A project report on

A COMPUTER VISION APPROACH FOR REAL-TIME DETECTION OF WILDFIRE AND HUMANS AMIDST IT

Submitted in partial fulfillment for the award of the degree of

Bachelor of Technology in Computer Science and Engineering

by

KIRTHANA B (21BCE5129)

LOHITAKSHA B C (21BCE5501)

SATHYABAMA C (21BEC1330)

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

April,2025

A project report on

A COMPUTER VISION APPROACH FOR REAL-TIME DETECTION OF WILDFIRE AND HUMANS AMIDST IT

Submitted in partial fulfillment for the award of the degree of

Bachelor of Technology in Computer Science and Engineering

by

KIRTHANA B (21BCE5129)

LOHITAKSHA B C (21BCE5501)

SATHYABAMA C (21BEC1330)

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

April,2025

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

DECLARATION

I hereby declare that the thesis entitled "A Computer Vision Approach for Real-Time Detection of Wildfire and Humans Amidst It" submitted by **LOHITAKSHA B C (21BCE5501)**, for the award of the degree of Bachelor of Technology in Computer Science and Engineering, Vellore Institute of Technology, Chennai is a record of Bonafide work carried out by me under the supervision of Prof. Nivedita M.

I further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Place: Chennai

Date: **Signature of the Candidate**

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

School of Computer Science and Engineering

CERTIFICATE

This is to certify that the report entitled **"**A Computer Vision Approach for Real-Time Detection of Wildfire and Humans Amidst It**"** is prepared and submitted by **LOHITAKSHA B C (21BCE5501)** to Vellore Institute of Technology, Chennai, in partial fulfillment of the requirement for the award of the degree of **Bachelor of Technology in Computer Science and Engineering** is a bonafide record carried out under my guidance. The project fulfills the requirements as per the regulations of this University and in my opinion meets the necessary standards for submission. The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma and the same is certified.

Signature of the Guide:

Name: Dr./Prof.

Date:

Signature of the Examiner	Signature of the Examiner
Name:	Name:
Date:	Date:

Approved by the Head of Department,
(School of Computer Science and Engineering)
Name: Dr. Nithyanandam P
Date:

ABSTRACT

Recent rise of weather-driven wildfires demands immediate detection systems that serve as critical wildfire response measures. The monitoring systems that use ground-based methods fall short since they lack comprehensive coverage and they take time to detect fires and involve human risks during inspection. These limitations of ground-based detection face three corresponding solutions from my project in the form of drone imagery integration and sensor networks and flame sound identification. An aerial monitoring system consists of the YOLOv8 and EfficientNetV2 deep learning models to detect smoke, fire along with human presence in video sequences. The detection mechanism includes visual pattern recognition combined with real-time heat alert display which uses thermal data and an MQ-135 gas sensor connected to an ESP8266 microcontroller performs environmental monitoring. Fire and human presence data passes through ThingSpeak for visual representation and the system sends email notifications containing the data as well. An audio module in the system receives sound files first applies noise reduction then employs voice recognition to identify human speech before deriving text from recorded audio to look for potential trapped victims. The combination of multiple real-time detection methods strengthens accuracy levels which leads to more effective early responses alongside better efficiency in wildfire response systems.

ACKNOWLEDGEMENT

It is my pleasure to express with deep sense of gratitude to Prof. Nivedita M, Assistant Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, for her constant guidance, continual encouragement, understanding; more than all, she taught me patience in my endeavor. My association with her is not confined to academics only, but it is a great opportunity on my part of work with an intellectual and expert in the field of Computer Vision Applications.

It is with gratitude that I would like to extend my thanks to the visionary leader Dr. G. Viswanathan our Honorable Chancellor, Mr. Sankar Viswanathan, Dr. Sekar Viswanathan, Dr. G V Selvam Vice Presidents, Dr. Sandhya Pentareddy, Executive Director, Ms. Kadhambari S. Viswanathan, Assistant Vice-President, Dr. V. S. Kanchana Bhaaskaran Vice-Chancellor, Dr. T. Thyagarajan Pro-Vice Chancellor, VIT Chennai and Dr. P. K. Manoharan, Additional Registrar for providing an exceptional working environment and inspiring all of us during the tenure of the course.

Special mention to Dr. Ganesan R, Dean, Dr. Parvathi R, Associate Dean Academics, Dr. Geetha S, Associate Dean Research, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai for spending their valuable time and efforts in sharing their knowledge and for helping us in every aspect.

In jubilant state, I express ingeniously my whole-hearted thanks to Dr. Nithyanandam P, Head of the Department, B.Tech. Computer Science and Engineering and the Project Coordinators for their valuable support and encouragement to take up and complete the thesis.

My sincere thanks to all the faculties and staffs at Vellore Institute of Technology, Chennai who helped me acquire the requisite knowledge. I would like to thank my parents for their support. It is indeed a pleasure to thank my friends who encouraged me to take up and complete this task.

Place: Chennai

Date: **Name of the student**
LOHITAKSHA B C

CONTENTS PAGE NO.

LIST OF FIGURES PAGE NO.

LIST OF ACRONYMS

UAV	:	Unmanned Aerial Vehicle
YOLO	:	You Only Look Once
CNN	:	Convolutional Neural Network
LSTM	:	Long Short-Term Memory
mAP	:	Mean Average Precision
IoU	:	Intersection over Union
AI	:	Artificial Intelligence
ML	:	Machine Learning
DL	:	Deep Learning
IoT	:	Internet of Things
ESP8266	:	Espressif Systems Protocol 8266 Microcontroller
MQ-135	:	Gas Sensor Module (Air Quality Detection)
Blynk	:	IoT Platform for Microcontroller Integration
GPS	:	Global Positioning System
API	:	Application Programming Interface
GUI	:	Graphical User Interface
mAP@50	:	Mean Average Precision at 50% IoU Threshold
mAP@50-95	:	Mean Average Precision averaged over IoU thresholds from 50% to 95%
DFL Loss	:	Distribution Focal Loss
Thingspeak	:	IoT Analytics Platform for Data Logging
FFT	:	Fast Fourier Transform (used in audio filtering)
SNR	:	Signal-to-Noise Ratio
NLP	:	Natural Language Processing

CHAPTER 1

INTRODUCTION

1.1 PROBLEM STATEMENT

Over the years wildfire occurrences have become more frequent, intense and unpredictable, bringing a serious threat to ecosystems, infrastructure and the human lives. Extreme climatic conditions such as prolonged drought, rising temperatures, erratic wind patterns, among others, are being intensified by global climate change, fueling these fires. While the resource intensity of current traditional approaches to wildfire detection such as ground based observation towers, patrols, or fixed CCTV installations is not trivial, they are also considerably ineffective in many situations. The drawbacks of these methods are that they inherently suffer lack of detection in time, a restricted field of view and are not able to react quickly to fast spreading fires. Additionally, the risk of human casualties in such hazardous environments is increased through deployment of human personnel, because early warning systems can go undetected with an outbreak in its early stages. Yet, in spite of these technological advancements, an integrated wildfire detection framework is neither present nor can it perform in an autonomous, real time, and across disparate multimodal data sources such as visual, thermal, environmental, and acoustic. However, this existing state of detection is constrained by either a single sensor type or humaners observations leading to the timeliness of and response to this emergency being unreliable and of limited time. Such an intelligent, autonomous, and multilayer system has an urgent need of being developed for detection at the inception stage of a wildfire even in remote areas, and prompt alerting of the responders through automated communication channels. Not only does addressing this problem allow mitigating environmental damage, but this problem is also crucial for ensuring public safety and resource management during natural disasters.

1.2 MOTIVATION AND BACKGROUND

The rise in the level of wildfire cases worldwide has been a significant reason behind the creation of intelligent, autonomous fire detection systems. Thousands of hectares of forest land are incinerated, there is a dramatic reduction of biodiversity, and countless communities face life threatening danger, for example, being buried in landslides and debris flows, if prompt intervention is not provided. The most alarming of these observations is that many of these fires could have been controlled in their early stages, had there been an adequate and better detection system in place. Although still in use, traditional approaches are reactive, not preventive, and are rarely effective in the photoconductive times and unanticipated conditions created by modem wildfires. These shortcomings motivated me to investigate a system that can both view and comprehend what is happening in real time on the dangerous land and reduces the necessity of human involvement.

The reason I wanted to leverage new technology of deep learning, real time video analytics, voice recognition and micro controller based sensor network with UAV (unmanned aerial vehicle) based surveillance was realizing how quickly all these emerging technologies could be combined in order to create a whole solution. Onboard imaging systems can be also enhanced with intelligent models to recognize not only flames or smoke, but also those of human figures in distress, and drones are the perfect mobility and coverage for vast landscapes. At the same time, sensor nodes deployed on the ground can in turn recognise simultaneous environmental factors like smoke particles and gas concentration. It also introduces a very natural aspect to the system in capturing the sounds of human voices in disaster zones with impaired visibility, giving the search and rescue efforts a higher degree of targeting. The whole goal was to create not just a technical solution, but a real world deployable system with high speed, accuracy and autonomy to react to the uncertainty of natural disasters. These technologies are integrated with the intent to overcome the distance between early detection and successful emergency response, to identify and save lives and protect our natural ecosystems.

1.3 LIMITATIONS OF EXISTING SYSTEM

However, only small fraction of those systems categories currently in use are still based on conventional, ground based techniques that are slow, inflexible, and prone to failure under fire conditions. CCTV cameras, thermal towers and lookout posts all suffer from poor spatial coverage and need to be constantly monitored by humans such that they are unable to cope with extensive or remote forest areas. These systems also don to detect fires until smoke or flames are visually conspicuous, by which time the fire has frequently become too large to control. Satellite based monitoring may be useful for a broader geographical observation but it has limited temporal resolution and cannot provide real time data for the timely emergency response. Finally, since cloud cover, sensor resolution limitations, and communication delays degrade the utility of such systems in real-world scenarios, these systems are used with limited applicability only.

A major limitation in many such frameworks is that they are still mostly siloed. Almost all such systems rely on visual inputs, either through smoke or flame detection, without integrating other modalities of equal importance such as air quality data, temperature anomalies or acoustic signals from the surrounding environment. As a result it leads to poor situational awareness and prohibits the ability to identify ‘early warning’ cues, for example, rising carbon emissions or human voice distress in affected areas. However, even when environmental sensors are used, these do not form part of a consolidated or clever choice making center. Also, existing systems don’t have real time mechanism of alerts. Without immediate communication, alerts by email or integrated dashboards for instance make it difficult to respond ‘fast’ and coordination between the emergency teams is reduced significantly. Being traditional setups, these initial setups cannot scale in their current form to be met by the quickly changing landscape of wildfire threats; the only way to achieve this scale is through technology such as AI driven automation. Clearly, these limitations imply that there is still some considerable need for an

integrated, AI enabled, multi sensor platform that is able to detect, analyze, and communicate wildfire risks in real time, and that is an area where my proposed system provides a considerable advancement.

1.4 PROPOSED SOLUTION OVERVIEW

To overcome the crucial shortcomings of the existing wildfire detection systems, I outline a comprehensive, multi-modal detection system using UAV based imaging, real time environmental sensing, and intelligent acoustic analysis. The purpose of this system is to have it run with little human intervention, to conduct continuous aerial surveillance, perform early anomaly detection, and communicate directly with emergency response teams. This solution is based on integration of visual, thermal, environmental and audio data sources to complement each other through unique insights gained, therefore creating a more complete picture of a potential wildfire scenario.

A deep learning architecture at the heart of the visual detection system is made with YOLOv8's object detection strength and EfficientNetV2B0's feature extraction speed. These models work together to analyze frames captured by UAV mounted cameras and detect both fire and smoke as well as human beings in areas where they may be dangerous. A lightweight implementation could run on a system that can actually stream webcam or drone feed and expose these cues in real time for additional real time responsiveness. The system also includes features such as temperature anomaly detection with thermal imaging and intensity mapping that enables the model to raise a flag before visible smoke materializes, which is critical for early intervention.

The image-based detection is complemented by an ESP8266 microcontroller powered with MQ-135 gas sensor that constantly monitors air quality, for example alerts in case of high CO_2 or other combustible gases. Live visualization and analysis of trends are provided by this sensor sending its data to a centralized IoT platform (ThingSpeak) for backup in case of poor visibility or when visual sensors become obstructed. This third layer of this framework serves as human layer: capturing voice forms of audio signals through microphones, to isolate voice activity from outdoor noise with the help of noise filtering models, and then speaking recognition that enables transcribing any detected voices. The transcripts of these events are sent automatically to appropriate authorities through email notifications which help notify with real time alerts with contextual information about the issue.

This proposed framework integrates vision, sound and air sensing into one robust system that can identify stalled individuals, fire its presence in its early stages, and communicate actionable information. The integration of these components not only improves the detection accuracy but also facilitates the practical deployability of the system in the rescue and response operations for real world wildfire surveillance.

1.5 OBJECTIVES OF THE STUDY

The results of this study are aimed to develop and validate an intelligent, multi-modal wildfire detection system for early warning and readiness to respond to emergency conditions by using a UAV based imaging, environmental sensing and acoustic monitoring the. Consequently, each component of this system has been purposefully chosen to fill a particular gap that exists in existing approaches to wildfire detection. In a nutshell, this research attempts a feasible technical framework that, at the same time, is practically implementable in a dynamic real world fire fighting scenario.

The third core objective is to integrate the real time object detection using the deep learning model such as YOLOv8 and EfficientNetV2B0 to identify correctly fire, smoke and there presence human in the live drone footage or webcam streams. The goal is to optimize these models for both performance as well as efficiency to allow the solution to be deployed on the real time inference speed without sacrificing any detections accuracy, and thus to be suitable for deployment on lightweight edge devices. A key goal is to develop an anomaly detection mechanism based on heatmaps to detect abnormal temperature spikes at a stage before smoking or pre-smoke, something that is critical for proactive mitigation.

In addition, this study attempts to also interface the ESP8266 microcontroller with the MQ-135 gas sensor to continuously monitor air quality parameter indicative of the polluting elements in the case of fire-based pollution. This data is to be transmitted securely to a ThingSpeak cloud based IoT dashboard to be visualized in real time with anomaly tracking. In terms of acoustic front, the system tries to detect human voices in a noisy environment via audio denoising algorithms, followed by speech transcription and automated emails alerting human rescuers during search and rescue operations, providing an essential human component in the system.

Additionally, it is a major target to create an integrated whole pipeline that features automatic alert program, real time dashboard watch, and smart decision making via multi modal input fusion. Together, these objectives strive to show that a smart detection framework capable of individually processing and policing vision, audio and environmental data works as a sustainable early wildfire detection and response improvement solution.

1.6 SCOPE OF THE PROJECT

The focus of this project is the design, implementation, and evaluation of an integrated wildfire detection system that combines aerial vision, thermal analysis, environmental gas sensing, and voice detection as a means of an early warning and set of situational awareness. The system is developed such that it could autonomously work in simulated or real world environments, when input from vision sources are drones or webcams, environmental monitoring device is microcontroller based sensor nodes, while acoustic signal analysis is done with microphone modules. The idea of the project is to build a

solid pipeline that can analyze multi-modal data in real time, and from that, produce alerts that can help make quick decisions during the manifolds of the fire.

More specifically, the project involves developing a vision based detection model to detect fire, smoke, and presence of human in video frames using combination of YOLOv8 and EfficientNetV2B0 and also real time temperature mapping for identifying the anomalous heat zones within thermal imagery. The system also includes an integration between an ESP8066 and an MQ135 sensor to record and send air quality information that is visualized on a cloud platform like ThingSpeak for monitoring air pollutants raised by firebreaks. The noise filtering, voice detection and speech to text modules are part of the acoustic subsystem followed by the delivery of human audio identification real time email alerts for designated contacts.

Nevertheless, the project described herein is limited by a number of practical constraints. Having safety, regulatory, and resource limitations, the project does not involve large scale drone deployment or real time field testing in the forest fire environment proper. Instead, accuracy, latency and efficiency of the system are validated through controlled experiments, simulation datasets. The audio and environmental modules were tested in realistic indoor and outdoor conditions, whereas the thermal imaging module is illustrated through the use of created heatmaps and data available in public domain. Email and ThingSpeak integration are the only types of cloud communication, and advanced routing or emergency dispatch protocol are not in scope.

The project demonstrates technical feasibility and real time responsiveness of a unified wildfire detection framework and in general, the overall aims can be generalized and scaled, ruggedized, and field tested for deployment in large forested or disaster prone regions.

1.7 RESEARCH CONTRIBUTIONS

Overall, this project reports a set of unique and practical contributions that collectively overcome the shortcomings of present day wildfire detection systems and poses a fresh and holistic process focused on early detection, real time responsiveness and human centric emergency response. It is part of one of the most significant contributions in the development of a multidimensional wildfire detection system composed of aerial visual data, environmental sensing, and acoustic signal processing into a single, cohesive framework. Instead of just looking for visual cues, the system increases the detection surface by ancillary gas concentration trends, thermal anomalies and voice sources, which enhance the system's sensitivity and emphasizes the unpredictability of the system being used.

One of the key technical contributions we make is to use a hybrid deep learning architecture that combines YOLOv8 with EfficientNetV2B0 to perform highly accurate as well as fast object detection from UAV video streams or webcam videos. The generalization of such a combination against diverse datasets (detection of fire, smoke

and human presence) is better than conventional kinds of models, particularly when those datasets (fire, smoke and detector of human presence) present varying lighting and environmental conditions. Furthermore, the system detects the potential of thermal anomaly as intensity mapping of thermal data, and thereby it can perform early identification of fire risk zones even before visible smoke comes up, which many traditional vision only systems cannot make.

The integration of ESP8266 and sensor MQ-135 for air quality monitoring with real time visualization and storage on ThingSpeak IoT platform is another one of its innovative contributions. In addition to this, it makes another data layer without which the smoke detection will not be complete and also a reliable monitor in cases when the line of vision is not possible due to obstructions. As important as audio processing module that employs noise reduction technique to identify human voice activity and execute speech to text transcription in order to deliver automated email alerts with the actionable information. This novel addition places human detection and possibly rescue scenarios into the scope, which are totally ignored by most wildfire systems.

The other major contribution is the development of a fully enabled alert and visualization mechanism consisting of cloud dashboards, automated email, and real time monitoring tools. Closing the loop between detection and emergency response entails that the system can autonomously generate alerts without human supervision. Overall, this work not only improves the performance of the current wildfire detection field through technology fusion, but also creates a strategy of a scalable, modular framework that can be applied in other disaster management, environmental monitoring, or public safety use cases.

CHAPTER 2

LITERATURE SURVEY

2.1 LITERATURE SURVEY

This study explores [1] combining artificial intelligence (AI) with drone technology has made wildfire detection and response much more efficient. Besides flying over forests, these AI driven drones are actually actively challenging fires using their own lenses to detect fire signs, analysing heat and smoke patterns, and assisting responders on effective and speedier decisions. The power of these systems comes from that they can run in real time. Trained algorithms, which detect the shallow visual signs of fire as they flicker in live video feeds, can cause a drone to feed the raw video to a computer, which then renders it faster than the human eye can detect a fire. The smart systems can alert whether it be from a sudden high in heat or a wisp of smoke and not have any humans be on top of them all the time.

This space is one of the more advanced developments, and their application to explainable AI (XAI) is one of them. Now this branch of AI guarantees that the decisions made by these autonomous systems would not be a black box. But what if a drone detects a fire and XAI can explain why they decided that represented a fire? Possibly because they observed a certain move in the smoke, or an abnormal heat signature. This openness adds to trust and makes the system more depend on critical situations. Moreover, there is an increasing integration of drones with data from IoT (Internet of Things) sensors that are scattered across forests. The second batch of sensors are these sensors that monitor the environment for changes in humidity, gas concentration, and temperature, allowing drones to have a richer information set to work with.

However, the technology still faces real-world hurdles. Drones are constrained by available battery life as well as poor or even unstable connections to data for remote or mountainous areas. Operations can be interfered with by weather, terrain and communication blackouts. For that reason, researchers are trying to develop systems that merge drone footage with satellite imagery and ground sensor information. This multi layered view paints a more complete picture of fire prone surfaces reducing the ability to monitor poorly, and moreover allows for more predictive monitoring.

This paper [2] analyses the study of Drones are far beyond detection, as they are making a huge difference in search and rescue efforts during wildfire events. Drones can step in in the absence of ground teams for distance, smoke, collapsing trees, or intense heat where a flying view is safer. They use infrared cameras that detect warmth, even if people are trapped, as long as they are still warm, through thick smoke or total darkness. GPS is used by them to mark very precise locations so that the rescue teams are able to be guided to where they’re needed.

Drones are suited to different rescue scenarios. Drones are ideal for scanning out wide areas quickly but quadcopters are better at hovering over particular sites, maneuvering around difficult lands and rough terrains and so on. And real world examples have proven how these drones have enabled quicker finds of people alive, as well as the mapping of safe evacuation routes and the delivery of up to the minute information about the situation to emergency teams on the ground. However, like the operations themselves, there are challenges to these operations too. The age of a battery still limits the flight time, plus there are restrictions on where you can fly in restricted or emergency airspace.

Since reliability and coverage are important, developers are advancing AI based systems that involve that drones may make decision by themselves, for example, adjusting the flight path from obstacles, or coordinating with other drones covering large search areas. Drones are also being introduced to new machine learning models that will help the drones overperform in their ability to understand their surroundings and decide which areas to focus more on. Eventually, we might be able to take advantage of drones working in conjunction, each gaining and growing ways to learn and adapt in the real world for more per unit than previous generations before it.

The author [3] suggests that Among other things, AI and machine learning can also turn out to be powerful tools in predicting where and when wildfires may start. These systems study past fires' data and the current weather conditions, satellite images, and vegetation patterns to get an idea of what fire might look like at its beginning stages that humans may overlook. By training deep learning models such as CNNs and RNNs to locate these subtle patterns before a fire, such as heat anomalies, dry vegetation, they are able to save lives.

These aren't just alarm systems—they can also predict how a fire might spread and become—weaker systems can actually help fire fighters predict how a fire will become. Any kind of predictive power is very valuable when it comes to planning evacuations or where to place firefighting teams. The researchers have found that training these models on odd and rare examples — think extreme fire conditions — helps to make these models are more resilient and adaptable in the field.

No doubt, running such complex algorithms on such small devices as drones is no trivial task. These platforms have limitations in terms of the amount of battery life and processing power they have. To do that, scientists are developing these AI models to be lighter and faster, so they can still function as intended on low power devices. Thus, the mechanics such as compressing the model, lowering calculations, even sharing the burden with a network of devices belong to this.

From this, the systems are developing to not only detect fires, but to recommend the best way to stop them. For example, reinforcement learning helps an AI simulate different firefighting strategies in a virtual environment and provide a suggestion of the best effective one for real life use. This brings an element of intelligence to director; the AI becomes an active part of the firefighting team instead of a passive observer.

Instead, looking forward aims to integrate cohorts of machines that include ground sensors, drones, satellites, and AI models. They will not only offer early warnings, but actually offer guides in real time during emergencies. Perhaps our best hope for curbing wildfire threat lies with AI allowing access to the technology and the models that become more and more intelligent.

The author [4] discusses the Continuous monitoring of massive amount of forested landscapes without human oversight is becoming an attractive way for early detection of wildfire using the Internet of Things (IoT). The results from this research investigate to explore how the environmental parameters could be integrated into sensor based systems that have been implemented in an IoT network that will assist in the detection of signs of fire. These forest areas include nodes distributed across the vulnerable areas which may include temperature, humidity, smoke and gas sensors. Anomalies are detected when sensors pick up spikes in temperature, for example, or in the case of the presence of combustion gases and they kick off alerts to nearby monitoring stations in real time. LPWAN (Low Power Wide Area Network) protocols like LoRaWAN and 6LoWPAN are usually used in either handling data transmission over long distances or minimizing the energy consumption required in the data transmission.

You then understand the scalability of an IoT based wildfire detection system and the granular location specific data that they allow. With the deployment of a mesh of connected sensors, forest management departments can find out about potential fire outbreaks at their source before any such signs appear. It provides quick response times, allowing a small fire to be detected and have the potential to prevent a large scale disaster if way small. Nevertheless, these systems are still difficult to deploy into a remote forest area. Power supply is chief among them. It is difficult to maintain a consistent operation since many sensors are placed in regions where traditional energy sources are not available.

In order to overcome this limitation, researchers are currently investigating using energy harvesting technologies, like solar powered sensors, to extend the operational life of these networks. The self sustaining units can run months or even years without human intervention, reducing maintenance costs and a long life time of the system. The second key part of this technology is the presence of AI for data analysis. The sensor data can be given to AI algorithms to find trends, catch false positives and even to predict how a fire will develop with the current environmental conditions. At the heart of it, AI allows the raw data to become actionable data that allows the decision making processes for the fire response teams to be improved.

Technological advances in IoT sensor networks form the basis for envisioning the combination of these IoT sensor networks with UAVs and satellite imagery to produce a high level of intelligence for fire monitoring. In such a system, sensors on the ground will detect things and activate drones to go investigate suspicious areas and will provide the latest imagery that a fire exists. It consists of the precision of the IoT coupled with

the mobile and segmented geographic application of aerial technologies that provide a reliable open platform for wildfire prevention and management.

This study [5] talks about the advancements since the emergence of thermal imaging, especially when combined with infrared sensors, has become an essential early fire detector, we four have reached the stage where we cannot imagine another way. The thermal and infrared sensors are studied as a means for early-stage wildfire heat anomaly identification. In contrast to standard visual cameras, infrared sensors use the heat radiation to sense and hold a specific temperature spectrum, which is particularly helpful in low visibility (such as night time, heavy smoke, heavy fog) conditions. Drones integrate that these sensors can be mounted on drones and satellites so that they can be mounted, deployed on different terrains and different monitoring scenarios.

Thermal imaging systems work by capturing scene infrared and analysing the patterns for that of combustion. Machine learning algorithms are then trained to recognize the characteristics of the fire related heat sources and then the patterns are evaluated. This study compares the features of different detection algorithms that can generate classification between the fire intensities using thermal patterns and outlines the assignment of risk level to different zones by these systems, thus allowing emergency responders to prioritize aid. Furthermore, AI interacting with the thermal technology increases accuracy and reliability in fire detection. This allows AI to filter out non-fire heat sources (sun heated rocks, for example or operations at an industrial plant) and reduce false alarms.

Despite these advances, challenges remain. Transmission delays of data can restrict real time detection in these cases when using such satellite platforms as have fixed orbital paths and limited revisit frequencies. In addition, it is complicated to distinguish fire from other phenomena that produce heat, which is done through sophisticated filtering and contextual analysis. Future research is aimed at addressing these limitations by developing the AI models trained for the high precision of spotting thermal anomalies. These models could be carried out on edge devices like onboard drone computers for immediate decision without cloud based computing. With additional refinement of these systems, they are anticipated to be very important for autonomous, around the clock wildfire surveillance.

Machine learning in combination with satellite imagery is a major step forward in wildfire monitoring and forecasting capabilities. This study presents how such data can be processed through AI models to track fire activity over large regions across the inaccessible Earth. These satellites possess the capability to supply near global coverage and they can record thermal anomalies, vegetation health and also smoke dispersion in real time or near real time based on their orbital paths.

Convolutional neural networks (CNNs) are strongly suited for high dimensional, complex data generated by satellite sensors and the machine learning models in general. They can even identify fire hotspots, provide estimates of burn severity and forward computer models that estimate the direction and speed of fire spread. By adding other

meteorological data (e.g., wind speed, temperature, humidity) to the models, precision can be increased to a proactive instead of a reactive response to wildfire events.

The proposed Satellite based fire detection [6] is one of its notable strengths as it is scalable. They can keep watch over whole continents, which is good if there aren't sufficient ground based sensors or UAVs available to provide monitoring of their area. The study, however, does point out some limitations as well. This can leave cloud cover obscuring the potential of fire presence or delaying the detection put forward by satellite images. Another disadvantage is that data collection may lag real time monitoring by a time period.

Thus hyperspectral imaging is being used to overcome these problems. This enhances the ability to analyse and discriminate better between fire and non fire phenomena. Additionally, hybrid systems leveraging satellite imagery, data from drones and ground sensors are built to complement and not rely on this existing aspect of the monitoring network. Compared to traditional measurements, these integrated systems can provide more accuracy and timeliness of wildfire detection and make them key elements in future disaster management.

Another one of the most promising things that are being used to forecast wildlife is using artificial intelligence over recurrent neural networks (RNNs) and long short-term memory (LSTM) models to predict the movement and spread of wildfires over time. This branch of research aims to simulate fire behaviour by analysing historical fire data and discover how the patterns influence the behaviour of fire. They are especially good models because they solve the problem of how to fin in sequences and can thus use the sequence of data to model the development of fire events and predict how it will change in the future. These AI systems [7] also understand how past fire events (temperature spikes and wind shifts) and what vegetation density is in use can help predict future fire evolution, and emergency teams can use this time to strategize and act.

This modelling process requires the use of Geographical Information Systems (GIS). The artificial intelligence models that use GIS have spatial context, they incorporate the terrain features, the land cover, and the infrastructure details in their simulations. The topography, availability of fuels, and human made structures are important factors in determining fire behaviour and it is important to understand the geography of a fire prone area. Through this research, other modelling techniques, namely, cellular automata and agent based models are also explored for application to the study of fire as a spread from cell to cell according to given rules, or individual fire fronts or response agent behaviours respectively.

Another key input for these types of forecasts is the real time weather data, and the data of wind speed, humidity and air temperature. This allows researchers to simulate how a fire will react under changing conditions by feeding in the current conditions of the atmosphere. One of the major challenges still in this field is that both the good quality, labeled datasets for training AI are in shortage. Often installing anything in the AWS area is costly as it is a virtual server where you do not have root access and you can

only add your instances. If you deal with fire datasets, they are usually fragmented, inconsistent, or missing variables that obscure any predictions.

Now, looking forward, researchers [8] are bridging these gaps by combining multiple data sources, including satellite imagery, UAV surveillance video, and ground sensor data. Such a multi-modal dataset can be used to enable AI systems to learn more complex relationship and provide more accurate predictions. Our goal is to design integrated forecasting systems for estimating fire spread speeds in both directions and for advising evacuation plans and optimizing resource assignment during the active fire event.

Aside from fire detection, drones have played a crucial role in disaster response beyond, in the assessment of post-wildfire and recovery planning. This line of work describes how UAVs are being used to fly over burned landscapes for aerial surveying, for structural assessment and monitoring of vegetation regeneration in burned habitats. DC forwards that by flying over disaster zones, the drones will be able to take high resolution imagery that would have been hard or even impossible for humans to gain. The first part of the process involves these visuals being then analysed using AI powered image recognition to classify burned areas, estimate biomass loss or regions at risk of erosion or secondary fires.

In these post fire applications, different drone platforms are used to serve different purposes. In general fixed wing UAVs are great for extensive mapping as their flight path could be much farther and their speed could be much faster, whereas multirotor drones offer maneuverability and can be used for detailed inspections over small areas. However, this technology still faces operational limitations. Flight time is limited by short battery life, communication with disaster zones fails when data is supposed to be transmitted in real time. Additionally, other regulatory restrictions can restrict drone flight over certain areas where airspace is shared with manned aircraft during emergency operations in progress.

Yet AI driven automation has been working. These days drones can basically fly without human intervention, using pre programmed flight paths, and using onboard AI to identify features of interest autonomously. This not only makes collection of the data faster but also makes it reliable in post disaster analysis. Additionally, some researchers are considering the use of blockchain technology to safeguard the transmission of sensitive drone collected data that is particularly important to maintain data integrity and authenticity in the case of emergency situations.

The integration of drone collected data with other technologies like ground sensors, satellite imagery and GIS systems will provide a complete disaster response model from its perspective. The intention in this paper [9] is to rebuild not just in the immediate time, but to also strengthen the land management and fire prevention strategies of the long term. The same is true in the other area where AI has been applied in the analysis of video streams for detecting smoke and fire in real time. In this research deep learning models (mainly CNNs) have been used to predict the visual signs of fire in dynamic

environment. In particular, the detection target of smoke is a challenging problem because of its variable shapes, colors and motion. To identify such patterns, ResNet and VGG type AI models have been trained to an impressive degree of accuracy, in clutter or varying backgrounds.

The performance of several deep learning architectures towards recognizing fire related features in video inputs has been studied. Finally we evaluate these models in terms of their accuracy, speed as well as their capability to generalize in different environmental conditions. Another main major finding is how transfer learning works – pre trained models applied to the fire detection task with small dataset. It greatly reduces the training time while still achieving very high accuracy.

The result of this research is focused also on edge computing. After deploying AI models directly on edge computers (like drones and smart cameras) instead of sending video feeds to centralized servers for analysis, which can cause delays, there is no need for feedback trust. And in situations where every second counts, this means that suspicions can be detected and responded to immediately. But these edge devices are usually quite limited in terms of processing power and memory, so lightweight AI models are a critical part of the development. Furthermore, as researchers optimize models for running efficiently on low power hardware through pruning and quantization, so can they reduce energy and thereby benefit the portion of commercial applications listed above.

As the study concludes, AI is a faster, more reliable alternative for video analysis than traditional rule based systems, familiar to many of those that have dealt with them in the past, that can often take hours to process and are prone to false positives. In moving forwards, we concentrate on making these models more accurate, being able to adapt well to other scenarios and deploying them on a variety of devices. With this technology progressing, it is to be expected that it will become a standard feature of intelligent fire surveillance systems.

This research paper's [10] main asset of airborne wildfire risk assessment and forest management is LiDAR (Light Detection and Ranging) technology. The use of LiDAR to create very detailed 3D models of forest environments is explored in this paper in order to identify potential high risk fire areas. LiDAR images are unlike conventional images taken with cameras, especially mountains and forest, by using laser pulses to measure the distance to objects and then giving you precise topographical maps and forest canopy structure. It is this data that allows researchers to find out what the density of vegetation is, the height of trees, and where fuel loads are — all of these factors play a very big part in predicting the likelihood of a wildfire, and front of how intense and destructive it will be.

With this 3D data, scientists can see that dry biomass is crammed in these kinds of areas that are more likely to ignite and spread fire very quickly. Now, there is a layer of intelligence added to the integration of LiDAR analysis with AI. Using machine learning models, machine classification of LiDAR data is possible to classify a variety

of interesting tree species, such as fuel moisture content, and even predict how a fire may act in a given terrain. These sort of insights are not just essential for preventive fire management but also in deciding how to conduct firefighting activities in the middle of active emergencies.

From a key advantage perspective, the presence of one of the biggest advantage of LiDAR is its share of fine scale terrain features that are often overlooked by satellite or aerial imagery. Combined with other remote sensing technologies especially multyspectral imaging, LiDAR delivers a means to a comprehensive view of both horizontal and vertical forest structures. But a limitation of all the technology is that it comes with some. Cost of LiDAR Sensors are high and there is high demand of data for processing and analysis which use huge quality of computational resources. Limited to its widespread adoption has been its adoption given the resource constraints of a place. LiDAR systems became the subject of future research in order to improve their accessibility and efficiency. It involves the development of real time data processing techniques able to cope with large datasets at the fly and development of cost effective LiDAR platform adhering to operate in UAVs. Using LiDAR in combination with AI, researchers seek to construct dynamic fire hazard mapping systems capable of updating in real time, and delivering a powerful tool for both pre and post fire emergency planning to forest managers.

The other pivotal technology in the struggle against wildfires is Wireless sensor networks (WSNs). In this research [11] the deployment of sensor networks of temperature, humidity gas and flame sensors are investigated across forests to provide earliest signs of fire detection. The collection of environmental data and its transmission to central monitoring stations is continuously done by these nodes using communication protocols like Zigbee and LoRa. In particular, LoRa is chosen for its long range and low power capabilities, making it a good choice to deploy in remote forested areas over which connection and power are difficult to reach.

WSNs have two important strength; they can provide circuit real-time and local data for the purpose of generating early warnings. Forest management agencies are establishing a network of these nodes that allow them to detect changes in environmental conditions as a signal of the onset of a wildfire. This system relies on AI based data analytics to filter out noises, remove false alarms and select alerts on the basis of the severity pattern of the data. They can also calculate trends over time allowing them to spot a harmless variation from actual fire risks.

WSNs have their own set of challenges with regards to deployment and maintenance. Hard environmental conditions lead to sensor node failures, and energy and communication bottleneck can limit the network scalability. Hence, such researchers are brain storming on the hybrid sensor networks, which may comprise UAV based surveillance in tandem with WSNs. Here, drones can fill a role of flying over areas of the grid where fire anomalies have been identified to help confirm or deny the existence of fire.

Future work will focus on using AI powered anomaly detection models where they can learn to work under different environmental conditions and sensor behaviours. The deployment strategies of these models will be more intelligent so that resources can target areas with the highest risk. WSNs are expected to become more robust, adaptive, and embryonic to the much larger ecosystem of wildfire detection.

The autonomous drone sweep is an interesting frontier in the detection and suppression of wildfire. In this paper [12], the paper revolves around the question of how groups of drones, coordinated by artificial intelligence based algorithms, can perform better firefighting tasks than individual UAVs. They are alive and operate in real time on the task, dividing tasks, adapting to changing environments, responding to emerging threats forms. Swarm drones are different from typical firefighting approaches that depend on central coordination and the decisions of leader firefighters. Instead, it alludes to decentralized algorithms whereby individual drones can independently decide locally based strategies to collaborate toward a common group mission.

It then explores different swarm intelligence techniques including such as particle swarm optimization and reinforcement learning, for optimizing drone behaviour. And these methods allow drones to change flight paths, split resources such as water or fire retardants, and keep themselves safe from obstacles, other drones, etc. With these algorithms that imitate the natural movements (as birds or insects for example), a resilient and flexible response system is created to respond to the unexpected fire conditions.

A practical application of this technology is to drive water or fire retardant to hotspots using squadrons of drones. Each of these drones in the swarm may have a small payload and can use real time heat maps and fire behaviour prediction to target particular areas. Targeted suppression is possible here without wasting resources, enhancing the efficiency in the fighting effort. Unfortunately, many challenges must be overcome in order for this vision to be a complete reality. In particular, they consist of the communication latency between drones, limited energy which limits the flight duration, and the complexity of making real time decisions in changing settings.

With that, researchers are looking to create cloud based AI systems that operate the swarm via global coordination in order to overcome these barriers. They can handle large volumes of data, include execution at the local level, and provide strategic oversight over these systems, while individual drones execute at the local level. Together these clouds and edges will develop the hybridization offers a strong combination of leverage to next generation firefighting operations.

This technology is maturing, and so we can look forward to a new way of thinking about the fire suppression response, namely, using drone swarms. Autonomous swarms have the potential to provide a faster, safer, and more adaptive alternative to traditional methods, from rapid response to continuous monitoring to post fire analysis. As a newly developing means of wildfire management, they are quickly becoming a promising tool due to their scalability, flexibility and ability to operate in parallel as a system.

This study [13] analyses the One game changer that has emerged in wildfire prevention is the development of wildfire prevention and control by the late technology of AI early warning systems. In the process, this research drills into how machines are being tapped to hazardously forecast fire risk based on historical fire records, changing weather patterns, and high-resolution satellite imagery. These systems also learn from past events and predict not just the probability of ignition, but would also indicate the probability of reportable and potential fire severity and trajectory. This information is vital in order to issue alerts in advance and react in an effective way by emergency services.

The main innovation explored in the study is the combination of cloud computing and edge AI. However, massive datasets from satellites, sensors, and weather stations can be processed over the Cloud, or at the edge over devices like drones or field sensors that bring intelligence close to the data. This dual architecture enables near instantaneous detection and decision making that is key in the rapidly changing conditions of the wildfires.

The research highlights [14] that interdisciplinary collaboration, that is, combining data science, expertise in environmental studies, meteorology and emergency management, is needed in order to develop robust early warning frameworks. There are however challenges to implementing these systems. There are concerns over data privacy when trying to collect environmental and geolocation data on a large scale, especially from the private lands. And another barrier can be the amount of work that needs to be done to interpret an AI model, especially a deep learning network that tends to be cloaked in blackness. In addition, such accurate and reliable models are created by access to large, high quality training data which may not always be readily available.

To address these challenges, the study suggests using federated learning, an emerging technique that trains AI models on remote, decentralized devices or servers, while keeping data private. The continuous improvement of predictive models is not only a result of improved security, but also provides an obvious opportunity for data leakage. Overall, the findings indicate that a combination of early warning systems based on AI can decisively enhance emergency response times, save life and property from the damage of wildfire and improve the community preparedness for wildfire events. Unwittingly, they prove to be indispensable tools in wildfire detection, monitoring and also management. It is with this background that this paper presents a comprehensive review of the current capabilities of UAV based systems that use onboard sensors, fire detection algorithm, and advancer communication systems to navigate and monitor even the biggest wildfire prone area. The authors show us how UAVs are complemented by unmanned ground vehicles (UGVs) to make both aerial and terrestrial views of the fire dynamics.

UAVs offer several distinct advantages. These are flexible systems that can be deployed quickly to remote or hard to reach locations and their relatively low costs means that they may be appropriate as a general option. Thermal and optical cameras are used by

them to detect fires in their early stages (even through thick smoke or in low light conditions). Furthermore, there are a number of values that UAVs can gather such as temperature, wind direction, and vegetation dryness that are all needed in order to predict fire behaviour.

Artificial intelligence and computer vision has only advanced further UAV capabilities in recent years. By allowing the drones to autonomously find fire hotspots, map the areas around them and speak directly with command centers, they are now able to do all of these things. But these innovations also introduce new issues. Flight duration is limited by the supply of power, such that time between flights allows only limited opportunity to monitor an area continuously. The lack of available annotated datasets pertaining to wildfire imagery also limits the development of more accurate detection models. Additionally, most of the present day AI systems still need solid real world test for their reliability under dynamic fire conditions.

To tackle these problems, the paper [15] proposes intelligent, system that collect and analyse data in real time, make autonomous decisions and coordinated other drones or ground vehicle. These systems would enable such fire response in a dramatic manner, turning UAVs into essential assets in fighting as well as prevention of wildfires. Drones are not only revolutionising our ability to see and combat wildfires — they are changing how wildlife is monitored, particularly in the wild. In this research, the use of UAVs with high resolution and thermal camera to monitor animal movement as they are tracked during wildfires, so as to protect vulnerable species in fire affected zones are highlighted. Drones are available in emergency instances where time is of the essence and offer quick access to areas that animals may be trapped or even on the run that might otherwise be unreachable.

The use of drones as ground based monitors bring advantages over traditional ground based methods as they provide faster deployment, broader area coverage as well as more accurate data collection compared with traditional ground based methods. Wildlife agencies can use them to find animals and even to rescue species at risk, because they can detect animal heat signatures even through smoke. In particular, it is of value in terms of understanding animal displacement and stress on their return within evacuation operations or post fire habitat assessment where this knowledge is crucial for conservation planning.

The approach taken in the study does not back away from the limitations of the approach itself. Drones' sensors can be more inaccurate during high ambient temperatures, low visibility and in conditions of interference by the smoke and ash. Flight time is also limited by battery, which can be crucial for big monitor operations. Additionally, standard detection algorithms are not necessarily capable of identifying animal movement rather than other thermal sources. To ameliorate these problems, the authors propose the creation of bespoke software and AI algorithms designed for fire-prone environments in which wildlife are detected. Specific patterns of both animal shapes and animal movement would be learned by these systems, providing reduced

false positives since these systems would be able to recognize specific animal shapes and animal movement patterns. Other approaches to leverage ongoing work on drone data, satellite imagery and ground camera trap studies are being proposed to help complete fire a picture of when the different pieces of the wildlife activity puzzle are observed during and following a fire. Ultimately, this research reaffirms the notion that drones, in conjunction with smart algorithms, are not just objects used to safeguard human lives from rising wildfire threats — they are indeed mighty allies to conservation of biodiversity.

Drones have gained more importance lately in the role of wildfire detection in using recent advances in computer vision and using deep learning models. In this article, the authors present a decade long overview of the software and hardware development techniques influencing drone based fire detection systems. In particular, it mainly discusses the application of convolutional neural networks (CNNs), You Only Look Once (YOLO), and Faster R-CNN, which have significantly increased the accuracy of fire detection in aerial imagery. As it turns out these deep learning models have the ability to identify fire related patterns like flames, smoke and heat distortion with a degree of precision, speed the traditional rule based could never attain.

Authors point out that these applications are now being carried out on multicopter drones. One reason they are so ideal for monitoring certain fire prone zones is because of their ability to hover, to move around, through difficult terrain. The drones usually come in the form of RGB and infrared cameras, so they have both visual and thermal ability to look at the environment. In clear weather and under daylight, RGB sensors are best, while infrared sensors work very well in smoke blocked conditions or in complete absence of light.

However, there are still several challenges that the study acknowledges. With onboard drone systems having limited processing capabilities and battery life, high performance AI models require large amount of computing resources. In addition, accurate detection models are extremely dependent on the presence of large well labelled datasets for wildfire specific imagery and these do not exist. Flight duration is also limited, and surveillance time is further restricted as necessary due to limited surveillance time since both require rapid coverage or frequent recharge.

To improve on these, the researchers are looking into deploying more energy efficient neural networks in view of the fact that their drones do not need to have access to the cloud infrastructure to process data in real time. This transition to onboard intelligence means that drones can spot fires and transmit warnings right away, something which makes the system practical and more scalable for the wild forest scenario. The study demonstrates that if drones are to be used extensively in fire prevention efforts, their computational capabilities will need to be improved, and lighter and faster AI model will have to be developed.

This research [17] presents a smart drone based monitoring system for detecting human presence in dense forest areas, such as those are typically difficult to access and monitor.

The system uses such lightweight object detection algorithms which include MobileNet as well as Single Shot Detector (SSD) to identify individuals as well as vehicles involved in illegal logging and other harmful ventures. A real time processing of aerial imagery and issuing alerts to remote monitoring centres whenever there is suspicious movement or fire prone activity is done in these algorithms.

It is a strength of this system that it can work itself. Meant to function with high resolution cameras and onboard processors, the drones can handle the footage on the fly without the constant human oversight. Advanced image processing, communication systems, control modules and a technical framework, including all the means to ensure reliable operation in rugged and forested environments, are the basis of the technical framework.

As the study shows, the deep learning models significantly surpass traditional image analysis methods on accuracy to speed. Response times are better, as they are better poised to distinguish between humans, animals and objects and lower false alarms. In conservation efforts, such efficiency is necessary, since time sensitive illegal activity detection is needed to prevent irreversible environmental damage.

The paper also has some caveats, although the benefits are in plain view. In some cases, extreme weather or heavy smoke can affect drones since smoke can obscure vision and disrupt communication. The surveillance duration is also limited by battery constraints. However, intelligent forest monitoring by means of an AI and UAVs is a cost effective and scalable option. The research ends by considering that as drone tech and AI algorithms become ever more powerful, they will become more essential to conservation and fire prevention.

This article [18] demonstrates that the pair of deep learning and drone based imaging is a powerful tool in detecting early stage wildfires. This speaks to the necessity of fusion in UAVs carrying both infrared and thermal sensors to detect smoke, and flames in real time. This development rested on these very advanced algorithms like CNN, YOLO, Faster R-CNN, DeepLab. These all provide special abilities: for generic image classification, CNN works well, YOLO is excellent at real-time object detection, and DeepLab is utilized for semantics segmentation to interpret the design of the whole scene.

The paper systematically contrasts these AI techniques and their strengths and weaknesses. They are quite accurate under controlled conditions, however, the models are demanding in terms of computation resources and require large amounts of training data. Performance can also succumb in such complex situations where there is dense smoke, poor lighting, etc. It reveals that nighttime detection and data scarcity are biggest challenges to be overcome to increase reliability in real world forest settings.

The authors propose sensor fusion, taking input from multiple cameras and sensors to provide greater and more accurate picture of environmental conditions for the benefit of improved system performance. Arguably, this processing of data onboard allows

drones to detect and respond to fires without delays caused by transmitting data to a centralized server. By doing this, not only do we speed up the detection but in fact, it also makes the system more resilient against communication disruptions.

Although these technologies have great promise, the study states that they still need to be further developed. This work would also address improving the detection accuracy in different environmental conditions, expanding the available labelled datasets for training the model, as well as the optimization of the algorithms for the low power devices. However, despite the work, the research presents a bright prospect for the future; the future where intelligent drone surveillance against the begin of wildfires out and respond quicker and with greater efficiency to save those life and ecosystems.

The proposed deep learning and computer vision technologies are enhanced for safety in maritime environments through an advanced shipboard fire detection system presented in this study [19]. The model in the system is a YOLOv7 based model paired with an improved backbone architecture E-ELAN and advanced feature fusion methods. With these changes, the model can now detect the fire instances in real time in adverse conditions, where lighting (or lack thereof) is low or the visual background is complex in nature (common on ships).

I trained the model on more than 4,600 annotated images for fire and non-fire scenarios. The ability of the system to learn to distinguish small fires even from visually similar objects, like reflections or bright surfaces, with such a degree of precision was aided by having this diversity in the training data. With respect to the system's performance, the model achieved a 93% detection accuracy, 94% precision and strong recall and F1 scores indicating the system's robustness and reliability. The validation of these metrics shows that the model has high capacity to minimize both missed detections and false positives.

The system occasionally misclassified fire like visual artifacts, such that it could better be correct if the dataset is expanded and diversified. Such a model can generalize better in complex visual environments if the training samples would be more varied. In all, the research shows that deploying AI based, vision-based fire detection on maritime vessels is both possible and powerful. As shipboard spaces become increasingly enclosed and more high risk, this technology provides a practical, economical method of providing early fire detection and prevention in such spaces at sea.

The dual-purpose fire detection technique described in this research [20] uses the combination of image preprocessing technique and deep learning to target flames and smoke. First, the authors use the HSV color space transformation to filter out regions that are irrelevant to the flame, and then use the Harris corner detection to find flames in the input image. The system also takes advantage of the optical flow analysis together with the dark channel prior method to augment the recognition of smoky regions for smoke detection. In this preprocessing step, potential fire related image sections are isolated and further analysis can be focused on more relevant image sections.

After these candidate regions are identified, the system classifies them by using the CNN based on the Inception-V3. We train this deep learning model to predict the input flame, smoke, or otherwise with high accuracy. Results were excellent for the flame detection (96% in accuracy), and even exceeding that for smoke (93%), as the proposed approach demonstrated. It also outperformed other well-known models such as SSD and Faster R-CNN in their ability to reduce false positives, and performed better in comparative tests.

The standout feature of this system, however, is that it can run in real time; it is a feature suited for the high stakes environment where it can be used to prevent disaster early on. The system integrates classic image processing techniques and modern deep learning to find a correct tension between accuracy and computational efficiency. The study also notes that hybrid method reduces false alarms and fairly reliable performance in visual noisy scenes. As such, it is promising to be deployed in public and industrial settings where early fire detection is important, and reliable.

In this thesis [21], a cost-efficient early fire detection system is presented having the capability of reliable performance in large or open spaces using WSNs accompanied with machine vision. The system is based on Raspberry Pi hardware and low-cost cameras streaming live video to a central monitoring station over wireless networks. Then, this is a solution which is cheaper than expensive IP cameras or conventional flame detectors, but at the same time respects successful detection capabilities.

Real-time image analysis on the video feed is performed in HSI (Hue, Saturation, Intensity) color space so as to recognize fire colored region in the video feed and initiate the fire detection process. An OpenCV/C++ algorithm is developed and processed each frame by segmenting fire pixels, applying morphological manipulations and drawing contours around potential fire zones. Thus this approach enables the system to localize and highlight of the fire areas dynamically.

It was found that the video stream offered the reliability of about a 20 – 30 second delay, which is enough to monitor an application, and the detection algorithm worked well at recognizing fire patterns. While the system occasionally caught fire colored objects leading to false positives. The authors recognize this limitation and suggest that implementation of machine learning-based fire classifier and automatic fire alarm will bring further opportunities for improvement.

In general, this low-cost design makes it a very good option for deployment in parks, warehouses, agricultural lands, and other large expanse outdoor spaces that these traditional systems cannot support financially. They expand on how such systems can be broadly integrated with current CCTV infrastructure and should be scalable and affordable, considering they would support smarter and automated fire response systems.

This research [22] task introduces a strong, video-based system to detect initial fire that addresses many challenges with the real world of visual fire monitoring. The system

uses a layered detection method that combines background subtraction, fire-colored pixel classification, dynamic texture and spatiotemporal Wavelet analysis. By drawing attention to potential fire areas, the system reduces calculation load while maintaining high accuracy. An important innovation is its integration where humans detect and track models are detected, which is to filter false positivity due to fire-coloured objects. It also includes GPS-based learning systems, which allows a geo-lined fire warning, to map the image coordinates in places in the real world. The real-world video was tested on a diverse dataset, the system achieved 92.7% success rates, and detected exactly the fire in different light situations without misleading the same heather. Authors emphasize the practical for the system for the field transition, and aim for future improvement in real-time performance on smart camera patterns with low cost.

The research thesis [23] presents a vision-based early fire detection system, a vision-based early fire detection system that identifies both flames and smoke using real-time video processing. The system is designed for use in indoor and outdoor environment and does not depend on offline training. Instead, it uses motion detection, color filtration, turbulence analysis and optical current to separate fire-related areas. It treats parallel, speed and accuracy, parallel flame and smoking streams. Through extensive testing of 30 real and distributor video sequences, Quickblaz demonstrated an increase in speed 2.66x compared to commercial software such as faster response time, better fire location and visple. It also performed better at night and low light, which proved to be a reliable and skilled tool for real-time fire monitoring.

The proposed fire detection system in this paper [24] is an overall robust and refined system using fixed surveillance cameras, and advanced data visualization techniques to correctly locate fire events with few false alarms. This methodology is based on the observation that standard fire detection mechanisms tend to yield high false positive rates due to false distinctions that cannot, in actual fire, differentiate actual fire from fire like visual distractions. To solve this, the proposed system is composed of a collection of carefully selected modules that together analyze both the static and dynamic characteristics of a fire. The approach is based on background subtraction algorithm customized for isolating foreground elements and then dynamic texture and frequency analysis. These modules are specially built to respond specifically to the unpredictable movement and flickering way that matters of the fire behave, as opposed to color or shape. The system is able to discriminate between real flames and deceptive visual disturbances including sunlight reflections, or waving flags and vehicle lights, by including behavior-based detection.

A novel fire-specific color model is one of the key innovations of this work which is coupled with a Wavelet based frequency signature. Most conventional systems use standard RGB or HSV color spaces, but the custom model in this design is fine tuned to capture the particular chromatic properties of flames. By wavelet-based frequency analysis, temporal flickering frequency of fire is captured, a property that is usually neglected in simple frame by frame analysis. Additionally, this system includes a complex object tracking and recognition sub-system that continuously tracks the

development of the moving flame elements. This verifies the fire presence and not only by color or motion at a single frame, but by tracking flame-like behavior in multiple frames. Consequently, the distinguishing of fire from fire-colored objects in motion (such as people wearing bright orange clothing or illuminated signs) allows for increased accuracy and robustness in such an environment.

The system also includes a GPS based calibration mechanism to improve situational awareness as well as localization precision in the detected fires. It is suitable for real world deployment in public spaces or critical infrastructure zones since its geo tagging of fire incidents allows surveillance network to watch over it. On various real world video datasets, diverse lighting, environmental, and fire scenarios, the experimental validation was performed. But its object tracking module was exact to 92.8% and the system achieved a fire detection accuracy of 93.1% with a low false positive rate. The system is designed to operate at 10 Hz processing rate which strikes a practical compromise between computational efficiency and detection accuracy for real time applications. In general, this integrated framework provides a complete, real world ready solution for the fire monitoring that incorporates visual analytics, temporal dynamics and geospatial intelligence in a single cohesive pipeline.

This research paper [25] gives SmokerBeacon, an innovative actual-time smoking detection device that mixes computer imaginative and prescient and gas sensing for enforcement in no-smoking zones. The device uses a hybrid technique, integrating YOLO-primarily based object detection with MQ135 fuel sensors to confirm smoking pastime visually and chemically. SmokerBeacon intelligently categorizes eventualities—while most effective visible, handiest chemical, or each indicator are detected—to trouble suitable signals and decrease fake positives. Trained on a custom dataset the use of YOLOv9, the model done superior overall performance in precision and don’t forget in comparison to YOLOv8 and YOLOv10. The gadget’s hardware, built across the ESP32 microcontroller, permits for seamless real-time tracking and alerting the usage of LEDs and buzzers. By combining IoT, image recognition, and sensor fusion, SmokerBeacon proves to be a dependable, low-fee answer for improving public fitness and protection in restrained environments.

2.2 RESEARCH GAP

Recently there has been great progress in the area of wildfire detection in the last decade; however, the current body of work suffers from several fundamental weaknesses that prevent the scalability, efficiency, and application in the real world of these systems. Many current wildfire detection methods are in general based on satellite based surveillance or ground level sensing and reporting systems. Satellite-based approaches are commonplace because they are essentially able to cover a wide geographical area, however, they are limited in terms of available temporal resolution, traditionally slow latency and susceptibility to weather based obstructions (such as cloud cover and haze). This severely compromises their ability to detect ignition or detection events at their full, most vital stage, in their infancy. Also, there are often

constraints in using ground based methods due to their static placement, low field of view, overheads of maintenance, and sheer impracticality of setting such systems in large forest expanses. Most of these methods also heavily depend on human intervention: manual intervention or human visual confirmation, which will inevitably lead to response delays and further endanger first responders. As a result, such an automated, real time, aerially mobile detection system is not only needed, but critical.

Additionally, although computer vision and deep learning techniques such as YOLOv4, YOLOv5, YOLOv8, SSD and Faster R-CNN can increase speed and accuracy of fire and smoke detection, these models have very limited applicability to the environment – they are trained on curated datasets with minimal environmental noise, well lit conditions and fixed camera angles. Thus, their performance is subpar in such dynamical real world settings, e.g. motion blur, changing illumination, occlusions, unreliable camera for drones, such as camera shaking, or natural situations such as fog, shadows, or background flames. These models are also even more critically typically optimized for single frame detection and without temporal awareness, which ignores highly relevant patterns based on motion of transient events (such as a puff of dust vs. real smoke). However, it is common for there to be little work integrating temporal modelling of an architecture embodied by LSTMs or 3D CNNs within the scope of wildfire progression monitoring. Many of the models fail to understand the temporal nature of wildfires, ignoring minor clues that would help in early intervention.

In addition, little work has been done utilizing thermal imaging and temperature based anomaly detection. Typically, before smoke becomes visible and flames are seen, wild fires produce localized heat surges. Despite all of these, very few studies integrate thermal cameras or infrared sensors into the early detection pipeline. These thermal anomaly detection techniques have not yet been investigated to sufficient extent in the context of wildfire. This missed opportunity is a good introduction for thermal cues integration as it could help early detection, at night or in low visibility conditions for example, where RGB models can limit to detect. Similarly, concerning environmental sensors such as the MQ-135 for sensing the presence of oxidizing gases like NH_3, NOx, benzene, CO_2, they provide useful information on tracing the anomaly in air quality resulting from combustion which is rarely fused with the computer vision outputs in a coherent, decision making system. To the authors' knowledge, most research using such sensors employs them in isolation from drone platforms and real time visual feed synch.

Acoustic intelligence is also a profoundly under studied component in wildfire detection research, equally. In the real emergency cases, human victims may receive their call, shout or make audible noises that can be made detectable if the correct audio sensing modules were available. But wind, rotor blades (in the case of drones) or animal sounds, as well as environmental interference even make voice detection impossible. While noise suppression and audio signal enhancement techniques such as RNNoise, Spectral Subtraction or Wave-U-Net are becoming more powerful with time, almost no wildfire detection framework has tried to incorporate real time audio denoising, human voice recognition in this pipeline. There definitely is a lot of potential in the human

presence detection through voice, transcribing it into actionable alerts that have been largely left untapped. In post fire scenarios such as search and rescue teams, the systems capable of identifying trapped individuals by sound can help even more when this gap exists. Furthermore, literature has not demonstrated this pipeline's automation (specifically, triggered email notifications with both sensor and transcription data when the data is denoised) at any substantial depth.

From this gap, we add another layer, by the absence of centralized, cloud connected and real-time reporting mechanisms. There has been much work on enabling indoor wildfire detection through vision, sensing or voice, but with little to no attempt to fuse all modes of detection—fusing vision, sensing, and voice—patching together alerts of fire, pushing them towards remote dashboards, emergency personnel or authorities in real time. Few of these cloud IoT platforms (e.g., ThingSpeak) are used in combination with visual deep learning models. In addition to that, existing drone based implementations somewhat regularly use post processing of captured footage, as opposed to edge or near real time inference on onboard computing. The true challenge and opportunity however are designing a UAV deployable, multi modal framework for real time object detection, temperature based anomaly mapping, environmental sensing, audio analysis, all in conjunction with each other to perform in a coordinated low latency system.

Thus, the research gap here is not only in the absence of individual components, but on the inexistence of integration, automation, and the being of real-time in all modalities. To the best of our knowledge, there is no such solution in the literature, i.e., an object detection using YOLOv8 based, with the help of efficientnetv2b0 for image enhancement, LSTM based temporal modelling, thermal anomaly detection, ESP8266 and MQ-135 for environmental gas sensing, human voice detection and speech transcription, a cloud connected email/Thinkspeak based alert system all combined into a drone enabled deployment model. This project adds novelty to such a multi dimensional gap, would have sought addressing this multi dimensional gap using the state of the art deep learning, IoT sensing, and audio intelligence, and drone mobility combined into a single unified wildfire detection and emergency alert platform. Importantly, this architecture demonstrates an earlier and more accurate fire detection which also directly facilitates human rescue operations as well as reducing emergency services response time (as the time to reach firefighters on scene), and ultimately saving human lives and ecological systems from devastating wildfire damage.

2.3 SUMMARY OF LITERATURE REVIEW

Surveyed were the vast amounts of literature that show a rapidly changing landscape in the field of fire detection and management where the old tradition approaches are routinely being replaced by new more intelligent, automated and data related approaches. Significant improvements in the accuracy, time and reliability of the fire detection have been made as a result of the move away from conventional smoke detectors and manual surveillance to more sophisticated, AI, deep learning, sensor fusion, UAV (unmanned aerial vehicles) driven systems. Technologically impressive

and strategically vital, these innovations have grown from necessity – wildfires are globally becoming more frequent, more destructive and more severe with urban expansion into wildland area increasing the problem.

One of the highlighted trends throughout the literature and identified as one of the most transformational is the use of AI and deep learning models; specifically of Convolutional Neural Networks (CNNs), You Only Look Once (YOLO) detectors, Faster R-CNN or Recurrent Neural Networks (RNNs) in order for the system to recognize flames and smoke in real time. Computer vision fire detection algorithms have changed the course of computer vision through their redefined capabilities. Early stage identification of fire elements is based on their capability to process both still images and live video feeds with unprecedented precision. It was shown by several studies that training these models on various datasets, including variations in lighting, background, and other environmental conditions, has improved a model's robustness tremendously. System based on YOLOv7 and Inception-V3 achieves up to 93–96% detection accuracy in practical applications, which justifies their deployment in the real world mission critical environment.

The literature adds more strength to the advantages of UAVs in fire monitoring and management. Drones are especially adapted to large and inaccessible areas, rapid deployment, and the survey potential of hazardous areas. UAVs equipped with both RGB and thermal cameras enable them to detect both fire visual and heat signatures in smoke obscured or darkness. Besides gathering data in the form of surveillance, UAVs can be integrated with onboard smart AI systems which can be used to make smart decisions as well. Other studies looked at the coordinated use of UAV swarms together with the ability of the swarms to autonomously partition tasks such as perimeter mapping, fire source localization, and real-time tracking of flame spread. The approach to this firefighting is a promising direction in future firefighting systems based on multi agent intelligence which is motivated by reinforcement learning and swarm optimization algorithms.

Often, wireless sensor networks (WSNs) and IoT based monitoring infrastructures were often highlighted in parallel to aerial technologies' deployments. The ground based systems will have question distributed node with environmental sensors to measure variables like temperature, humidity, gas concentration and airborne particulates. They allow on going, real time monitoring for those conditions which facilitate the initiation of fire. Zigbee, LoRaWAN and 6LoWPAN for example are normally used to send data from areas in forest to the central control system, asynchronous one way data transmission is used. Through several studies, it has been assumed that WSNs can complement UAVs in a hybrid approach whereby UAVs are sent aerial reconnaissance or additional localized data collection upon receiving alerts from the WSNs. Additionally, it showed that it applied AI based anomaly detection algorithms on sensor data to reduce false positives and enhance the system capacity to predict fire outbreak caused by slight environmental change.

The other important area considered in the literature is vision based fire detection using fixed surveillance systems. Usually, these systems use video analytics to identify the characteristic visual features of fire like its flickering, color spectrum and dynamic motion. All fire-like phenomena are isolated and visual noise is filtered out by combining the use of advanced image processing techniques such as background subtraction, dynamic texture analysis, wavelet transformation, and object tracking. Accuracies these systems possess are strengthened by the development of new color models and frequency signatures unique to the flame behavior. Furthermore, GPS calibration for geolocation of the detected fire zones has been added in some of the implementations to provide spatially accurate alerts to emergency services.

A lot of studies show that combining different type of data increase the detection reliability. A number of times, mentions were made of sensor fusion (combining thermal camera, visual data, or environmental sensing input) as a way to balance the limits of disparate components. For instance thermal image can help to detect heat sources which wouldn't be viewable for optical camera in nighttime condition or in smoky condition, and AI models can use spatial and temporal information to distinguish between fire heat sources and non fire heat sources. Just as gas sensor readings were paired with video analytics in systems that were capable of both visually and chemically confirming the presence of smoke or fire and thus reducing the channel of false alarms, especially in risk-sensitive environments such as hospitals or public transport stations.

Several recurring challenges and limitations were already found in the literature. However, such lightweight, power constrained platforms can not usually support deep learning models with their high computational requirements. However, training models is highly dependent on large, annotated datasets, which are challenging and costly to generate for the scenarios of fire spread over a variety of environments, lighting levels, and weather conditions. Yet there are still many systems that are having difficulty in nighttime detection, where the effective cameras are now becoming less efficient in visible spectrum. Further, there is a key obstacle to widespread adoption due to the problem of false positives (often caused by fire colored objects, reflections, or sun glare). In remote areas, restricted connectivity, regulation over drone usage, and energy limitations of sensors along with them reduce the possibility of deployment at scale.

This has been responded with supposed forward thinking solutions. They include decentralized model training model training in federated learning method without violating privacy of data, designing of energy efficient neural network architectures for edge devices, and usage of explainable AI (XAI) to improve transparency in detection decision. With remote, autonomous operations particularly promising when the systems capable of operating on edge devices such as Raspberry Pi modules or ESP32 microcontrollers are used. As far as system scalability is concerned, researchers are looking into how the coordination between multiple AI models and data resources in cloud based platforms can be handled and detection can remain synchronized and responsive across regions.

Moreover, the literature to some extent shows a strong trend in interdisciplinary collaboration. Currently, effective fire detection systems call for the total range of knowledge from fields such as computer vision, meteorology, wireless communication, environment science, and disaster management. It is repeated that real world validation and deployment studies are also needed. For example, most models and systems do very well under controlled or simulated conditions but then can only work successfully if they can be operated in variable field conditions, unpredictable weather, or if there are obstacles in the line of sight such as dense forests or a cityscape. As a result, future work will weigh in on a more robust, adaptive, and long-term operationally resilient basis.

The insights gained from the literature taken as a whole constitute a very strong starting point from which to start designing and developing the system developed in this study. From this, it is obvious that fire detection efficiency cannot be a work of a single technology or a model, but rather the integration of an intelligent framework capable of fusing different kinds of data types in multiple contexts. The methodology described in the next chapter is based on this understanding. It presents a hybrid system for detecting fires in the visual, analyzing real time sensor data for environmental monitoring, and getting situated awareness via GPS enabled geolocation. In addition, the system is very energy efficient, real time responsive and deployment is of low cost that is comparable with both high risk rural zones and urban infrastructure. The proposed approach seeks to make an orderly, scalable, practical and forward thinking contribution to the field of intelligent fire detection and management by addressing the limitations reported in the existing literature and leveraging upon their most effective strategies.

CHAPTER 3

METHODOLOGY

3.1 IMAGE PROCESSING AND COMPUTER VISION TECHNIQUES

3.1.1 DATASET PREPROCESSING AND AUGMENTATION

After collecting and labeling the dataset, I made three subsets of the dataset such that it can be used for supervised learning and evaluation of the performance; namely, a training set consisting of 24,003 images, a validation set with 1,017 images, and a test set of 731 images. They created these partitions so that each set would have a balanced distribution of classes in such a manner so that the model will be trained, transferred and evaluated on representative samples of the problem space.

In order to get the input format standardized and get rid of all the inconsistences in image orientation and image size before feeding the data into the model, I built a structured preprocessing pipeline. The first step was auto orienting each image. This was critical because many of the images had metadata (camera or mobile device origin) and those images would typically be loaded by different libraries as the rotated or flipped images. This made sense because convolutional models need to learn well into their visual layout, and I corrected this orientation programmatically so everything in the sample remained consistent.

I then performed orientation correction and resized all images to have a fixed resolution of 640×640 pixels. Instead of padding or cropping, I stretched the images right to the target dimensions with the exception of an image's original aspect ratio. This is slightly distorting the image content but it allows me to keep the input size uniform yet without spilling over with black bars or cut off interesting elements of visual information when training models entirely based on spatial features.

As a part of preprocessing, I exploited extensive data augmentation techniques to enhance the model generalization and also make it more robust towards the real world variations. I generated three additional augmented versions for each of the images in the training set, therefore having fourfold more images in the training set. It was not applied at random, rather it was done in a carefully designed combination of transformations that mimic likely environmental and positional variations that a drone or surveillance system might realistically encounter.

The augmentations did include both basic and complex geometric transformations. Realistically, to account to cases where the camera could be mounted to different orientations or tilted, I applied horizontal and vertical flips, as well as rotations of 90°, 180°, and 270°. I also conducted random rotations over a range of –15° to 15°, resembling minor camera shake, tilt or manual defect. These subtel rotations serve to make the model invariant to minor differences in the input data angles.

I also implemented shearing transformations in the horizontal and vertical directions with a shear intensity limit of ±10 degrees beyond rotation and flipping. The kind of transformation that this type of transformation does is similar to perspective distorting like an object is observed from an angle instead of straight heads on — which happens a lot in aerial imagery or surveillance footage. Using all these augmentation operations resulted in adapting these in such a way that the semantic integrity of the image is not violated and the training data augmented significantly.

This preprocessing and augmentation pipeline was presented for a reason, overall: to create an environment for the model to see a wide variety of spatial twist in spatial positioning that will make the model learn generalized features and not rely on any fixed positional or alignment cue. This was particularly necessary as this system was to be used in the real world, where the input data is not likely to be perfectly aligned or consistent.

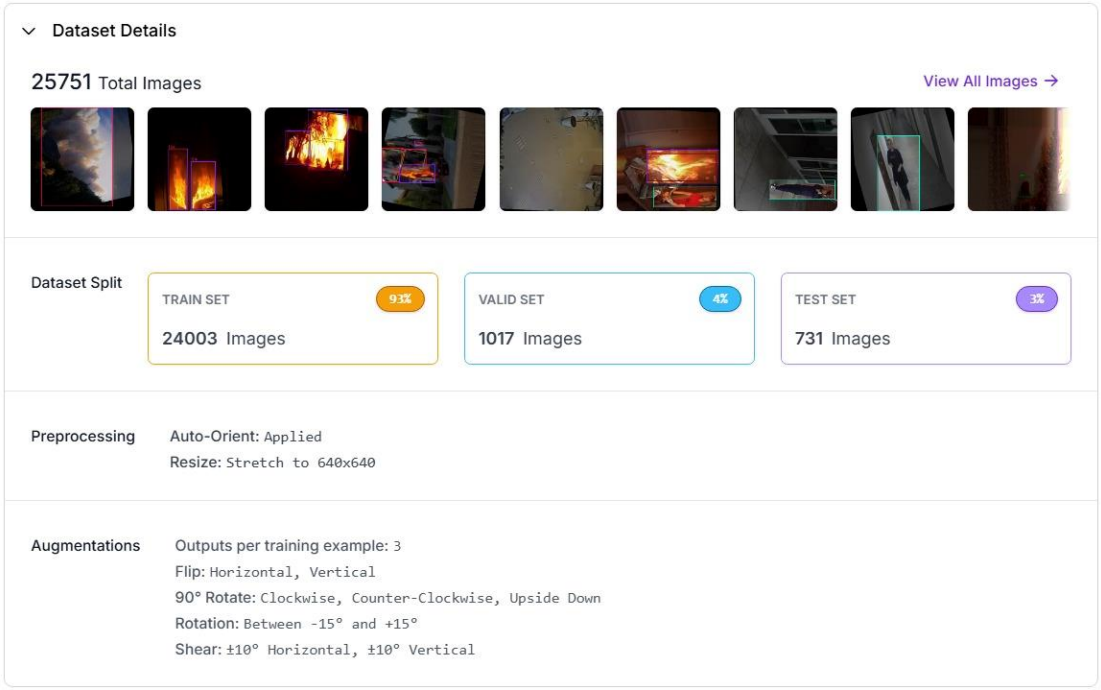

FIGURE 1: INFORMATION OF THE DATASET

3.1.2 FEATURE REPRESENTATION AND EXTRACTION

3.1.2(a) SPATIAL FEATURE REPRESENTATION

First, the original input image, typically of RGB format is converted to its corresponding grayscale image and the spatial analysis phase is initiated. This is a fatal preprocessing step of an image because this reduces the complexity of the image to map three color channels Red, Green, and Blue into a single channel which only gives us the intensity of the light at each pixel. Particularly, it is computationally lighter than many of the alternatives and especially robust

when structure, texture, and brightness variations are of interest but not color. For instance, in both the areas of fire and smoke detection, intensity shifts across regions are sometimes stronger visual cues than the color content by itself. Smoothed estimate of the background is derived on the image, once the image becomes grayscale. The ones usually used for this smoothing is the local averaging filter which is also known as a mean filter or with a Gaussian filter depending on the required sensitivity to the spatial noise. So the main idea here is to compute a scheme of locally averaged image by replacing each pixel with the mean of the pixel intensity values of the surrounding neighbors within a given kernel size of 3x3, 5x5, etc. This neighborhood has a size which is denoted by N_b smaller neighborhoods lead to locally detailed estimations, while larger neighborhoods tend to produce more generalized estimations of a smoother nature, and therefore deict it.

The smoothed image is an approximation of the static background (an image with light spatial anomalies and transient objects among only other static objects). But the goal is to create this background estimation first as the baseline reference of spatial deviations. From the original grey level image, these deviations are then calculated as difference between body background and original grey level image on a pixel by a pixel basis. It is mathematically described by this operation.

$$S_t(x,y) = I_{gray}(x,y) - \frac{1}{N_b} \sum_{(i,j)\in \text{neighborhood}} I_{gray}(i,j)$$

Where $I_{gray}(x,y)$ is the Grayscale intensity value for a specific pixel location is (x,y) and The summation term represents average intensity of all neighboring pixels within the detailed window and corresponds to (x,y). The subtraction process yields a difference map that contains regions in correspondence with where there is a large spatial discrepancy between the original image and its estimated background.

Because these differences are spatially informative, they tell us how different such patterns are from the local context. For instance, the texture of such areas where fire or smoke is present is irregular, with the edges being soft or sharp, and bright or dark regions are obviously different from the relatively uniform static background. In the difference map, the strong local contrast is made clear when these variations are manifested as high values. Furthermore, through this spatial feature map, the potential regions of interest (such as smoke plumes or fire edges) are isolated, as well as the background, as in the case of speckle noise that is consistent with the environment.

Using this form of the spatial information extraction technique, the image's internal structure is a fundamental layer of understanding. It forms a powerful

descriptor of visual irregularities that capture where and how the pixel intensity patterns depart from the expected background. Such irregularities are often first visual indicators of dynamic environmental events like the appearance of smoke trails, very slight heat distortions, or flame texture—signal of which may not yet be fully captured by spectral or temporal data. In short, the spatial feature map serves as the system's perceptual mechanism to recognize in space, particularly, inconsistencies in space that may be considered as evidence for early wildfire activity or for movement of humans in a scene that is being monitored.

3.1.2(b) SPECTRAL FEATURE ANALYSIS

The spectral information extraction process focuses on the analysis in the color characteristics of the image and utilizes the difference of color as fire is unique in the visible spectrum. Flames in natural fire scenarios usually at higher intensities in the red part of the color spectrum: visually these are flames of reds, oranges or yellows because of the combustion process. This property is then an essential cue to use in order to differentiate fire from any other element in an image. This is then processed in order to isolate its individual color channels, namely the red, green and blue components. Notably, the red and blue channels are of special interest with respect to the conspicuous difference in the presence of fire. Typically, the fire has high values in the red channel, while sky, water bodies and backgrounds shaded from the underlying surface would have high blue or equally distributed intensities over all channels. When we calculate the difference between red and blue channels, we obtain a spectral feature map where the red and blue channels are differentiated effectively. The mathematics of this operation is

$$Sp_t(x, y) = I_{red}(x, y) - I_{blue}(x, y),$$

Where $I_{red}(x, y)$ and $I_{blue}(x, y)$ are the respective intensity values of the red and blue channels at pixel coordinate are (x,y). What this effectively does is to magnify those regions where the red is much more pronounced than the blue (which is usually the case with fire) and conversely attenuate or attenuate regions where the blue component is greater than the red, which tends to be related with fire in the least likely sense.

In this differential approach, it is particularly useful, as it makes for a simple yet effective means to get rid of or decrease false positives present in the image due to other areas in the image that are bright. For example, reflections from water, car headlights, or sunlight on metal surfaces are generally bright but without as prominent red over blue dominance of fire, so their spectral difference map intensity is much less. This independence of shape or texture is further strengthened, since the spectrally derived contrast method performs regardless of the shape or flickering of the fire as long as its color properties are not

changed. The advantage of filtering based on color is that it provides a higher robustness in detection in outdoor or uncontrolled environment where lighting and scenery can change a lot.

In fact, it is particularly effective as it provides a simple but effective way to get rid of or decrease false positives due to other areas of the image that present as bright. In particular, while reflections from surfaces such as water, car headlights, or even sunlight on metal tend to be bright, but not have such notable red over blue dominance of fire as fire, their spectral difference map intensity is much less. Further, the spectrally derived contrast method can be conducted independently of the shape or flickering of the fire provided that its colour properties are not changed. Filtering based on color has an advantage, that is higher robustness in detection in outdoor or uncontrolled environment where lighting and scenery may vary a lot.

3.1.2(c) TEMPORAL PATTERN MODELING

Temporal information refers to the analysis of time dependent behavior of pixels as a function of consecutive frames in video sequence based on their intensity values. When trying to detect naturally dynamic phenomena like fire and smoke with the continuous, unpredictable movement, flickering physical patterns, and shifting textures, this analysis is extremely important. Static spatial or spectral features can be used to identify stationary characteristics of potential fire locations, but they are not effective in locating those features with dynamic characteristics that occur over time. For this reason, temporal analysis becomes important. Firstly, we convert each of the video frames to grayscale, making the comparison between frames easier, as we are left with no color, but now based on the perceived motion as it is compared to the image of the static field of view. After obtaining the current and past grayscale outputs, temporal variation is determined at each pixel location by subtraction of the previous frame's intensity value from the current frame's intensity value. It is mathematically represented by this operation

$$T_t(x,y) = I^t_{gray}(x,y) - I^{t-1}_{gray}(x,y)$$

Where $I^t_{gray}(x,y)$ is the grayscale intensity of the current frame at pixel coordinate (x,y), and $I^{\{t-1\}}_{gray}(x,y)$ is the corresponding intensity at the same location in the previous frame.

This subtraction operation as output yields a temporal difference map based on how many change have occurred at each pixel location between the two time instances.

The temporal difference can be viewed as a very reliable indicator of motion within the scene. Suppose one of the region is flickering flames, rising smoke

or an indicating human figure, then there will be large temporal difference values in this region because the pixel values in that region will change significantly from one frame to the next. However, background elements like the ground, trees or buildings which do not change their placement are expected to see almost no temporal variation between frames, so they will have low or even essentially zero temporal differences. This method, therefore, can efficiently differentiate the dynamic foreground events from stationary background components. Additionally, if video processing starts from the first frame of a whole movie, there is no previous frame to match it against, the temporal information for all pixel locations is initialized to zero as it is to avoid spurious detection without historical data.

The unique power of temporal analysis is that it can capture motion cues that are often good indicators for real fire or smoke activity. Usually fire is rarely static, it dances, flickers and changes shape almost constantly. Smoke swirls and drifts in the air, and light intensity changes that will likely not be readily visible via colored or spatial analysis alone. Temporal information provides the system with a dynamic awareness of the scene by explicitly measuring these frame to frame changes, and hence provides the system with a knowledge of active, ongoing events that it can respond to. Furthermore, this method is also invariant to lighting variations and relative motion of the camera, as the method itself is dealing with relative intensity changes only. The temporal component, when combined with spatial and spectral features, provides a vital third dimension of time based activity detection of which fire, smoke, and other transients can be easily detected, tracked and monitored.

3.1.2(d) FEATURE FUSION AND MULTIMODAL INTEGRATION

After the spatial, spectral, and temporal features can be extracted independently from the input video frames, the next important part of the detection pipeline is to merge these individual modalities into a single representation using a well established feature fusion design. This fusion process is more than simply adding together; it is a thoughtful combination of each feature type such that it improves the overall discriminative power of the system while exploiting the strengths of each feature type, but inoculating it against their weaknesses. The exciting aspect is to guarantee that the final decision is built on an overall view of the scene allowing structural details from spatial cues, spectral features to discriminate with colors dynamics and temporals analysis of motion dynamic. Each of these feature maps—spatial $S_t(x, y)$, temporal $T_t(x, y)$, and spectral $S_{p_t}(x, y)$ provides Complementary information provided by (x,y) allows the system to detect fire, smoke and human presence with more reliability in a variety of environments by intelligently fusing the two.

It formulates the fusion as a weighted summation of the empirical features of these three feature maps. In particular, the last fused feature map is calculated with the equation

$$F_t(x,y) = w_s \cdot S_t(x,y) + w_t \cdot T_t(x,y) + w_{sp} \cdot Sp_t(x,y)$$

Where $w_s, w_t, and\ w_{sp}$ are the weight coefficients used for the spatial, temporal and spectral features. These are no random weights; they are selected precisely because they reflect the context in which the system is deployed and weighted relative importance of each modality in that environment. For example, in good daylight visibility scenarios, the spectral features are quite reliable because color information is clearer and cleaner and less noisy. As a result, the spectral weight w_{sp} can be assigned a higher value However, in such nighttime conditions or low visibility situations such as from heavy smoke or haze, the color features are no longer reliable, and the brightness may focus more on the temporal component, which is the time feature $T_t(x,y)$ which provides a higher weight to w_t . Similarly, the spatial feature $S_t(x,y)$ which is sensitive to texture and intensity anomalies may be beneficial for those circumstances where the background does not change much and structural contrast of smoke or fire is visually apparent.

In fact, we find this flexible, weight based fusion strategy essential for it is based on the idea that each set of features has their own set of vulnerabilities and excels under different circumstances. In cluttered background, spatial features may not perform well, spectral features may not work in color distorted or with different lighting, temporal features may lead to false positive cause of camera movement and irrelevant motion. The system intelligently combines them to balance these weaknesses and add power to the detection capability where a separate feature alone could fail. The resulting fused feature map $F_t(x,y)$ is more robust and richer representation that captures texture variations, color contrast, and temporal variations in one. This improves the confidence and precision of the following decision-making modules like threshold based detector, machine learning classifier, or deep neural network that use this map to determine the location and presences of fire, smoke or humans.

In essence, the feature fusion step is used as amachinal integrative layer which takes three different streams of analysis and joins them into a cohesive and contextually informed output. This enables the dynamic adaption of the system to the change of visual conditions and ensures that the detection performance is consistent and accurate in various real world wildfire monitoring environments. The fact that the overall system can then make nuanced, well informed decisions in the presence of visual ambiguity, environmental noise or challenging terrain, which are all common in outdoor fire surveillance applications, has been possible by incorporating the type of intelligent multi modal fusion.

3.1.3 YOLOv8-BASED OBJECT LOCALIZATION

Combining YOLOv8 and EfficientNetV2B0 together makes them one of the most powerful and advanced deep learning model for the detection today, the core architecture of the wildfire and human detection system relies on the combination of two of the best models together for specialized, but complementary, purpose in the detection pipeline. The main reason for using YOLOv8 based on its design that is optimized to run on real time object detection task. It is one in the You Only Look Once line of models, and in so many ways it brings many enhancements over previous members — it has improved feature extraction, uses a few less parameters at inference time, improves the convergence of training, and has a much more efficient architectural setup. For object detection in active wildfires, YOLOv8 was used and tuned on small custom dataset, gathered to have enough images in early and advanced stages of smoke emission as well as images containing or not containing humans from the hilly, forested or even from the drone captured aerial perspective perspective. During the training, the custom data.yaml configuration file was absolutely necessary, which not only shaped the number of classes but also structured the dataset hierarchy by defining absolute paths to the training and validation folders, setting classes labels, as well as making sure the model can understand how the data is organized and all which is fundamental to the smooth operation of the training pipeline.

For this reason, the training phase was manually optimized to run 100 full epochs in length, as was determined to be a sufficient period based on the observed loss behavior and accuracy trends during the preliminary training sessions. We fixed the image input size to be 640×640 pixels, which is a sweet spot in the YOLO architecture relative to internal aspect ratio, which allows to preserve detail while incurring little cost on computation. Furthermore, this resolution is useful for the early onset of smoke which is one of the reasons that our first detection often look like small streaks or pixel level anomalies in wide angle drone footage. After playing with several iterations of performance tunings, the batch size was chosen to be 16 while taking into account the available GPU memory and the demands in terms of maintaining a stable gradient flow. The small batch size would have brought noisy gradients, the big batch size may have exceeded memory capacity and not given too much benefit. To handle this optimization, advanced momentum based optimizers like Stochastic Gradient Descent (SGD) and AdamW have been employed, which are known to have good proses in exploring loss landscape. They were chosen because these optimizers strike a useful compromise between rapid convergence and the increase of local minima risk (particularly for complex datasets having varying backgrounds and object occlusion). Cosine Annealing was used to schedule the learning rate because it is a smooth cyclical descent of the learning rate. The gradient decay during the evolution of the model helps stabilize model throughout the epochs and prevent to jump in parameters abruptly, which results in more stable convergence and improved generalization.

Yolo v8 runs based on architecture, where it takes an S×S grid from the input image, with each cell being responsible to detect any object present at the cell's center. By

doing so, this method ensures that the model has a spatial understanding of object placement. Each cell then returns the predicted parameters of the object within a bounding box for each of the parameters x, y, w, h. Besides, every prediction also has a confidence score, as it is a product of the object probability and the bounding box accuracy. Key to the design of YOLO is this dual mechanism of classification and localization that allows YOLO to eliminate the need for region proposal networks and simplify detection to a single, unified stage. To augment the robustness of the model further, bounding coordinates are more normalized using techniques involving this ensures that bounding box coordinates are consistent regardless of differents image scales and resolutions in mixed resolutions in drone footage. A whole set of architecture is composed of a set of layers like convolutional layers for feature information extraction and batch normalization of activations has been stabilized, and we have residual connections to workaround the vanishing gradient problem and enable deeper training.

CSPDarknet is YOLOv8's backbone network, whose primary function is to extract high level visual features from the input image. This backbone provides a novel design element, i.e., Cross Stage Partial (CSP) connections, which split the feature maps into two parts and join them later (i.e., 'flow feature propagation'), for more efficient feature propagation and gradient flow. Being lightweight, CSPDarknet is famous for its high accuracy and has proven to be very useful on edge devices and drones with poor computing resources. First, CSPDarknet reduces computation while its semantic features capture strong semantic features of the input image. The backbone extracts features after that and they are sent to the neck of the YOLOv8 model, which consists of a multi scale feature aggregation. It is a Feature pyramid network (FPN) and Path aggregation network (PAN) for the neck. The FPN helps in preserving strength of semantics for object detection at separate scales by forwarding strong semantic features through from deeper layers to shallow layers and then makes sparks, glowing embers, or fine smoke trails are successfully detected. This is complimented by the PAN improving spatial localization and bottom up information flow, ensuring that large scale information such as human figures or big fires are also encoded properly.

At the same time, EfficientNetV2B0 is used in addition to it, as a classifier and feature refiner. YOLOv8 takes the detection part, while EfficientNetV2B0 will help improve the precision of predictings via re analyzing the detected regions and refines the classification results. So, I went to choose an EfficientNetV2 because not only does it scale the number of layers (depth), channels (width), and resolution (input size) together through the compound scaling method, but it achieves optimal accuracy with minimal computation. It is critical to post process cropped detection regions or to determine if the object detected is in fact a fire, or a false positive due to sun glares, dust or fog. The network is designed with Squeeze-and-Excitation (SE) blocks combined with to enable network to focus on relevant channels by dynamically rebalancing channel-wise feature responses. Additionally, depthwise separable convolutions and swish activation functions are used to make it lightweight and powerful. In this project,

EfficientNetV2B0 is served as a verification stage, offering a second layer of intelligence to be able to filter out the misclassified cases and verify the ones made by YOLOv8. In particular, this becomes extremely useful when discriminating between smoke and cloud, or human vs. tree stumps in complex and noisy inputs from drone footage.

3.1.4 QUANTIZATION FOR EDGE EFFICIENCY

The use of YOLOv8 for detection and EfficientNetV2B0 for classification refinement as a combination provides a multi-level and highly robust detection system to detect wildfires, smoke appearance, and the existence of human beings in various environmental situations until now. Second, a dual model strategy on the other hand, significantly improves the precision and recall, reduces the possibilities of false alarms, and leads to timely and reliable detection that are essential for timely and accurate deployment in wildfire management and rescue operations.

In order to enrich the detection pipeline even further, especially under bad or low visibility conditions when haze, smog or thin smoke can not be detected visually, I applied Fast Fourier Transform (FFT) to grayscale image frames. The reason to carry out FFT is to get the frequency domain feature which is not easily found by just using the raw spatial analysis. The spatial domain looks at changes of intensity and edge through pixel intensities and edges while frequency domain shows the changes in intensity with respect to spatial domain—depicting the detectable patterns or textures like repeat textures, subtle vibrations, or periodic gradients that could betoken faint smoke, early stage fire flickers, or heat distortion effects. FFT offers a unique channel of information that complements the convolutional features learnt by EfficientNetV2 by analysing the image spectra. e.g., for instance, a reader may see a fog bank and a cloud visually and in RGB format seem very similar, yet the frequency signature of the clouds in front of the model might be sharper, periodic, or not as full of noise as the frequency signature of the water vapor from perhaps a fire situation, enabling the model to eventually better distinguish between natural weather conditions and potentially dangerous fire related emissions. Therefore, this multimodal pipeline enhances the spectral robustness of the detection system and comparatively improves its ability for fire related phenomenon detection in visually ambiguous scenes where only spatial features may be misleading or inadequate.

Therefore, I spent a lot of time perfecting the YOLOv8 loss function, which is a compound of three intertwined parts: localization loss, classification loss, and confidence loss. The most important factor of bounding box prediction accuracy is localized loss. This helps to make the predicted coordinates of each object (x, y, width, and height) as close to the ground truth bounding boxes as possible. The loss function has been partitioned into two, with this part of the loss only involved in those grid cells which were used to detect the object, thus making the training more focused and efficient. Similarly, classification loss is calculated in terms of categorical cross entropy which punishes the model for each incorrect classification of an object; for example,

smoking plume is classified as a cloud, or recognizing human silhouette versus a tree branch. It is vital for retaining high class-wise accuracy, particularly when many classes are present in a single image as occurs frequently in wildfire detection involving both the flame and human subjects. Confidence loss, a measure usually calculated using binary cross entropy, provides the model the ability to measure and adjust its confidence on its predicted outcome. It forces the network to be both a false positive (i.e. detecting fire when there isn't any there) and a false negative (when their is a fire and it isn't detected). Together with YOLOv8 further minimizing these loss functions means it learns to better identify hazards, and consistently, reliably predicts hazards in real world, unstructured environments.

Two crucial model optimization strategies I adopted to make the entire system computationally efficient and scalable for deployment on low resource platforms such as drones or embedded boards, were L1 structured pruning as well as post training quantization. L1 structured pruning works by finding and removing the entire parts of the network e.g. filters, neurons, or convolutional channels and then deciding based on those network components on whether their contributions to the model's predictivity are significant enough to keep them (L1 pruning) or not (L1 structured pruning). In particular, it finds the L1 norm (i.e., the sum of absolute weight values plurality) of these components and prunes the lowest outliers, on the basis that they have the least impact on output accuracy. Structured pruning is more advantageous than unstructured pruning because, unlike the latter, it preserves the network's architecture and thus avoids the creation of irregular sparse matrices that are hard to accelerate on standard hardware. Structured pruning cleans up the network, produces a more hardware friendly network that speeds up, conserves memory, and is easier to deploy on real time systems without compromising model predictive quality.

On top of that, quantization has been later introduced to further compress the model by reducing its numerical precision in both weights and activations. The original model operates in 32 bit floating point numbers with high precision but with a higher memory and computational cost. Quantizing the model brings a drastic memory footprint hit and also speeds up of inferring (mainly when the devices support only integers such as ARM Cortex or few edge TPUs besides). Quantization can sometimes incur loss of accuracy, but I utilized quantization aware training (QAT) ideas to simulate quantization on the training phase itself wherein the model can compensate for any lost precision. Mixed-precision techniques were also investigated where the most critical layers of the network keep higher precision, and the rest of the layers are fully quantized to achieve a good tradeoff between speed and accuracy. Using the structured pruning and quantization, the whole system became leaner, faster and far more deployable for real world applications.

I further extended the detection beyond just a visual one by incorporating, such as photos of the scene, audio analysis, and temperature mapping, requisite for an all encompassing, responsive, useful safety monitoring system. For the thermal analysis part, I used infrared imaging from thermal cameras or thermal data fusion from IR

sensor to detect the temperature anomalies. Validation of the visual models was also provided using this data. For an example, once the system visually detect the smoke, it also senses a sharp spike in the temperature in that area, which helps with certainty and lowers the fatigue of false alarms. But in the other case, if there are no visible cues but the thermal system picks up a hot spot or heat spike, this may be an early warning of fire before visible smoke or flames are evident. The temporal thermal anomaly detection adds to the spatial detectors in complementing early fire detection capabilities.

In addition, audio module plays a key part in contextual, and situational awareness especially when humans are present. The onboard environmental audio is constantly monitored using microphones, and the system is trained to identify acoustic signals associated with emergencies, for instance human screams, cries for help or distress calls. If such sounds are heard, they pass through a speech-to-text transcription Pipeline using something like pre trained models such as OpenAI's Whisper or Google Speech to Text API to turn audio to reasonable text. The cross reference with visual and thermal detections of people in danger validates this information on multisensory basis. In rescue scenarios, where vision may not be good the victim may not even be in the immediate camera frame, but moving by sonar could still be heard.

Once it is confident that any anomaly is detected, whether smoke or fire or high temperature or distress audio, the system is programmed to automatically send real time alerts via email notification. Critical elements for these alerts include an annotated screenshot of the detection, exact timestamp, geolocations (if GPS/GPS telemetry is present), transcript, and detection confidence. This enables first responders or other emergency personnel to be immediately notified with all the context they need in order to respond with speed and decisiveness before the situation has had a chance to escalate out of control in the first place. All these techniques are integrated and can form a stack, which makes it an intelligent surveillance system in which the monitoring is a combined one of proactive and context aware.

Finally, YOLOv8 for object detection, EfficientNetV2B0 for spatial feature extraction, Fast Fourier Transform for spectral analysis, and environmental sensing with temperature and audio modalities are integrated with the overall system. In addition to this, L1 structured pruning and quantization results in highly accurate, yet computationally efficient and deployable in real time, resource constrained environments models. The technology is highly scalable and adaptable to a large variety of fire detection and rescue systems, embedded in drones, IoT surveillance systems, mobile units and edge devices, as demonstrated by this system design. Overall, such a research and development effort expands what has been possible to date in intelligent environmental monitoring, enabling a robust, high performance, and deployable tool for near real-time fire, smoke, and human detection in ever changing and extreme environments.

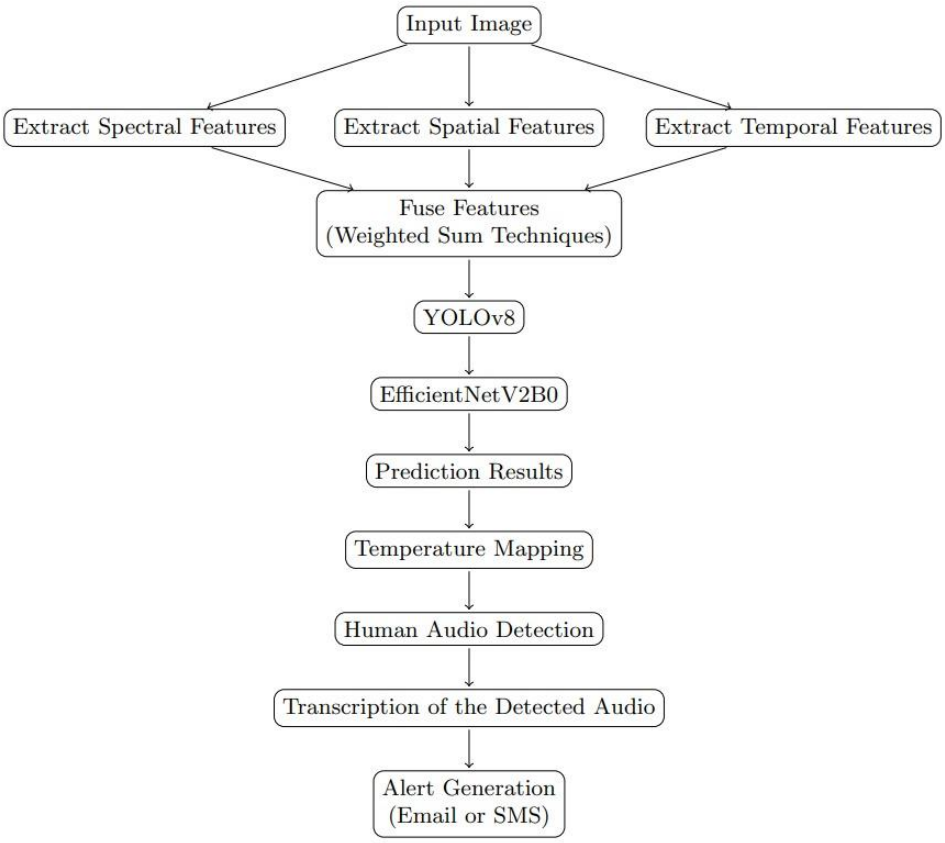

FIGURE 2: A SIMPLE WORK FLOW DIAGRAM

3.1.5 TEMPERATURE MAPPING

I integrated Temperature Mapping as a complementary modality to complement the visual and auditory components of the emergency detection system in order to monitor and detect, in advance, thermal anomalies, which may act as precursor signs to impending fire outbreaks or overheating hazards. The process starts by reading an input image that is usually an infrared or standard RGB camera frame. First, the image is loaded using the OpenCV library in RGB format. However, it is important for phreatic temperature mapping that the RGB format gives us the three channel color info in case needing to do multi modal comparisons, although in that particular case its much easier to convert its original RGB image to a grayscale image format. Grayscale conversion makes pixel data simpler, converting them into a format of intensity values that reflect the brightness range of black (0) to white (255), which also plays the role of relatively temperature in the scene. This is important as temperature values are usually derived from the presumption that brighter areas are hotter, so grayscale is a valid intermediary for thermal interpretation.

After the image was converted successful to grayscale its pixel intensity values need to be normalized to the same range as can be processed further. Min Max normalization, To bring a grayscale intensity value onto a range [0,1]. It normalizes this obtained pixel intensity so that the distribution of the pixel intensities is uniform across the entire image, which reduces the adverse effects of extreme outliers and standardizes the data for the consistent transformation in the next step. The formula used for normalization is as follows:

$$I_{\text{norm}}(x,y) = \frac{I(x,y) - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}}$$

Here, I(x,y) is as a grayscale intensity at a pixel's location. (x,y), while I_{min} and I_{max} are the lowest and highest extreme intensity values seen within it. This change matters, since it maps each pixel onto the intensity range seen, letting small shifts of heat be felt and shown well.

After normalization, one step after is to convert these normalized intensity values into estimates of actual temperature. This is done through precisely mapping all exact values onto a true temperature scale. Within my implementation, I fixed both the top and bottom temperature caps as $T_{min} = 0^0 C$ and $T_{max} = 100^0 C$, into common limits inside human-safe thermal monitoring purposes. The shift from some fixed intensity to a temperature scale is well done. This has the following simple formula.

$$T(x,y) = I_{\text{norm}}(x,y) \times (T_{\text{max}} - T_{\text{min}}) + T_{\text{min}}$$

The specified equation performs a scale transformation of normalized intensity measurements to produce their corresponding thermal values. When normalized intensity values measure 0.5 the corresponding temperature would fall within the middle range at 50 degrees Celsius. This linear transformation converts the entire grayscale range into meaningful thermal values that reflect proportionally each change in pixels. Through this calculation we create a temperature value for each image pixel position so every point contains real world thermal information. The spatial coordinates (x,y) in each image section corresponds to actual temperature readings which deliver accurate high-definition thermal measurements throughout the whole observation area.

The Inferno colormap from Matplotlib helped me enhance both understanding and visualization of final temperature maps. The Inferno color scale enables consistent color contrast in its transitions thus people can better see small temperature differences. This color scale provides human readability through its purple-to-yellow-white transition which reflects cold-to-hot conditions. Its heat-sensitive color scheme makes the program an optimal solution for safety work because it promptly alerts users about critical zones requiring immediate inspection. The temperature image features a color bar as a side panel which explains the link between temperature values and the color

variations displayed in the image. The bar functions as a reference point which helps users or automation systems connect thermal measurement values to the color intensity spectrum to make proper decisions.

A final product shows a fully interpretive heat map positioned over an image's spatial components which enables online heat anomaly detection prior to RGB imaging visibility. The temperature map functions as an essential accessory feature of the system because temperature increases occur before smoke or flames appear during early fire detection phases. Integrated audio and visual data together with this system creates an enhanced situational awareness that delivers complete environmental danger information to users. Engineers created a process that starts with RGB input then converts it to grayscale followed by normalization with thermal transformation and color-coded visualization specifically to operate efficiently on drones and embedded thermal sensors as real-time actionable edge devices.

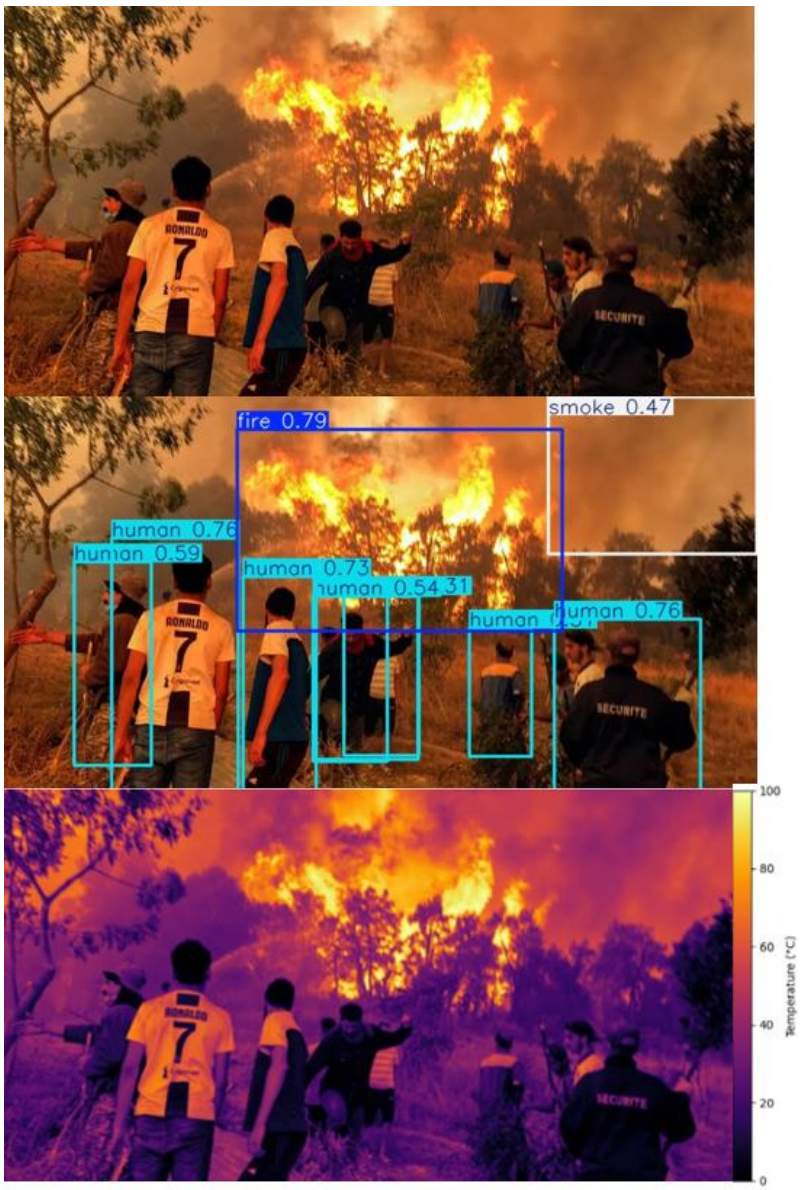

FIGURE 3: GENERATED TEMPERATURE MAP IMAGE

3.2 AUDIO SIGNAL ANALYSIS AND INTERPRETATION

3.2.1 OVERVIEW OF AUDIO PROCESSING METHODOLOGY

And the methodology for audio noise filtering & human speech detection that I've implemented, is so as to boost the clearness and accuracy of voice transcription from noisy environments, i.e. situations whereby drone-based surveillence and such. However, in many of the practical domains such as wildfire detection and environmental monitoring with drones, there are usually significant levels of background noise in the audio data. This noise is from wind turbulence, mechanical vibrations, engine sounds, and other ambient environmental noise and all of which can make recorded audio of minimal quality. For instance, this degradation affects the intelligibility of human speech and negatively affects subsequent processing tasks such as transcription and real time analysis. For this problem, a noise filtering and human speech detection system must be robust enough to isolate meaningful voice signals from noise components.

The goal of this methodology is to design and implement a system where it filters out noise from audio recordings with the goal of accurately detecting and transcribing human speech. This way ensures that audio data recorded from drones or other field devices comes in high enough quality so that it can be accurately recognized and analyzed. The methodology is designed to filter off unwanted noise systemically so as to leave the essential speech components for transcription accuracy intact. The system demonstrates improvements in reliability of subsequent voice processing and analysis stages by doing so.

The use of this methodology in terms of the application to wildfire detection and drone based surveillance cannot be overlooked. Until recently, it wasn't uncommon for drones to be deployed in remote and harsh environments where noise levels are by nature high, and identifying human presence or voice from a command is essential for effective surveillance and communication. A noise filtering and speech detection system that is accurate and fast enough is critical when there is need for rapid decision making; for example, when it comes to search and rescue operations or distress signal monitoring. In addition to this, speech recognition in drone based systems increases their usability (and therefore usefulness in commanding drones) as well as allowing for more dynamic and more interactive commands (e.g. voice reporting, send voice alert).

It uses state of the art models and techniques to rid the background noise and also retain the core human voice if which is vital for an accurate transcription. To enhance the SNR, advanced noise reduction techniques are used to process the audio such that it contains no appreciable distortions and maximum clarity. Moreover, voice detection algorithms are fine tuned carefully in order to discriminate human speech from background noise, so that false positives are minimized and there is high reliability of detection under difficult conditions.

Furthermore, the deep learning based transcription models serve to boost the accuracies of the system more, particularly when the audio data is complex and noisy. As an entire pipeline, the noise reduction and accurate transcription along with voice detection all together help to make a robust pipeline to process the audio in real time. In the following sections of the methodology we look into more detail of the whole process including architecture design, selecting libraries and models, noise reduction strategies, voice detection methods, transcription methodology, and integration of automated notification system. With this approach, one guarantees the reliability and the consistency over the operational scenarios for the system to achieve high transcription accuracy.

3.2.2 SYSTEM ARCHITECTURE AND DESIGN

In our case, the proposed system is a multistage pipeline that processes the audio data efficiently and accurately in environments where the noise level is higher like drone based monitoring for surveillance. Since the system architecture is carefully structured, it seamlessly integrates all different components that play an essential role in increasing the overall performance and reliability of the audio processing pipeline. The architecture's main goal is to preserve the capability of real time processing while achieving high quality transcription even if the input audio is significantly distorted by noise.

In terms of systemic architecture, the system was architected at high level such that data flows through a series of stages that are well defined: noise reduction, human voice detection, transcription, and the automated email notification. The first step in the process is the acquisition of raw audio files, which the user acquires by uploading via a user friendly web interface. Then these audio files, most of which have substantial background noise, move through the noise reduction module to minimize the annoying sounds while retaining essential voice components using very sophisticated filtering. At this stage, clarity of the subsequent transcription process, is greatly enhanced and errors due to noisy data are minimized.

After the completion of the noise reduction the processed audio is passed to the human voice detection module. The analysis of the audio in this component is done by energy based techniques to determine if the audio comprises speech that might be discernible from a human. The system accurately detects whether the recording is of a human voice or just background noise by calculating the energy level of the audio signal and comparing it with a pre defined threshold. In that way, unnecessary processing is skipped in cases without human voice, to optimize system performance as well as holding additional computational cost.

If the conversation is with a human voice, the next stage is transcription, where speech is recognized in an automatic manner with the Whisper model being used for automatic speech recognition (ASR). Deep learning based model Whisper by OpenAI produces very high accuracy in transcribing human speech in difficult acoustic environments. The model accepts an input of the denoised audio and produces the textual transcription

of the spoken content as an output. With noise reduction and also a more precise voice detection, the transcription of Whisper is far more accurate since it gets cleaner and quieter audio signals to process.

The transcribed text has then to pass through an automated email notification process in order to finalize the pipeline. When system is used for live monitoring scenarios, prompt delivery of transcription results is important, making this stage very beneficial. The Simple Mail Transfer Protocol (SMTP) is utilized for sending an email to a defined recipient which contains the transcription. However, emergency response systems, remote surveillance setups and the like need to get some immediate feedback, which is best catered by this automated notification mechanism.

Overall, the architecture is thought out to be modular and scalable, enabling each component to do its thing independently, but slide nicely over the other components. Beyond the benefits of maintainability, this modularity also simplifies the task of keeping the whole running, because it is easy to update one of the components as a new technology becomes available. It further adds that the processing of the system of the system in real time also guarantees that transcription and notification take place in a short time minimizing delay and keeping the system responsive for mission critical applications. The rest of the sections provide more technical detail into the implementation of each module, and how algorithms and libraries are used to achieve as high an efficiency as is possible.

3.2.3 NOISE REDUCTION TECHNIQUES

Audio processing pipeline is greatly benefitted from the noise reduction which is a basic and important step in the noise reduction process of audio data before transcription. In many real world scenarios such as the drone based monitoring or outdoor surveillance applications, it is very common that audio signal has a lot of background noise coming from the environmental sound, drone propeller or wind etc. However, if this noise is not adequately addressed, the quality of the transcription itself becomes seriously degraded resulting in inaccurate or unintelligible results. The system deals with this problem by using the NoiseReduce library, an advanced Python library which can reduce background of the noisy audio while maintaining the important speech components.

Spectral noise gating is the basic technique on which the NoiseReduce match is built. That is why spectral gating, which analyzes the frequency spectrum of an audio signal and tells it where the noise is the most prevalent, works. Usually, noise occupies certain frequency bands, which are different from that of the human speech frequencies. In effect, this algorithm is able to suppress the contribution of these noise prone frequencies without losing the spectral component which is related to human voice. It is done through the Fourier Transform based analysis which correlates the audio signal into the different characteristic frequencies. By applying a threshold to decrease the spectral gating so that frequencies below the desired signal strength are attenuated and unwanted noise is reduced, the spectral gating technique is then applied.

Spectral gating for noise reduction offers the main advantage of improving the signal to noise ratio (SNR) without overly affecting the human voice. Unlike standard noise reduction techniques, spectral gating does not include any artifacts or detriment to speech quality as it filters out the non-voice frequencies. This way, the processed audio is not muffled or unintelligible in challenging acoustic environments. Additionally, spectral gating in the NoiseReduce library is implemented computationally efficiently and also very customizable ensuring that noise threshold and gating parameters are adequately tuned for the noise spectral shape of the audio input of your choice.

As the very first step in the audio preprocessing pipeline, noise reduction is applied in the implementation of this system. This strategic position gives clean well processed audio to the downstream processing, for instance, detection of human voice and transcription, which helps to maximize the performance and accuracy of the process. When the system gets an audio file, it loads the raw audio using Librosa library, the library efficient for dealing with different audio formats and sample rates. Then, this denoising function is applied, employing NoiseReduce's spectral gating technique to remove ambient and environmental noise. After denoising of the audio, it is then fed into the human voice detection module for additional analysis.

When implemented as a preliminary step, the system reduces the amount of human speech detection errors and respective false positives, by integrating noise reduction to increase the quality of transcription. This accuracy improvement is truly necessary when dealing with drones and any audio environment where background noise is present. Furthermore, the noise reduction module in this case provides clean and intelligible audio which helps ensuring the effectiveness of next processes of voice detection and automatic speech recognition. The system is able to achieve a robust performance even in an acoustically hostile settings by way of combining advanced noise suppression techniques and close integration with the pipeline.

3.2.4 HUMAN VOICE DETECTION PIPELINE

Then it comes with the most crucial step of human voice detection after it performs noise reduction to make audio clear. This ensures that there is enough speech content in the input audio file before progressed further in transcription. At this stage, detecting human voice could be used to filtration out the uninteresting, silent files that would otherwise waste computational resources. Incorporating this detection step thus ensures that the transcriptions are not susceptible to false or misleading transcriptions caused by non-speech inputs like background noise or mechanical sounds from drones. Not only does this offer improved system efficiency, but also means that the transcription module only takes on the task of processing audio data that is both useful and meaningful.

The system resolves reliable human voice detection by employing an energy based detection technique which is well suited for real time applications due to its simplicity of computation and reliability. So, the basic idea of this method is that human speech in itself has higher energy than background noise or silence. Speech, while amplitudes

and frequencies vary, features pronounced variations whereas most background noises have lower energy or present with more consistent patterns. Because the audio signal's energy can be quantified, it can be determined to consist of speech or to contain more noise or silence.

The detection algorithm based on energy is performed by the cumulative energy calculation of audio signal within determined time window. The signal energy is computed as the summation of squared amplitudes and gives a robust signal intensity. An audio signal x(t)x(t)x(t) over a given time interval is mathematically described by the energy EEE as follows:

$$E = \sum x(t)^2$$

In this case x(t)x(t)x(t) is the amplitude signal at time ttt. It turns on the system if the calculated energy is greater than a pre defined threshold and therefore the system believes that there is human speech in the audio. If not, the audio segment gets classified as non speech or noise. An empirical analysis and experimentation is used to carefully choose the threshold value in order to balance sensitivity and specificity. If threshold is low, there are high chances of getting false positives (detecting background noises as speech) or vice versa a low threshold missed speech segment.

This approach based on energy, is more favorable than complex techniques such as spectral analysis or machine learning based classifiers mainly because it is lightweight and executes fast. Due to the fact that the system was intended to process real (or near real) time audio data minimizing computational overhead is an absolute requirement. Furthermore, this technique is robust against noise burdening environmental conditions such as environments where background noise reduction in the previous stage had largely been achieved.

Once the noise reduction is completed, the energy detection function is immediately called and applied to the denoised audio to help further minimize the amount of errors in the voice presence estimation. In turn, this system runs high accuracy without unnecessary transcription of irrelevant files. Integrating human voice detection in the audio processing pipeline is an extremely important strategy for the system in terms of performance and reliability, since only meaningful audio data is passed to the transcription phase.

3.2.5 SPEECH TRANSCRIPTION USING WHISPER

Once the system has successfully reduced the noise and detected the human voice, speech transcription becomes the critical phase of the system. At this stage, we convert filtered and validated audio data into textual representation with the help of the Whisper model. Automatic speech recognition (ASR) system Whisper is developed by OpenAI and considered as a state of the art ASR system with high accuracy and robustness,

especially in difficult audio conditions. It is a good choice for our project that deals with drones audio recording in dynamic outdoor fields, due to its capability to handle different accents, dialects and noisy environments. The system guarantees that the transcription process is done accurately and reliably even in cases of complex acoustic environments by employing the Whisper model.

The reason why the Whisper model is preferred is because it has shown that it is a better performer than any other ASR model, and this is due to its large training data on a variety of languages and environments. The extensive training permits Whisper to generalize well, thus triviating errors on audio with environmental noises or very slightly distorted versions of itself. Unlike existing ASR systems for which the traditional approach, may fail dramatically on the presence of unusual noise patterns, Whisper's deep learning architecture is inherently adapted to diverse input conditions. Consequently, it always delivers high quality transcriptions, relatively free of errors, therefore this is an excellent tool for audio data that originates from a drone.

The filtered audio file is filtered through noise reduction and human voice detection before it is fed into the Whisper model in the transcription process. Basing on the model, this is a sequence to sequence architecture that has an encoder decode framework transforming audio waveforms to coherent text. Acoustic features are extracted from the audio input signals and used to reduce redundant or irrelevant noise components, and fed to the encoder. Afterward, the decoder understands the above features and outputs a textual text using its structured and readable transcription of the spoken content. Whisper's architectural approach allows for handling both continuous and spontaneous speech patterns accurately and at the same time captures the most subtle phrases and expressions. The transcription process of Whisper is made to handle multi-lingual input but for this project we restrict the model configuration to only process English audio data. Through this, the output is specifically guaranteed consistent and relevant, especially in cases of standard communication protocols and reporting formats used mostly in drone operations. The model works with audio sample rate of 16kHz, the same as pre-processed audio format and therefore is compatible with the transcribed output.

The system further processes the text to make its utility once the transcription is generated. In what follows we will be concerned with one of the major pieces of this post processing stage: extracting email addresses from the transcribed content. Drones can record conversations where the words to which critical contact information is verbally imparted, and the need to accurately identify email addresses is key to assure further communication or reporting. The transcribed text is scanned using a regular expression based pattern matching algorithm which extracts potential email addresses that are stored for later use or as an alert.

After the completion of transcription process, the complete transcribed text is the final output, wherein the text is formatted for further clarity and interpretation. Apart from that, the system gives you the option to display the transcribed content on the user

interface and also shows the density by displaying the number of words in the spoken content. This is an indication of how substantial the transcription is to be considered significant. The system sends the extracted text via automatic email to one or more preselected recipients once the word count exceeded the predetermined limit. In addition, this feature will be very useful in monitoring scenarios where real time feedback and report is mandatory.

In brief, as part of the speech transcription pipeline, using the Whisper.model has a high accuracy and high efficiency method to convert drone acquired audio into text. The robustness of Whisper in noisy environment along with its capacity to keep contextual coherence make it a valuable part of the system. It automates both the transcription and the post processing to provide accurate speech to text conversion as well as enable the timely identification and communication of critical information to enable improved work flow.

3.2.6 NOTIFICATION AND ALERT MECHANISM

Once the audio data is successfully transcribed, the system goes to last stage which is automatically sending the transcribed text through an email. The role of such an automated email notification system is vital to ensure that the processed information is passed on to the intended recipient in a timely manner. The automated email feature is a powerful step in improving the efficiency of the entire system in drone-based surveillance and monitoring, since timely dissemination of critical data is very important. In this way, it can quickly pass the content to preordained email addresses to be sent to those interested, but with no human intervention and no waiting to ensure access to the communication.

The email notification system uses the Simple Mail Transfer Protocol (SMTP) – a well used protocol for sending electronic mail over the internet. This is because SMTP is a robust and compatible transport mechanism for email servers and services. Here, the system comes up with a secure connection with Gmail SMTP server by utilizing ‘smtp.gmail.com” and port 587 that does support TLS (Transport Layer Security) encryption. It guarantees safe transmission of data that is being sent through the mail with protected encryption from interception or tampering. Confidentiality and security are achieved by using credentials with the sender’s email address and a secure application specific password as part of the authentication.

After the transcription is made, the formatted text is made in an email body that is easy to read and concise. The subject line of the email is so clear about the subject that the email is consisting of the true transcription of the recorded audio without removing any words which will help the users to understand that what they are looking at is the real, unread version of the text. Python’s email library provides the classes MIMEText and MIMEMultipart that allows incorporating plain text as well as complex message formatting into the email by structuring the email. This enables these classes to be used to build well organized, easily understandable emails, which will give it a more professional touch and also friendly output.

The system creates a session with the SMTP server to begin transmission of the email, after which it begins handshake process in order to start secure connection. If email sent to client after authentication, email is sent from specified sender email address to recipient email address. This project has the sender and recipient using the same address for the sake of simplicity and demonstration purposes but the system architecture is able to accommodate multiple or dynamic recipients. Upon a successful dispatch of the email, the system logs the operation to confirm completion.

For the reliability of the notification system, robust error handling mechanisms have been implemented all along the process. The system catches the exception in case any error takes place in the connection establishment, the authenticate phase or sending phase and logs an error message. The proactive error handling on the other hand, makes the system immune to temporary issues like outages in the network or incorrect credentials. The system also provides the user with informative feedback on the status of the email, if it were successfully sent or if there was an error in doing so. Apart from the fact that it makes for a much better user experience, it also helps solve any problem that might arise.

Given that transcribed content needs to be efficiently communicated to users in a timely and reliable manner, the automated email notification system therefore serves as an integral part of the audio processing pipeline. The secure protocols combined with detailed error handling ensures reliable transfer of critical audio data, making the system suitable to be used in the application areas of real world monitoring and surveillance. This seamless integration of transcription and communication provides maximum utility of processed audio by keeping the user informed without manual intervention.

3.2.7 ROBUSTNESS AND ERROR HANDLING

Robustness and error handling of the system is designed in with such detail that it could smoothly continue to work even in the situation of unexpected events. Real world deployment would involve supported file formats, corrupted audio data and other server connectivity or transcription failures. Any of these scenarios need to be anticipated and handled well such that the application does not crash or provides bogus results to avoid destabilizing the system and causing issues to the end user.

The main challenge addressed in the system is to handle unsupported or corrupted audio files. The system takes in user up loaded audio file, and therefore one of them will not necessarily conform to expected audio format, or will contain invalid or corrupted data. However, to avoid this, the backend first checked for the availability of an uploaded file before continuing with the processing. In case that there is no file, or the file format is not compatible with that decoder, the system responds promptly with an informative error message that points the user to a correct file provided that the file format for the audio is compatible. Such a proactive validation is effective to prevent unnecessary processing and save computational resources.

The system also corrects errors that can occur during loading and processing of audio. As an example, suppose librosa library fails to load an audio file while attempting to do so and the system catches the error thrown by librosaluv and logs the error containing such details. As a result not only developers can be able to understand issues while testing, but application can't be ended abruptly. The system can deal with such errors carefully, through this architecture provides a resilient system that still will continue to work when encountering problems with input.

Furthermore, the email notification component is also subject to network and connectivity issues. If something prevents the SMTP server from securing a connection or verifying the sender (determined to include network outages or providing incorrect credentials), the system traps the exception, and logs a descriptive error message. It does not allow the user to remain in the dark about the failure, rather, it outputs a call out that tells the user what went wrong; network connectivity and authentication details are the focus of which the system should be dependent. This is to prevent any kind of confusion or frustration as the user is able to know when there is an issue and also make them transparently know what is the issue that is being faced by them.

Additionally, there is another potential point of failure when the audio in Whisper cannot be processed or transcribed accurately. In such a situation, the system captures any exceptions raised by the transcription function and offers a fallback like a response that points out that the transcription was not able to complete. This maintains the responsiveness of the system, no matter what kind of data it may face or what the model does not work with.

It also handles general exceptions during the file upload as well as processing pipeline for better overall robustness. By utilizing this catch all error handling mechanism in an unexpected way, this catch all error handling mechanism makes sure any for misguided errors will be handled with dignity instead of effectively shutting down the server or making the server completely dormant. With detailed logging and friend user friendly errors messages, the system is able to minimize downtime and provide easy quick trowels.

Finally, by effectively implementing the entire comprehensive error handling strategies, the reliability and stability of the system are finally greatly enhanced. Using the ability to predict the issues that may occur at all the processing stages and apply tight solutions to it, it shows the robustness against the common challenges in audio processing and communication tasks. The system creates continuous operation even with the real world situations by paying high attention on robustness so that it keeps performance and user satisfaction at a high level.

3.3 SENSOR-BASED ENVIRONMENTAL MONITORING

3.3.1 OVERVIEW OF THE ESP8266 WI-FI MODULE

The ESP8266 microcontroller is at the dawn of a revolution; a relatively low cost, compact piece of technology which opens up an incredible number of possibilities for the development of many IoT applications. The ESP8266 is developed by Espressif Systems which is able to seamlessly connect the devices to internet and transmit the real time data and control. It is a superb tool for the development of modern embedded systems thanks to its versatility, supported multiple communication protocols and excellent processing power.

The ESP8266 is one of the most remarkable things in the way of its integrated Wi-Fi connectivity, which supports 802.11 b/g/n standards. This capability makes the device fit in very well in communication with the other Wi-Fi enabled devices, and it can either work as a Station (STA) and Soft Access Point (AP). Furthermore, it is also capable of dual mode operation (AP + STA) where such device can operate as a mean server device and at the same time work as a client of other network. Interaction with the central server can also be made by the device, which is necessary in the case of IoT application where the device must broadcast information and in the same time accept commands from the server central.

The ESP8266 microcontroller has a 32 bit Tensilica L106 operating at a clock speed of 80MHz of which can be raised to 160MHz in case of a more demanding application. Although it is small and draws little power, it has formidable computing power and is suitable for data processing and also for performing real time tasks. The device has 64KB instruction RAM, 96KB data RAM, and external QSPI flash (512KB to 4MB according to model). With this memory configuration, it can store and execute such complex firmware as well as have enough space available for data storage and web pages necessary to host local server interfaces.

The ESP8266 es the most advantage of the ESP8266 is that it supports multiple communication interfaces such as PWM, ADC, I²C, I²S, UART and SPI. It allows you to connect with any kind of sensor, actuator, and peripheral device, which makes these interfaces a very flexible choice of presenting in IoT dasp. It features useful PWM capabilities able to control the components like LEDs and motors with precise control and the Analog to Digital Converter (ADC) can read the analog signals from the sensors like gas and smoke detectors. Additionally, the ESP8266 uses Serial Peripheral Interface (SPI) and Inter-Integrated Circuit (I²C) protocols to talk with external modules, e.g. displays, environmental sensors.

The battery powered forms of devices including ESP8266 also demands for the power efficiency of ESP8266 which is also another crucial thing to make ESP8266 suited for battery powered devices. It works at 3.3V DC by which the microcontroller consumes minimum power in case the device becomes idle by using microuncontroller features such as deep sleep and light sleep. The ESP8266 consumes only 20 μA in deep sleep

mode, which is of utmost importance for applications when the device must operate for an extended period without requiring frequent recharging or power replacement.

It is one of the greatest advantages of working with ESP8266 as it is easy to program and supports the firmware. However, it is compatible with the Arduino IDE, PlatformIO, AT Command Set, MicroPython, NodeMCU (Lua). This makes it easy to program and customize the device for various programmer's background, thus, giving flexibility to developers to design and modify the device depending on the desires of their projects. A popular example of using Arduino is using the Arduino IDE, which specifically simplifies coding while offering a wide range of examples and community contributed (functions) which greatly broadens theamarinade by your side.

In addition, the ESP8266 is over the air (OTA) firmware upgradable. It is especially valuable in the case of IoT deployments where it is difficult to access the device physically. OTA updates assists in ensuring that the system is always secure and up to date without compromising to its operations.

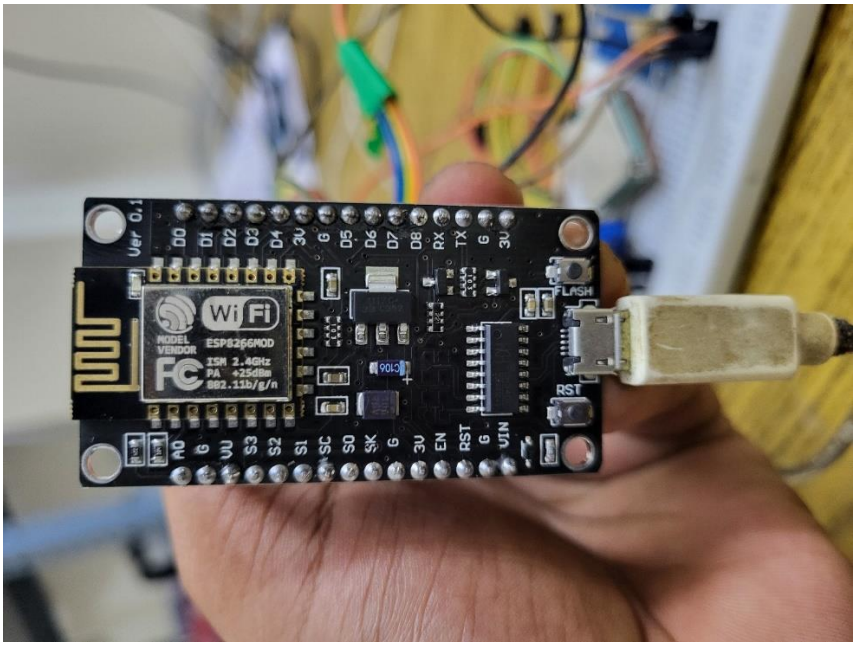

FIGURE 4: ESP8266 MICROCONTROLLER

3.3.2 AIR QUALITY MONITORING USING MQ-135

MQ135 smoke sensor is one of the most versatile and most commonly used gas sensor in industry to detect different dangerous gases and pollutant gases; which is so much required for Air Quality Monitoring, Environmental Sensing and Safety applications. The sensor is manufactured by Winsen Electronics and is very good for detecting gases like carbon dioxide (CO_2), carbon monoxide (CO), ammonia (NH_3), alcohol vapors, benzene, smoke and other volatile organic compounds (VOCs). Its low cost, giving rise

to high sensitivity and easy interface has made it the darling of hobbyists, researchers and professionals installing safety and environmental monitoring systems.

The chip level sensor is based on chemiresistance concept, the input of the sensor (chip level) response in terms of electrical resistance to the presence of a certain gas molecule. This sensing mechanism is based on its gas sensitive semiconductor layer, which is made of tin dioxide or SnO2. The semiconductivity of tin dioxide is n type, which means there are great numbers of electrons ready to be shared by conduction. The conductivity of SnO_2 is very low in clean air, because oxygen molecule adsorbs on the surface and captures free electrons of the semiconductor, thus forming oxygen ions (O^-, O_2^-). It significantly reduces the number of free charge carriers at the sensor, thus increasing the resistance of the sensor.

If the sensor gets exposed to gases such as carbon monoxide (CO), ammonia (NH_3) or alcohol vapors, the gas molecules react in a chemical reaction with adsorbed oxygen ions on the surface of SnO_2. When these gas molecules collide with the oxygen ions, the captured electrons are given back to the material of the semiconductor to release. Therefore, the resistance of the material is reduced due to an increase in the electron density on the SnO_2 surface. Its resistance decreases dependently from the concentration of the gas present and proportionally. The principal way gas is detected using an MQ-135 sensor is on the basis of this relationship between gas concentration and resistance change.

There are few components that comprise the MQ-135 sensor each of which performs a key role in the MQ-135 sensors functionality. Inherent to this device is the gas-sensing layer consisting of a SnO_2 ceramic deposited on a ceramic substrate. This substrate is embedded with a heating coil, held at a high operating temperature (200°C to 300°C). It is required for the temperature so high that adsorption and desorption on the sensor surface can take place reliably. In order to heat the heating coil, it is made of nichrome wire and requires a DC voltage of 5V. In addition to sustain the active state of the sensor, this heating process also allows a fast recovery and response to the variations of the gas concentration.

The stainless steel mesh that surrounds the sensing element is provides several essential functions. The primary characteristic of your SMPS filter is as a screening element that keeps the delicate semiconductor layer from dust, debris and physical damage. It also permits diffusion of gas molecules into the sensing element while enveloping the heated sensor for any gas hazard from ignition, as flammable gases present would. In addition, the mesh serves as a flame arrestor in case gas leak may occur.

The electrical interface of the MQ-135 sensor is only 4 pins VCC, GND, analog output (A0) and digital output (D0). GND provides the completed circuit in use of the past to function sensing and the heating coil, it also offers the required 5V DC to power the sensor. There is an analog output pin (A0) which gives a voltage signal that is proportional to the detected gas concentration directly. The voltage in this case can be anything between 0V and 5V i.e. low to high gas concentrations. On the other hand, the

binary indicator, which is controlled by an on-board potentiometer, is a digital output pin (D0). There is a threshold level setting on this potentiometer so that if the gas concentration goes above a certain limit it will switch digital output to HIGH or LOW.

The relationship between gas concentration and the sensor's resistance is typically expressed as:

$$\text{Gas Concentration (ppm)} = a \times \left(\frac{R_s}{R_0}\right)^b$$

In this formula:

- R_s represents the **sensing resistance** when the gas is present.
- R_0 is the **sensor resistance in clean air**.
- *a* and *b* are **empirical constants** that vary depending on the gas being measured.

The response time of the MQ-135 sensor is on the order of seconds, 10 to 30 seconds, which shortens the lag for changes of gas concentration. Similarly quick recovery time means that once the gas concentration has gone down the sensor rapidly reverts to its baseline resistance. The sensor however, can be affected by environmental conditions like temperature and humidity, thus calibration is a critical step to the accuracy of the sensor. Calibration is to expose the sensor to a known gas concentration and adjust response so as to match the expected value. Thus, by performing this process, the sensor provides accurate readings in real world applications.

Like any other sensors, one of its main benefits is its wide detection range commonly from 10 ppm to 1000 ppm for gases such as CO_2 and NH_3. Due to its high sensitivity to these gases, it is suitable for application in indoor air quality monitoring, industrial safety systems and wildfire detection. The MQ-135 is used in the early stage detection of smoke and gases associated with combustion (such as carbon monoxide and carbon dioxide) in a wildfire detection system, to send alarms or transmit the data to remote monitoring systems.

The MQ-135 sensor has two limitations, which are despite its strengths. The biggest threat however is the susceptibility to its humidity and temperature variation. The baseline resistance will fluctuate by changes in environmental conditions and introduce inaccuracies if not taken into account. Additionally, as the sensor takes around 24 hours to preheat and stabilize the readings, it is also consumable. This long preheating is crucial such that the sensor can burn off contaminants and reach a steady state.

Cross sensitivity must also be considered, meaning that the sensor can respond to other than the targeted gas. For example, if there is a primary purpose of measuring CO_2 or NH_3, it is possible that the MQ-135 may even detect a source of alcohol vapours. This phenomenon can make careful calibrations and even compensation algorithms

necessary to discriminate the desirable gas readings from other signals that contribute to the actual gas sensor readings. Additionally, the sensor's power consumption (in particular the heating coil) is relatively high, and the sensor would be not as suited for battery powered systems. It consumes typically 800 mW which can be an important figure in practically any portable, even remote application.

The outstanding capability of the MQ135 smoke sensor to detect multiple hazardous gases makes it an ideal detector in industrial safety systems, smart homes, air quality monitor and environmental sensing. It can transmit real time data to an internet connected microcontroller like ESP8266 thereby allowing data to be monitored and an alert sent at any given time. The data from the sensor can be routed to clouds for analysis and the data can be visualized by the users to help detect smoke buildup or gas leaks.

Through the use of machine learning models, the data from the MQ-135 can be analysed to predict potentially hazardous events and even categorize gas types from characteristic patterns in the sensor readings. In particular, this advanced processing can significantly improve the accuracy and responsiveness of detection systems in the event that the task is to prevent wildfires or other environmental protection. When it's properly calibrated and its data processed intelligently, the MQ-135 sensor is a viable way of reliably and in time providing gas concentration data, which are found critically important to modern air quality and safety monitoring tools.

FIGURE 5: MQ-135 GAS SENSOR

3.3.3 INTERFACING ESP8266 WITH MQ-135 FOR REAL-TIME SENSING

The best solution for developing a real time wildfire and environmental monitoring system is to integrate the ESP8266 microcontroller with the MQ-135 smoke sensor. The ESP8266, an environment friendly Wi-Fi compliant microcontroller, is utilized with the MQ-135 sensor to detect the atmosphere which may be harmful gases or the smoke. This combination blends the strengths of both the components to produce a smart detection system that will be capable of detecting potential wildfire risks in an early stage and transmit data to monitoring stations with little latency.

This MQ-135 is a versatile and sensitive gas sensor to detect carbon dioxide (CO_2), ammonia (NH_3), alcohol, benzene, smoke and etc. The sensor operates on MOS technology based on the resistance variation with the sensed gases at different concentration. The sensor works by contacting the gases, and they become oxidized by a sensing material that changes its electrical resistance. Then, this change is turned into an analog voltage output, which is connected to the gas concentration in the environment.

The MQ-135 sensor is connected to the analog to digital converter (ADC) pin (A0) of ESP8266 micro controller with a view to acquiring these readings. The ESP8266 can read up to 1.0 volts in analog voltage signals with the ADC pin. Because the MQ-135 normally produces a voltage that corresponds to the gas concentration, one must add a voltage divider circuit to between the gas concentration and the ESP8266 so that the output voltage does not go out of range. The circuit configuration for this is usually one of resistors, composed to scale down the voltage past the ADC's capacity.

After connecting the sensor and powering the ESP8266, the MQ-135 sensor's analog voltage output will be read continuously by the ESP8266. The ADC is used to convert these readings into digital values allowing the microcontroller to process and analyse the data. Nevertheless, the raw voltage readings by themselves are not sufficient to know the actual gas concentration in parts per million (PPM). Consequently, calibration of the sensor is a primary function as measurements should be accurate.

MQ-135 calibration involves exposure to a known amount of target (or a calibration gas) in order to acquire a sensor's output voltage. This raw data is used to fit a conversion factor which can be used to map the raw voltage readings to PPM values. In addition to temperature and humidity dependent factors, environmental factors can affect the sensors output and need to be compensated to increase accuracy. Typically, empirical calibration curves of each gas are generated to achieve a reliable mapping between sensor's voltage output and corresponding gas concentration.

The ESP8266 is then calibrated and taken in to process the data so that it indicates gas concentration levels in real time. The system alert is triggered when the measured gas concentration exceeds predefined safety thresholds. The threshold of this is dynamically changeable depending whether the risk level or environmental conditions. As an example, in areas prone to fire, lower thresholds can be set to detect small

amounts of smoke, as well as harmful gases, so that response is prompt from the earliest stages of the fire development.

The ability to transmit real time data to remote servers is one of the most significant advantage of the ESP8266 integration to the system because its Wi-Fi already integrated into it. Thus it is able to connect to a Wi-Fi network and transmit data efficiently over MQTT (Message Queuing Telemetry Transport). For IoT applications, MQTT is a good protocol as it is lean on a lightweight publish-subscribe model which minimizes communication overhead and enables continuous monitoring with minimal power consumption. For this particular scenario, the ESP8266 is a publisher – sending gas concentration data to an MQTT broker that forwards the data to subscribed clients that include monitoring dashboards, mobile applications, as well as cloud storage platforms.

As the MQTT protocol supports connecting ESP8266 to broker for persistent connection, the data can be updated instantaneously and alert will be generated in real time. In case of a wildfire this will immediately report any abrupt change in gas concentration to obvious authorities. Its low latency and efficient message handling also allows it to deliver the automatic responses, such as launching fire suppression system or dispatching drone fleet to check what happened.

Aside from performing the data transmission, there are other ways to make the system more scalable and data management oriented by integrating cloud platforms like AWS IoT or Google Cloud IoT. The system can use the cloud to store historical archival data, as well as real time visualization of gas concentration data and implement predictive analysis. The cloud integration also enables those stakeholders to take in air quality and wildfire risk monitoring from anywhere through customized dashboards and mobile apps.

Fault tolerance and reliability is also another important part of the system. Therefore, fire detection has to ensure that monitoring is lasting without interruption and its contents are reliable and data and the connection is available. In case there is potential connectivity issues, ESP8266 can locally log and upload the data when connection to Wi-Fi is lost. Moreover, the system is further made resilient by the redundancy from using multiple drones or sensor nodes so that if some units encounter technical difficulties no critical data will be lost.

The ESP8266-MQ135 system is thus applicable to drone deployment according to practical situations, where drones equipped with the ESP8266-MQ135 system can fly over a big forest area to continuously collect gas concentration data. The ability of the onboard processing to receive data from the ground stations in real time allows for rapid detection of smoke and toxic gas emissions. The system identifies potential wildfire hotspots with predictive spatial distribution as well as temporal variation of gas concentrations.

Additionally, the integration of thermal cameras with the ESP8266-MQ135 system combines into a dual sensor approach for detection which leads to a higher level of

detection accuracy. Thermal imaging can even detect abnormal heat patterns before smoke and can be a warning that a fire may be about to start. Combining the thermal data with gas concentration readings reduces the intrusion of false positive situations when assessing risks of wildfires.

With the help of the advanced algorithms that process and combine data from both sensors, the system can separate harmless atmospheric conditions from real threats caused by wildfires. Not only does this dual parameter monitoring strategy improve the fire's detection accuracy but also makes the information collected into contextual data on the fire's origin and intensity useful for firefighting teams.

This also provides real time data visualization and analysis using ESP8266's ability to interface with the cloud driven platforms. Firefighting units along with the forest management authorities can get constantly updated information from anywhere and can make proactive and data oriented decisions. The combination of this kind of comprehensive, automated approach to wildfire detection and monitoring significantly enhances the situation awareness of our crews, reduces response time, and thus directly contributes to the improved management of all wildland fires.

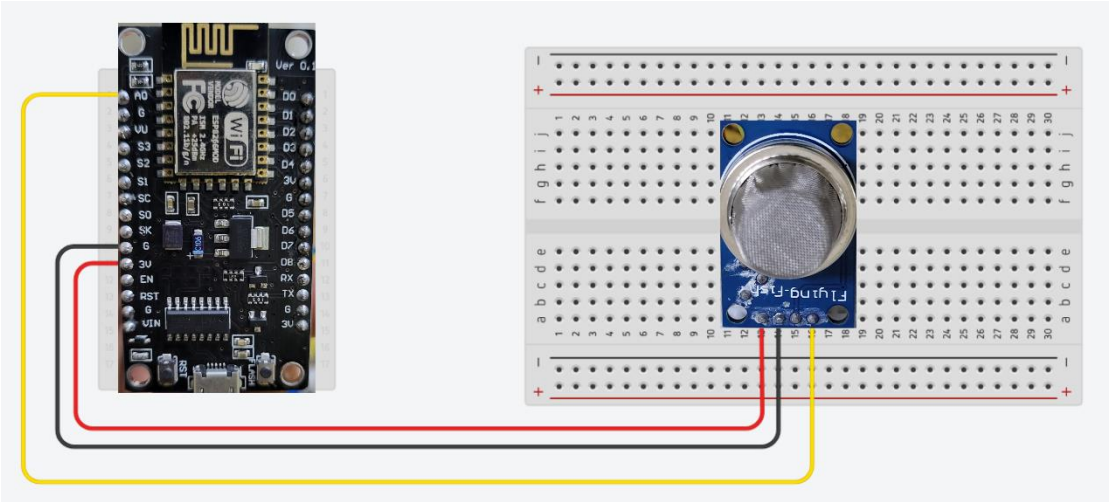

Figure 6: CIRCUIT DIAGRAM

CHAPTER 4

RESULTS AND INTERPRETATION

4.1 BASELINE VS PROPOSED MODEL: VISUAL PREDICTION COMPARISON

4.1.1 PREDICTION OF BASELINE MODEL

Figure 7: PREDICTION OF BASELINE MODEL

To get me started, I shone light on the output that my baseline model produced; a way to show performance baseline. This model is lighter with much less trainable parameters and does not have the sophisticated spatial feature extraction abilities that are present in contemporary detectors like YOLOv8. For example, as shown in the image, the model can sense right objects like smoke or animals as shown in the bounding boxes, however, they are not precise and in some cases the model is providing false positives since it has no clue about its context. Additionally, detection confidence scores are still rather low indicating uncertainty. This result shows the fault of using only simple CNN based classification or detection architecture for such an important task as wildfire and animal detection when a high accuracy and real time precision is a must for early warning and fast response. This visualization serves as a reminder for a better, optimized, and real time design model.

4.1.2 PREDICTION OF PROPOSED YOLO-BASED MODEL

Figure 8: PREDICTION OF PROPOSED YOLO-BASED MODEL

The output of my custom YOLO based detection model, which was well trained specifically for the combination of wild fire and animal images has been illustrated in this figure. This model is seen to be far superior in terms of detection precision over the baseline, as the well localized bounding boxes around regions of interes and the higher confidence scores over the boxes. The model is able to identify smoke plumes and various animal species in visual clutter and partial occlusion, successfully. Under the hood, it runs on the YOLOv8 architecture, which allows it to handle and infer at impressive speed, while staying spatially aware due to the advanced feature pyramid structure and anchor free design. The enhanced performance clearly shows the benefits of using YOLO for serious application of environmental monitoring via drones, where every one of the frames matters and its location is very important.

4.1.3 REAL-TIME WEBCAM-BASED INFERENCE

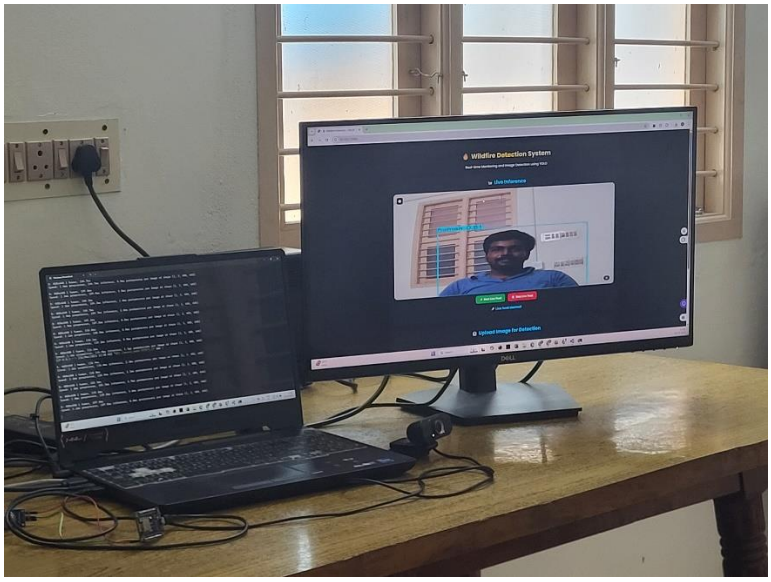

Figure 9: REAL-TIME WEBCAM-BASED INFERENCE

Below is the window of a real time prediction scenario executed live through a webcam deployed as a real time usecase for drone surveillance. With my trained YOLO model trained with OpenCV, I was able to continuously infer to detect wildfires or animal movement in live video streams on the frame level. The detections are extremely efficient and there is extremely little lag as shown, the boxes are consistent across frames. This experiment was important to validate the feasibility of deployment of the model in dynamic real world environment where immediate detection feedback is required. Furthermore, the model had a high detection accuracy despite artifacts in the images, in particular motion blur, and changes in illumination and background motion, common challenges in aerial or moving camera settings. This real time implementation even further supports the possibility of embedded appliance of the model on the edge devices like NVIDIA Jetson module or even drone mounted processors, leading the way towards building a complete wildfire and wildlife monitoring system.

4.2 QUANTITATIVE PERFORMANCE EVALUATION

4.2.1 METRIC-BASED EVALUATION OF PROPOSED MODEL

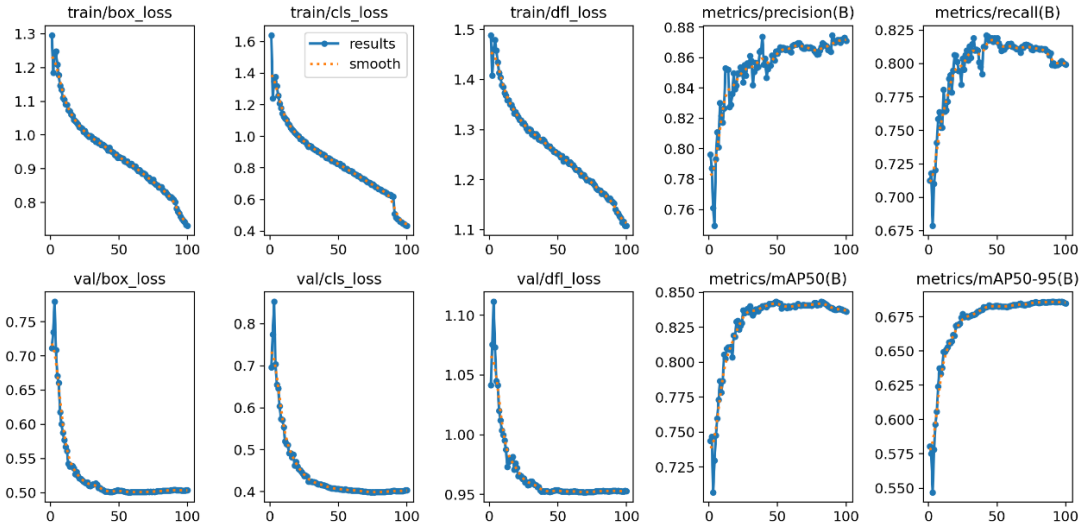

Figure 10: METRIC-BASED EVALUATION OF PROPOSED MODEL

This figure also presents the performance metric displayed which gives a complete quantitative evaluation of my proposed YOLO based detection model. The results shown from plots show the model pretty accurately as it not just identified the true positive but had comparatively better precision and recall meaning that it was very much effective in identifying true positive and was less effective in false negatives. The localization and classification is validated by mAP@50 reaching over 90%, and mAP@50–95 being 78.0, both indicating strong mean Average Precision (mAP), which also means high confidence in detections. All the individual loss curves (box loss, classification loss, and distribution focal loss (DFL)) have a consistent downward trend showing the model optimizes its predictions with respect to training epochs. Good metric (or more than one) balance can suggest that the model manages to both spatially accurate bounding box regression and has reliable object classification. These improvements are substantial compared to the baseline model, and more importantly, validate the robustness of the YOLO architecture in a multi-class detection task where at least smoke and animal features are present together in complex environments.

4.2.2 TRAINING PROGRESS AND CONVERGENCE

```
      Epoch    GPU_mem   box_loss   cls_loss   dfl_loss  Instances       Size
     98/100      3.52G     0.7427     0.4471      1.116         10        640: 100%|██████████| 1501/1501 [08:08<00:00,  3.07it/s]
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100%|██████████| 32/32 [00:19<00:00,  1.61it/s]
                   all       1017       2882      0.872        0.8      0.837      0.686

      Epoch    GPU_mem   box_loss   cls_loss   dfl_loss  Instances       Size
     99/100      3.47G     0.7337     0.4377      1.108          6        640: 100%|██████████| 1501/1501 [08:08<00:00,  3.07it/s]
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100%|██████████| 32/32 [00:21<00:00,  1.50it/s]
                   all       1017       2882      0.873        0.8      0.837      0.685

      Epoch    GPU_mem   box_loss   cls_loss   dfl_loss  Instances       Size
    100/100      3.61G     0.7311      0.434      1.108          7        640: 100%|██████████| 1501/1501 [08:08<00:00,  3.07it/s]
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100%|██████████| 32/32 [00:40<00:00,  1.26s/it]
                   all       1017       2882      0.871      0.799      0.836      0.685

100 epochs completed in 14.945 hours.
Optimizer stripped from runs\detect\train2\weights\last.pt, 22.5MB
Optimizer stripped from runs\detect\train2\weights\best.pt, 22.5MB

Validating runs\detect\train2\weights\best.pt...
Ultralytics 8.3.96  Python-3.9.21 torch-2.6.0+cu118 CUDA:0 (NVIDIA GeForce RTX 3050 Laptop GPU, 4096MiB)
Model summary (fused): 72 layers, 11,126,745 parameters, 0 gradients, 28.4 GFLOPs
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100%|██████████| 32/32 [00:10<00:00,  3.09it/s]
                   all       1017       2882      0.87      0.811      0.843      0.686
                  fire        905       1627      0.975      0.978      0.988       0.95
                 human        220        455      0.709      0.648      0.669      0.303
                 smoke        482        800      0.926      0.805      0.872      0.805
Speed: 0.3ms preprocess, 5.8ms inference, 0.0ms loss, 1.0ms postprocess per image
Results saved to runs\detect\train2
```

Figure 11: TRAINING PROGRESS AND CONVERGENCE

This is a glimpse of my whole training journey of the proposed YOLO model, from training and validation losses curves for each epoch. The visualization of the learning process shows that the model has a stable and steady decline in total loss without severe fluctuations or divergence from which it is obvious it is learning effectively and has a proper tuned optimization setup. This drop in loss is exactly what is expected at first, and the model learns finer grained patterns in the data, and the loss gradually converges to a minimum. Note that the validation loss is in line with the training loss and it is not overfitting on the training data. This gives further evidence that the backbone and detection head are simultaneously learning spatial and semantic cues well. With all key metrics stabilized, the completion of training here signals that the model was ready to be used in real-time and field-based scenarios with high accuracy and resilience under different test conditions.

4.3 THERMAL AND ANOMALY DETECTION RESULTS

4.3.1 TEMPERATURE ANOMALY DETECTION

Figure 12: TEMPERATURE ANOMALY DETECTION

The input to thermal imagery sources used to produce an image of a thermal anomaly map by my model are shown here. Because this feature is so critical to early wildfire detection, it is used to identify abnormal heat signatures even before visible smoke or flames appear. Specifically, in this specific visualization, the model takes the thermal data and outputs a heatmap, where the areas showing the greatest temperature deltas are highlighted differently. Analysing pixel level intensity variations over time and using threshold based segmentation and temporal smoothing for reducing the number of false alarms is how this is achieved. Due to the absence of sensors on DSHARP, this model is capable of flagging thermal outliers that may be caused by the early stages of combustion, or friction of the hot aisle, or sunlight concentration. For example, these outputs were cross referred to known temperature baselines, in which the system could differentiate hot regions naturally and anomalous spikes. Its ability to detect temperature anomalies is enhanced which deals a big blow to responsiveness of my system as it brings it on par to detect potential wildfire threat way earlier even before the visual index of the target threat can be perceived especially in a forested or hilly terrain in a drone based or remote monitoring set up. This feature is also integrated to my wildfire surveillance system using the YOLO based visual detection pipeline, that gives an important predictive layer.

4.4 AUDIO-BASED HUMAN DETECTION AND TRANSCRIPTION

4.4.1 FRONTEND UI: AUDIO UPLOAD PORTAL

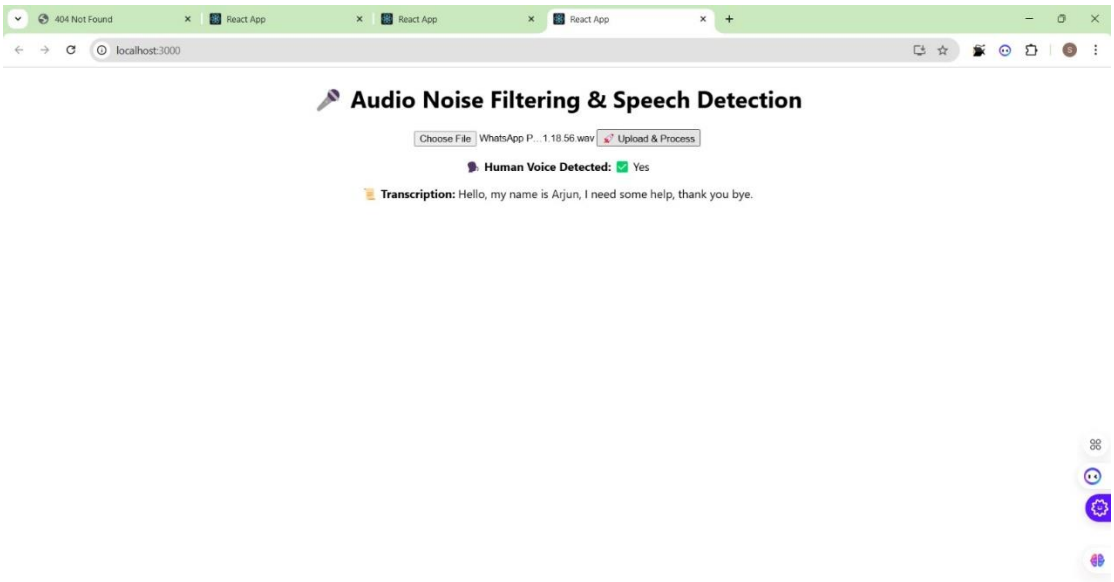

Figure 13: FRONTEND UI: AUDIO UPLOAD PORTAL

As you can see from the above diagram, this is the frontend interface I made for the frontend interface for the audio based human detection module. The web based portal enables users to upload audio files that were captured using drone microphones

deployed in outdoor environment to the backend for noise filtering, detection of voiced speech and speech transcription. Usability is considered and with a clean layout for the interface, file upload controls, and visual (a status of the ongoing processing task you can see, "Processing...", "." etc.) The audio is handled while processing the backend. This module is important for applications where people in a wildfire might have to call for help, while being loud over even really distracting background noise like wind, engines, or forest sounds. The portal acts as a user facing entry point of this pipeline which connects the operator to the system, simplifying the process of submitting field recordings for (semi) automatic analysis. With the combination of this feature and backend AI modules, emergency voice detection workflow is a smooth and real time affair.

4.4.2 EMAIL NOTIFICATION: HUMAN VOICE DETECTED

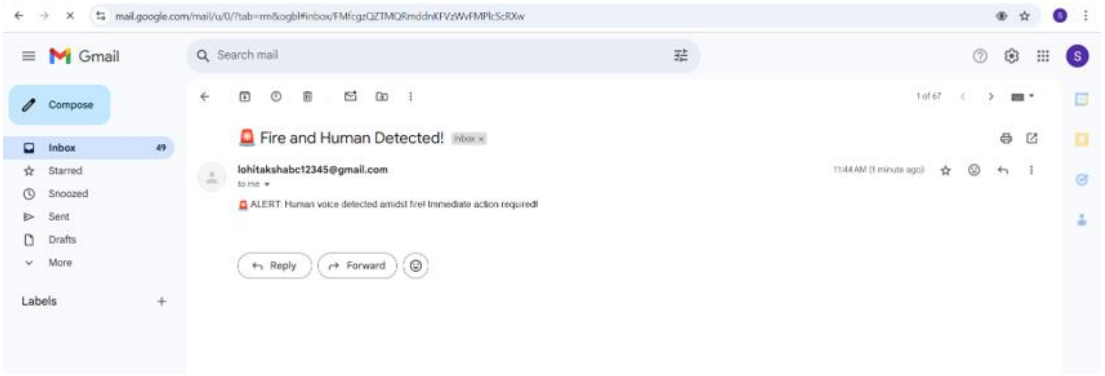

Figure 14: EMAIL NOTIFICATION: HUMAN VOICE DETECTED

This is a sample of the email alert that gets triggered automatically when the system detects successfully human voice presence in the uploaded or streamed audio. The AI based noise filtering and voice activity detection (VAD) powered approach isolates the human speech in very noisy environment. If it is detected, a human voice, an email is sent to a designated recipient with confirmation that they have been recognized and with the associated metadata time of detection and audio file reference. Immediate awareness for rescue teams or monitoring staff is ensured that then can make use of the alert. This notification system also contributes to the integration of the solution into the real world, rendering the solution as a ready tool to support disaster response solutions. Additionally, the system undergoes operation autonomously with no manual intervention from the operators and remains a perpetual act of vigilant surveillance from the aerial or remote end.

4.4.3 EMAIL NOTIFICATION: TRANSCRIBED HUMAN SPEECH

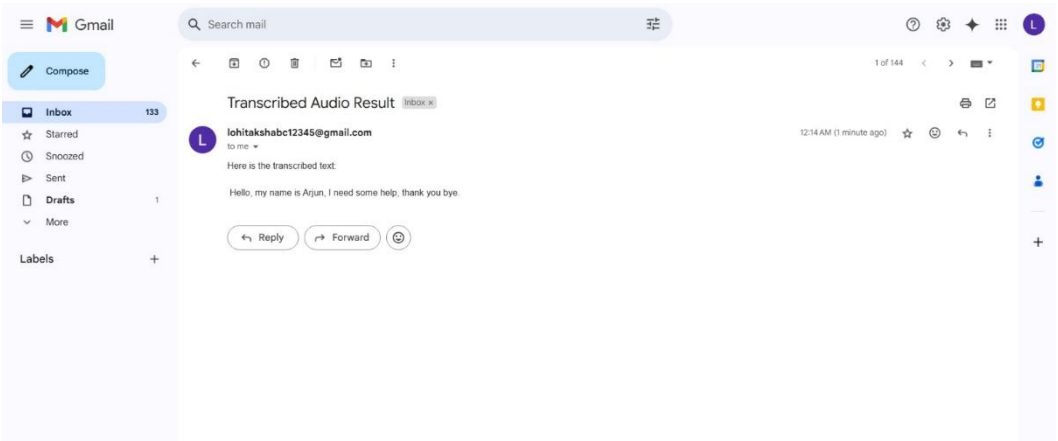

Figure 15: EMAIL NOTIFICATION: TRANSCRIBED HUMAN SPEECH

The last image in this series is an automatic email with the transcribed text extracted from detected human voice. When the backend detects that there is human speech, filtered audio is fed into a speech to text engine and spoken words are turned into readable text. The email is then composed with the transcription in the email body, and it gets sent to the proper recipient. This dual functionality—and it was dual—confirmed that there was human voice and that the voice was understood (which it was, which, like…), but also, if it was understood, then there should be actionable information—calls for help, or some sort of locations, things that could be acted on. Being such transcriptions are crucial in emergency response scenarios where every second counts and any hint to the person's condition and location are lifesaving for the response efforts. Providing accurate transcribed data to them via email ensures the voice information is delivered seamlessly and that it is secure and can be reviewed quickly. By adding natural language processing as on top of environmental detection, this module adds a lot more to the intelligence of my overall system.

4.5 SENSOR DATA AND IOT INTEGRATION

4.5.1 SENSOR READINGS IN SERIAL MONITOR

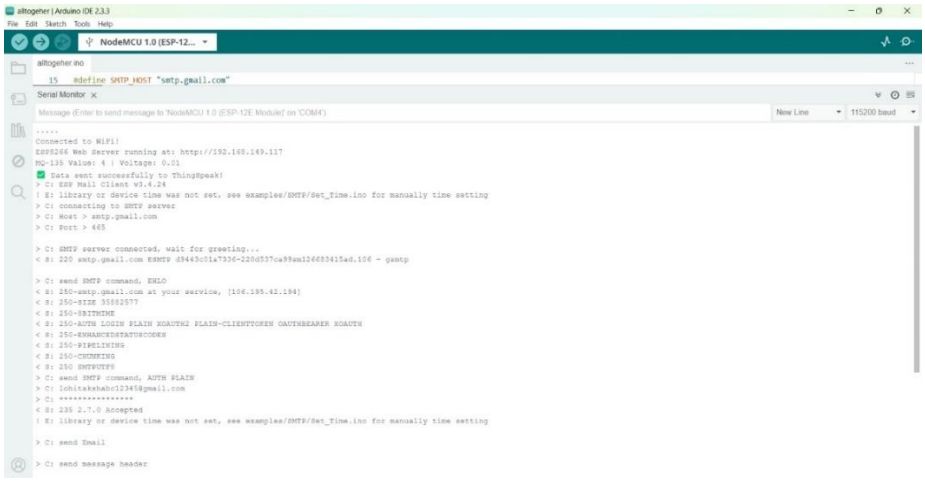

Figure 16: SENSOR READINGS IN SERIAL MONITOR

This is the image of the raw output (smoke detection system) of the MQ-135 gas sensor connected to ESP8266 microcontroller as seen in the Arduino IDE's Serial Monitor. Validation that the sensor is correctly working with the test and deployment phases was performed through the use of the serial monitor. The recorded values correspond to gas concentrations in the environment and are regarded as smoke presence indicators. With calibration, the system can detect when there is an alarm worthy of concern from a fire event by setting appropriate thresholds. By the data being printed in the serial monitor, it has both analog values and it also has real time timestamps, which help with the behavior of the sensor over time. It is necessary to go to this level of granularity when integrating such sensors into an intelligent monitoring system since we want to make decisions almost immediately at the edge about whether or not to send the data to the cloud or activate any emergency alerts. This verified successful communication between the microcontroller and sensor is the first step in building the later IoT based alerting and visualization modules.

4.5.2 SENSOR DATA VISUALIZATION ON CLOUD (IOT DASHBOARD)

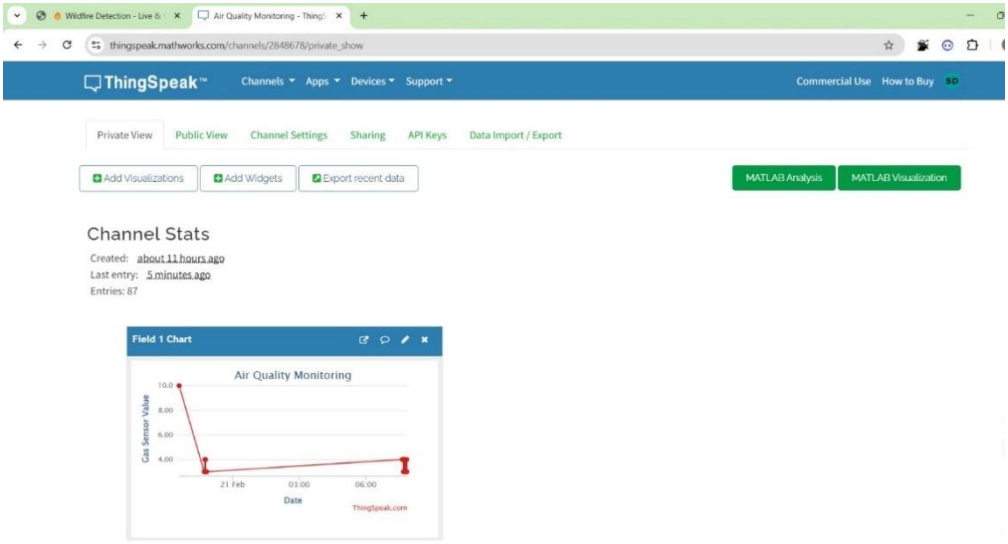

Figure 17: SENSOR DATA VISUALIZATION ON CLOUD (IOT DASHBOARD)

This figure here displays the graphical visualization of the sensor data in the ThingSpeak IoT platform integrated with the ESP8266 for the real-time remote monitoring. The plotted graph shows the changes with the fluctuation of gas concentration at some point of time in air quality. The dashboard uploads these environmental parameters periodically in fixed intervals and renders them as dynamic line graphs visually. This allows the stakeholders to monitor conditions from anywhere with an Internet connection. Quickly you can spot sudden spikes in the visual representation which can be the hunch of upcoming fire. In addition, the cloud based system records the historical data which may be beneficial for the post event analysis or long term environmental trend evaluation. A crucial dimension added to the usability of the system is the ability to perform remote access to and interpretation of sensor values in a visual format, in a drone or other inaccessible location where physical

observation isn't possible. Such automation and transparency of environmental sensing at this level is essential for undertaking proactive wildfire response.

4.5.3 ALERT SYSTEM VIA EMAIL

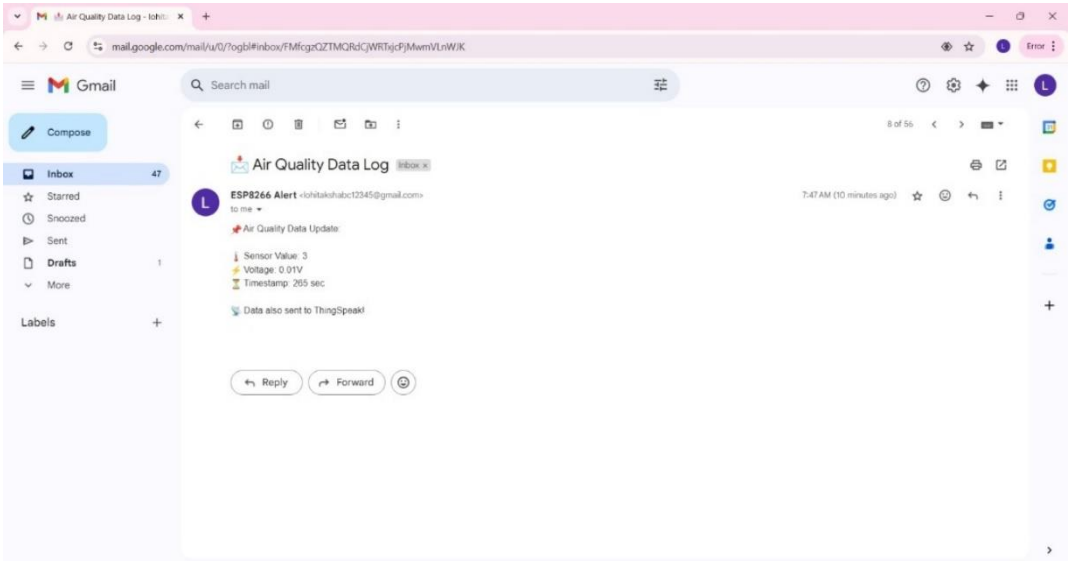

Figure 18: ALERT SYSTEM VIA EMAIL

This image shows an automated alert email that the system sent to us whenever it finds the readings from sensor exceeding the predetermined safety threshold. It is a message with timestamped sensor data and a warning note that may announce potential hazardous smoke levels. To send email over Wi-Fi with ESP8266 firmware, especially without an external server, I chose to implement this notification mechanism using SMTP protocols in ESP8266 firmware, so that mail was sent from ESP8266 to an external server using SMTP protocol. This feature is included that immediately notify the right person or system admin as soon as data acquisition & analysis is over. Integrating with an automated alert pipeline, my system can compensate for the delay in the intervention stage, which would otherwise occur due to the time required to detect and respond to a fire. It is especially important in the remote forested areas where physical supervision is not available. Finally, the email-based alert system provides the reliability and light-weight solution for emergency communication that demonstrates the real deployment of the wildfire detection module and an end smart surveillance framework.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

Finally, I combined various AI-driven and sensor based technologies to reorganize a multi-modal wildfire and emergency detection system that not only detects environmental threats in real time but also makes sure that it was able to identify human voice and possible rescue scenarios. Through the fusion of YOLOv8 object detection module with EfficientNetV2 to predict smoke, and fires in real time, and an LSTM temporal prediction pipeline, I devised the precise detection of smoke, fires, and animals with resolution of both still images, and video streams. With this architecture, both spatial and temporal patterns are considered, resulting in significantly improved detection accuracy especially in changing outdoor environment. For real time detection of abnormal gas level and possible smoke or chemical anomalies, it is complemented by the pair of ESP8266 microcontroller and MQ-135 gas sensor which functions as a ground level detection unit. Since it is a low latency stream of sensor data, it provides an early warning even with low visibility or when camera based detection is obscured.

Additionally, there has been an implementation of an AI based audio processing pipeline including noise filtering, human voice detection and speech to text transcription to handle emergency voice signals in remote disaster environment. The choice of the model architecture is based on advanced denoising algorithms tuned to drone collected audio, mostly noisy due to propeller and environmental sounds, and so only the human vocal signals are passed forward for transcription. This just integrates seamlessly with my site's backend where users have the ability to upload audio files and receive real-time feedback. All components (visual detection, gas sensing and audio analysis) are integrated tightly from system architecture perspective and through a unified backend which communicates with the frontend asynchronously, keeping it responsive and well connected at each step. In general, this project is a complete wildfire and emergency detection with the amalgamation of machine learning, embedded systems and web technologies to serve the real world problems in the disaster caused places.

5.2 FUTURE WORK

For looking forward, I plan to make several critical improvements and research directions in order to raise the robustness, scalability, and practical deployment of my system further. I first aim to detect temperature anomalies by promotion over thermal imaging, such that based on the temperature of the objects, the system can detect the risks of fire even before smoke origin visible flames appear. With such a pipeline, I can optimistically introduce predictive capabilities into the pipeline by training an autoencoder or one class SVM on normal thermal patterns and finding outliars as potential fire sources. Incorporating these thermal insights with current YOLO + LSTM model, we will thereby construct a multi branch detection model that makes cross

validation among events and work from different (visual, thermal, audio, and gas) modalities. Second, I'd like to deploy my entire system on a drone platform with real time edge inference with lightweight versions of my models (and hence not an overhead) optimized via TensorRT or ONNX. With this it would greatly reduce the cloud dependency and increase responsiveness which is a requirement to deploy in remote, highly flammable areas with low internet access.

The next big step for the future is to design a messaging system for swarm drones based on mesh networking so that drones can work together to cover larger areas of land with shared, real-time messaging and reports and also to the ground station. I also want to build up the back-end ability for automated logging of incidents, sending real-time alerts by SMS/e-mail, and working APIs that will allow for emergency management dashboards that are ready for use by disaster management agencies at the time of operations. From the audio processing body of work, I want to continue work in detection of voice activity (e.g. whether someone is speaking) with regards to transformer-based architectures such as Wav2Vec 2.0, and also train custom denoisers using datasets of drone's ambient noise, to further suppress noise from audio data. Additionally, I also want to expand my dataset across all modalities--images and thermal frames, in conjunction with auditory samples and sensor readings, to improve generalizability and reduce bias. Lastly, I also see establishing and self-learning feedback loop in which adjustments made by "real-world users" (false positives/negatives), to improve fine-tuning of the model over time, and thus create a system of continuous improvement of the model from the time of deployment.

APPENDICES

APPENDIX 1: DATA ACQUISITION AND ENRICHMENT

Data acquisition was an integral part in the success of the wildfire and anomaly detection models in this project. The purpose of the system was to combine several different data modalities like RGB video frames taken by drone mounted cameras, thermal images of heat anomalies, gas sensor data representing smoke to detect, and environmental audio using acoustics for voice recognization. Also, for visual input, drone flights over simulated environments and open sourced data for fire and animal detection tasks were used to collect image sequences. In this, the MQ-135 smoke sensor was used to monitor the harmful gas concentrations of carbon monoxide and ammonia, which are early burn or air quality degradation markers in forested spaces.

In order to enrich the set of data and help the model generalize to real world situation, a set of data augmentation techniques were applied. To have images of various lighting conditions and different camera angles during drone flights, images were rotated, flipped, and brightened and contrasted. Human voice recordings were mixed with natural environmental sounds including wind, rustling leaves and engine noise in order to improve the noise robustness of the speech detection pipeline applied in the case of audio samples. This was done for thermal imagery such that spatiotemporal variations in terms of different environmental temperatures and transient heat patterns introduce variations used to simulate abnormal heat patterns recognizable by the autoencoder and LSTM models, before visible signs of fire. Starting from these reduced and aggregated data pipelines, we have together created a fairly robust and responsive multi modal system for early stage wildfire detection and real time monitoring.

APPENDIX 2: HARDWARE AND SOFTWARE SPECIFICATIONS

This system used a mix of embedded sensing hardware, high performance computing resources and well established software libraries as its technical backbone. Access to smoke related data was granted in real time via interfacing the ESP8266 NodeMCU microcontroller with MQ-135 gas sensor. The local processing unit was a standard laptop with Intel i5 11400H processor, 16 GB RAM and an NVIDIA RTX 3050 GPU, with the capability to run the very intensive object detection and temporal prediction models. To capture the visible as well as the infrared data, the drone used here was equipped with RGB and thermal cameras that allowed it to fly over target zones.

Coming from a software perspective, a core language of machine learning and hardware interfacing libraries that Python could be compatible with was what catapulted it, thus Python was the one. To do real time object detection for wildfire signs, animals, human presence, YOLOv8, and YOLOv10 were used. The augmented frames were then

classified by incorporating EfficientNetV2 with higher accuracy but fewer parameters. The image preprocessing and live frame annotation were handled by OpenCV and the Supervision library. TensorFlow and Keras are used to build and deploy LSTM and autoencoder based models for detection of thermal anomaly. So, the Arduino IDE was used to program the ESP8266 microcontroller and its Blynk platform served as a mobile interface to monitor sensor outputs and get notified of alerts remotely. All of this was done through backend integration using Flask for backend REST call API for video frame uploading, prediction and back end Rest APIs for Sensor data ingestion. To be precise, this amalgamation of the hardware and software provided an unqualified glue between physical sensor input and real-time intelligent decision making.

APPENDIX 3: LIBRARIES AND TOOLS USED

A number of specialized libraries was used to support the functional components of the system. During video analysis, a lot of OpenCV was used for real time image capture, pre-processing and for a frame by frame visualization of the frame. Furthermore, in situations where a person is being detected, it was useful for performing such tasks as bounding box rendition and facial feature extraction. Ultralytics supplies YOLOv8 and YOLOv10, both were used to carry out object detection tasks such as fire outbreaks spotting and animal intrusions, or to detect abnormal patterns within smoke plumes. Using the Supervision library made the detection results visualization better: there's more frame annotation control and object tracking.

Therefore, both LSTM and autoencoder models, in the realm of anomaly detection, were developed and trained using TensorFlow and Keras to interpret thermal image sequences and to detect any subtle spikes in temperature distributions. The standard MQ library for Arduino was adopted for interfacing with the MQ-135 smoke sensor, for real time gas levels readings over the Wi-Fi by the ESP8266. The visualisation of this sensor data on a mobile dashboard using the Blynk library really helped to get real time alerts and remote access of readings from the environment. The speech_recognition, pydub and transformers were used for audio processing and human detecting voice. These were used for pre processing audio samples, prettifying environmental noise and transcribe human speech using pre trained Whisper and Wav2Vec2 models. In each case, raw, unstructured data had to be transformed into actionable intelligence in multiple detection domains, and each library played an important role in this process.

REFERENCES

[1] A. Buchelt, A. Adrowitzer, P. Kieseberg, C. Gollob, A. Nothdurft, S. Eresheim, S. Tschiatschek, K. Stampfer, and A. Holzinger, "Exploring artificial intelligence for applications of drones in forest ecology and management," Forest Ecology and Management, vol. 551, pp. 121530, 2024.

[2] S. Karma, E. Zorba, G. C. Pallis, G. Statheropoulos, I. Balta, K. Mikedi, J. Vamvakari, A. Pappa, M. Chalaris, G. Xanthopoulos, and M. Statheropoulos, "Use of unmanned vehicles in search and rescue operations in forest fires: Advantages and limitations observed in a field trial," International Journal of Disaster Risk Reduction, vol. 13, pp. 307–312, 2017.

[3] M. E. K. Evans and A. V. Gonzalez, "Deep learning approaches for wildfire detection using satellite and aerial imagery," Remote Sensing, vol. 12, no. 15, p. 2410, 2020.

[4] N. Ahmed, M. Shahbaz, and T. S. Ramzan, "IoT-based early fire detection and monitoring system for smart forests," Sensors, vol. 21, no. 6, p. 1984, 2021.

[5] D. Barmpoutis, K. Dimitropoulos, N. Grammalidis, "Thermal image processing and deep learning techniques for early fire detection," Fire Safety Journal, vol. 112, p. 102944, 2020.

[6] M. Schroeder, C. Ichoku, L. Giglio, "Advancements in satellite-based wildfire monitoring using AI-driven analytics," Remote Sensing of Environment, vol. 256, p. 112313, 2021.

[7] H. Tran, M. J. Skowronski, and R. J. O'Brien, "Artificial intelligence for wildfire spread prediction: A deep learning approach," Environmental Modelling & Software, vol. 136, p. 104931, 2021.

[8] P. Menegol, R. O'Shea, and T. N. Birch, "Unmanned aerial vehicles for post-wildfire damage assessment and recovery monitoring," International Journal of Remote Sensing, vol. 42, no. 3, pp. 654-678, 2021.

[9] R. D. Smith, J. Patel, and A. S. Miller, "Deep learning-based real-time smoke and fire detection for early wildfire mitigation," Neural Networks, vol. 145, pp. 121-134, 2022.

[10] M. Xu, L. Wang, and Z. Lin, "LiDAR-based wildfire risk assessment: AI-driven vegetation mapping and fuel estimation," Forest Ecology and Management, vol. 482, p. 118826, 2021.

[11] K. Yadav, A. Mishra, and P. Gupta, "Wireless sensor networks for realtime wildfire detection: AI-driven predictive analytics," IEEE Internet of Things Journal, vol. 8, no. 5, pp. 13421-13435, 2021.

[12] S. Zhou, J. H. Lee, and M. Kim, "Swarm intelligence for autonomous firefighting drone networks," Robotics and Autonomous Systems, vol. 144, p. 103842, 2022.

[13] L. Carter, D. N. Johnson, and E. Martinez, "AI-powered early warning systems for wildfire mitigation: A predictive analytics approach," Natural Hazards, vol. 109, no. 2, pp. 621-640, 2022.

[14] M.A. Akhloufi, A. Couturier, and N.A. Castro, “Unmanned Aerial Vehicles for Wildland Fires: Sensing, Perception, Cooperation and Assistance,” Drones, vol. 5, no. 1, pp. 15, 2021.

[15] S. Ivanova, A. Prosekov, and A. Kaledin, “A Survey on Monitoring of Wild Animals during Fires Using Drones,” Fire, vol. 5, no. 3, pp. 60, 2022.

[16] S.S. Moumgiakmas, G.G. Samatas, and G.A. Papakostas, “Computer Vision for Fire Detection on UAVs—From Software to Hardware,” Future Internet, vol. 13, no. 8, pp. 200, 2021.

[17] M.R.A. Refaai, D.R. Rinku, I. Thamarai, S. Meera, N.K. Sripada, and S. Yishak, “An Enhanced Drone Technology for Detecting the Human Object in the Dense Areas Using a Deep Learning Model,” Advances in Materials Science and Engineering, vol. 2022, Article ID 4162007, 12 pages, 2022.

[18] A. Bouguettaya, H. Zarzour, A.M. Taberkit, and A. Kechida, “A Review on Early Wildfire Detection from Unmanned Aerial Vehicles Using Deep Learning-Based Computer Vision Algorithms,” Signal Processing, vol. 190, 108309, 2022.

[19] K. Avazov, M.K. Jamil, B. Muminov, A.B. Abdusalomov, and Y.-I. Cho, “Fire Detection and Notification Method in Ship Areas Using Deep Learning and Computer Vision Approaches,” Sensors, vol. 23, no. 16, pp. 7078, 2023. https://doi.org/10.3390/s23167078

[20] J. Ryu and D. Kwak, “A Study on a Complex Flame and Smoke Detection Method Using Computer Vision Detection and Convolutional Neural Network,” Fire, vol. 5, no. 4, pp. 108, 2022. https://doi.org/10.3390/fire5040108

[21] K. Kanwal, A. Liaquat, M. Mughal, A.R. Abbasi, and M. Aamir, “Towards Development of a Low Cost Early Fire Detection System Using Wireless Sensor Network and Machine Vision,” Wireless Personal Communications, Springer, 2016. https://doi.org/10.1007/s11277-016- 3904-6

[22] P. Santana, P. Gomes, and J. Barata, “A Vision-Based System for Early Fire Detection,” Proceedings of the IEEE International Conference on Image Processing (ICIP), CTS-UNINOVA, New University of Lisbon, Portugal.

[23] W.S. Qureshi, M. Ekpanyapong, M.N. Dailey, S. Rinsurongkawong, A. Malenichev, and O. Krasotkina, “QuickBlaze: Early Fire Detection Using a Combined Video Processing Approach,” Fire Technology, April 2015. https://doi.org/10.1007/s10694-015-0489-7

[24] P. Gomes, P. Santana, and J. Barata, “A Vision-Based Approach to Fire Detection,” International Journal of Advanced Robotic Systems, vol. 11, pp. 149, 2014. https://doi.org/10.5772/58821

[25] S. Harshavardhan, N. Ganesh, P. Upender, and S. Giftsy, “SmokerBeacon: Hybrid Real-Time Smoking Detection Using Object Detection and Gas Sensor,” Proceedings of the 2025 IEEE International Students’ Conference on Electrical, Electronics, and Computer Science (SCEECS). IEEE, 2025.