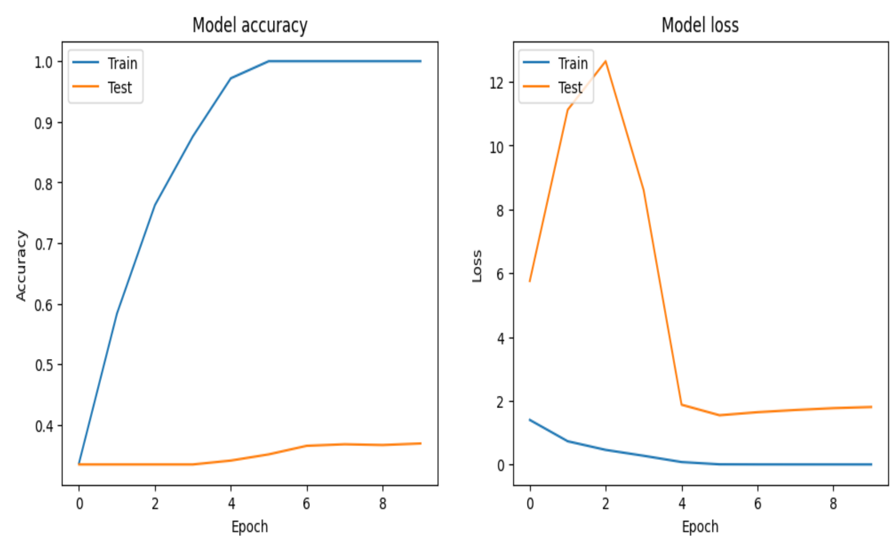
	2. Deep neural network		that of standard CNNs.	
[95]	SqueezeNet Model	Classification of fire	Very good result with accuracy = 0.93	Very good results
[96, 97]	MODIS	Tracking Burned Severity Data Set Trends.	Provide good result	Automatically and efficiently identify regions of disruption.
[98]	YOLOv5	Functions for unbalanced features on various YOLOv5 scales	85 FPS Detection rate	4.4% more speed

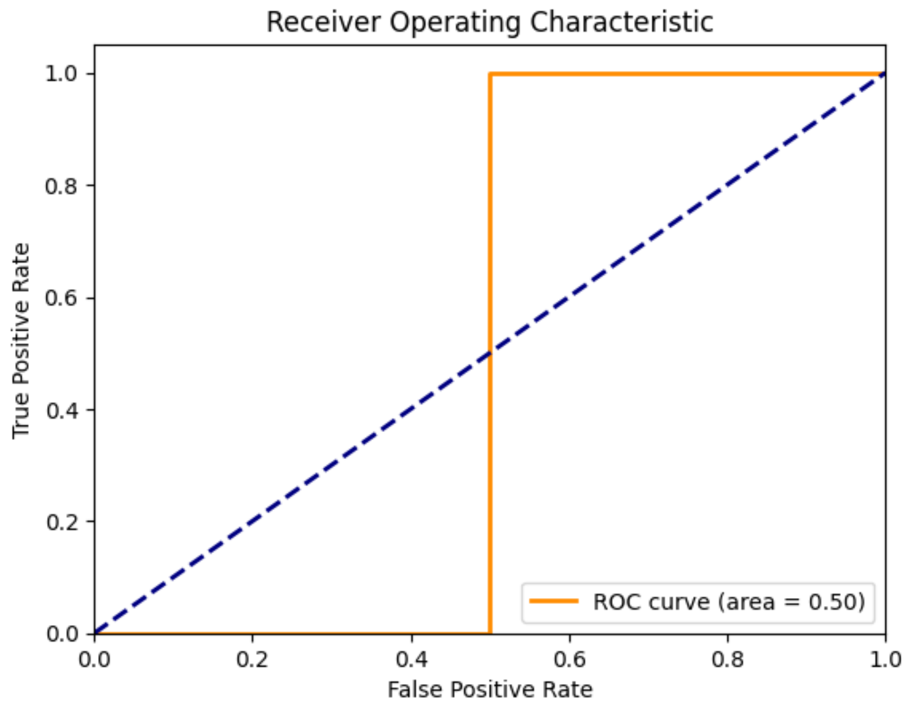
APPENDIX B

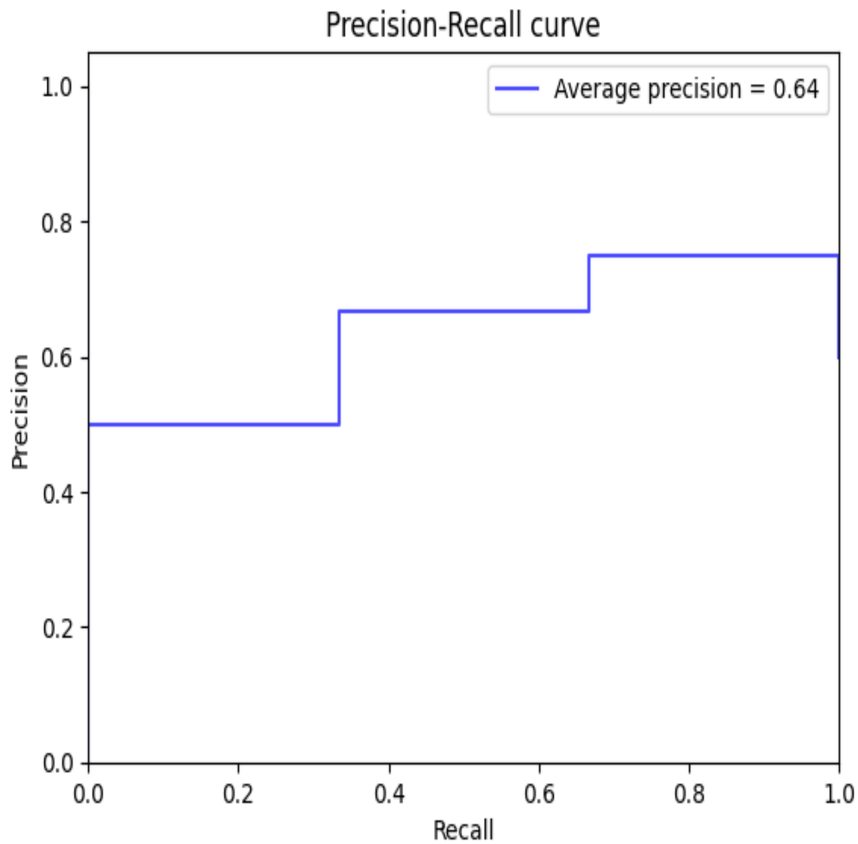
B. Results, and Confusion Matrix

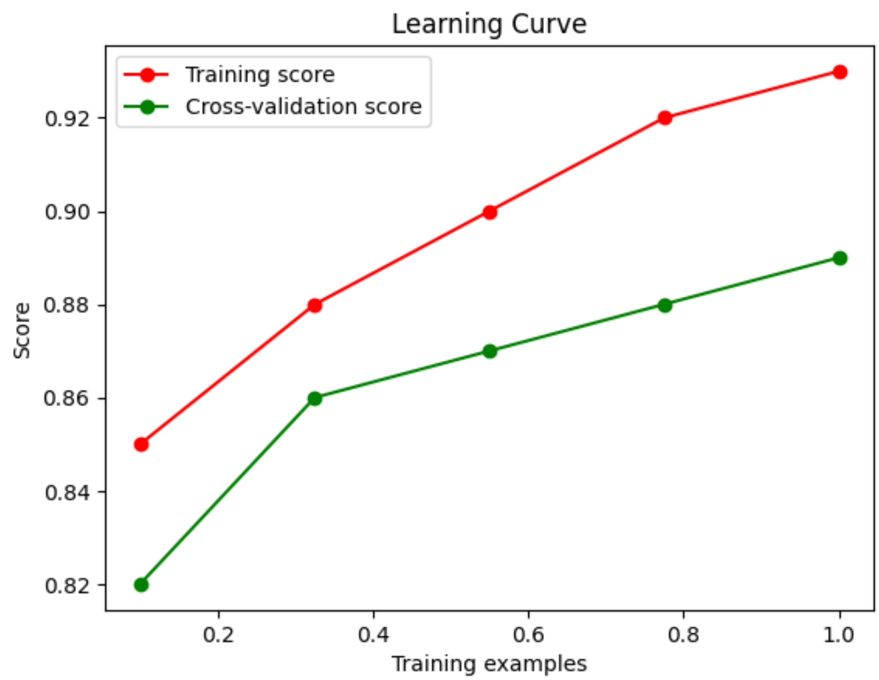
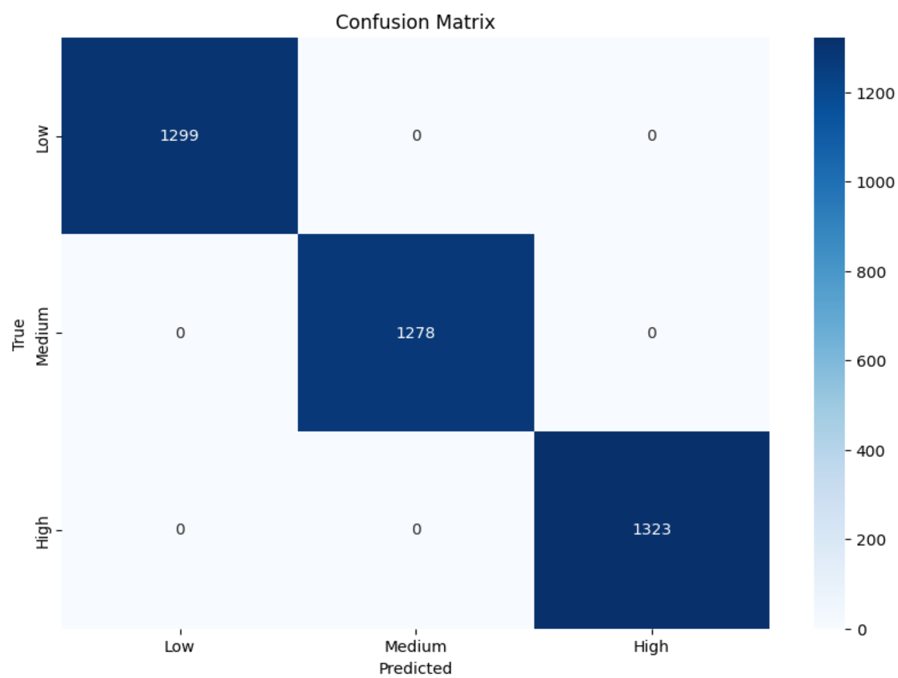
This section presents a comprehensive analysis of UAV images dataset for experimental methodologies. The visualizations include plots of training accuracy, training loss values, Receiver Operating Characteristic (ROC) curves, and confusion matrices. These plots provide critical insights into the model's performance and its effectiveness in accurately classifying the data.

Table 2: The Visualization of Results and Confusion Matrix

Accuracy Results	

ROC	 Receiver Operating Characteristic 1.0 0.8 0.6 0.4 0.2 0.0 True Positive Rate 0.0 0.2 0.4 0.6 0.8 1.0 False Positive Rate ROC curve (area = 0.50)

Precision-Recall	

<table>
<tr>
<td>Learning Curve</td>
<td>
</td>
</tr>
<tr>
<td>Confusion Matrix</td>
<td>
</td>
</tr>
</table>