TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PURWANCHAL CAMPUS

A MAJOR PROJECT PROPOSAL ON PROJECT TITLE

SAFEZONE AI : SMART SURVEILLANCE FOR URBAN AND RURAL SAFETY

SUBMITTED BY

BIBISHA BASNET (PUR078BCT022)

BIMLENDRA PANDIT (PUR078BCT026)

MANOHAR JHA (PUR078BCT047)

RESHMI JHA (PUR078BCT066)

SUBMITTED TO

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

PURWANCHAL CAMPUS

DHARAN, NEPAL

JUNE,2025

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to everyone who supported and guided us throughout the development of this project. Special thanks to our supervisors, faculty members, and peers whose insights, suggestions, and encouragement were invaluable in the successful completion of this work.

This project, titled “SafeZone AI: Smart Surveillance for Urban and Rural Safety,” has been a challenging and rewarding journey. It provided us with the opportunity to work with real-time video and audio analysis, explore advanced AI technologies, and develop a system aimed at enhancing public safety. We are truly thankful for the learning experience and for being able to transform an ambitious idea into a meaningful and practical application.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

API	: Application Programming Interface
CCTV	: Closed-Circuit Television
YOLOv8	: You Only Look Once version 8
AI	: Artificial Intelligence
GPS	: Global Positioning System
SMS	: Short Message Service
CNN	: Convolutional Neural Networks
PC	: Personal Computer
SQLite	: Structured Query Language Lite
CSV	: Comma-Separated Values
SMTP	: Simple Mail Transfer Protocol
USB	: Universal Serial Bus
SSD	: Single Shot MultiBox Detector

CHAPTER 1

INTRODUCTION

1.1 Background

In today's world, keeping people safe in both cities and villages is very important. Many public places like roads, markets, schools, and parks use CCTV cameras to watch what is happening. But these cameras only record videos — someone has to keep watching them all the time to notice if something goes wrong. This is difficult and sometimes important events like accidents, crimes, or fires are missed.

To solve this problem, we will create SafeZone AI, a smart surveillance system that uses Artificial Intelligence (AI) to automatically watch CCTV footage and listen to sounds. Our system can detect unusual activities such as:

-Accidents

-Weapons

-Fires

-Loud crashes or screams

-Large crowds

-Unexpected animals in public areas

It works in both urban (city) and rural (village) areas. When the system sees or hears something suspicious, it quickly sends alerts to the right people — like nearby security guards or the police.

SafeZone AI uses advanced AI models like YOLOv8 for video analysis and sound classification models for audio detection. It also includes a user-friendly dashboard that shows live video, detection results, and alert messages.

1.2 Problem statement

Despite the widespread use of the CCTV system in both urban and rural areas, most of them function only as passive monitoring tools. They require human operators to

constantly watch multiple video feeds, which is time consuming, error prone, and not scalable. In many cases, critical events such as accidents, criminal activities, fires, or unexpected animal intrusions go unnoticed or are identified too late.

Several AI-based surveillance systems exist, but they:

-Primarily focus on urban settings and ignore rural scenarios.

-Often specialize in detecting only one type of anomaly (e.g., just weapons or just fire).

-Do not integrate both video and audio data for a more reliable decision.

-Lack automated alert systems (such as notifying local authorities or police in real-time).

-Usually require high-end hardware not suitable for rural deployment.

-These limitations reveal a clear gap in the current surveillance landscape —- the need for a versatile, intelligent, and responsive surveillance system that works effectively in both urban and rural environments and can detect multiple types of threats through both video and sound analysis.

So, SafeZone AI aims to fill this gap by providing:

-Real-time multi-object detection using YOLOv8.

-Audio anomaly detection (e.g., screams, gunshots, crashes).

-Integration with GPS for area-specific alerts.

-A scalable dashboard interface for remote monitoring and quick response.

1.3 Motivation

Every day, accidents, crimes, fires, and other dangerous things happen in both cities and villages. Even though there are CCTV cameras in many places, they only record videos — someone has to watch them all the time. This is hard and not always possible.

Sometimes, help arrives too late because no one noticed the problem quickly. We thought, “What if a computer could watch the video and listen to sounds, and alert

people when something is wrong?”

That idea motivated us to create SafeZone AI , a smart system that can watch CCTV videos, hear unusual sounds, and quickly tell the right people when there is danger. This can help save lives and make places safer for everyone.

1.4 Objectives

-To analyze CCTV video and audio in real time using AI to detect an unusual and suspicious activities.
The system will detect unusual or suspicious activities such as accidents, presence of weapons, fires, loud crashes, screams, large crowds, or unexpected animals using advanced video (YOLOv8) and audio (CNN) detection models.

-To ensure quick response through automated alerts and GPS mapping:
Upon detecting any threat, the system will immediately send alerts via SMS, email, or alarms, along with GPS-based location details, to notify nearby authorities or emergency services, ensuring public safety in both urban and rural areas.

1.5 Scope Of Project

The SafeZone AI project is designed to make places like cities, towns, villages, schools, roads, and public areas safer using smart technology. This system can watch CCTV videos and listen to sounds to detect unusual or dangerous situations such as accidents, fire, weapons, loud noises, animals, or large crowds.

It works in both urban and rural areas, where sometimes help arrives late due to lack of monitoring. Our system can automatically detect problems and send alerts to the right people like security guards or police.

The project includes real-time object and sound detection using AI to identify unusual activities such as accidents, weapons, fires, or loud noises. It features a smart dashboard that displays live CCTV footage along with alerts, allowing security personnel to monitor incidents as they happen. An automatic alert system is integrated to send notifications through SMS, email, or sound alarms, ensuring quick response. Additionally, GPS-based mapping is used to pinpoint the exact location of detected incidents,

helping authorities take immediate action. In the future, the system can be expanded to detect a wider range of threats, such as floods, physical fights, or gunshots. It can also be integrated with drones and smart street sensors for enhanced coverage and mobility. Moreover, support for multiple languages and different regional settings can be added to make the system more adaptable and inclusive.

CHAPTER 2

RELATED THEORY

To build SafeZone AI, we use several important theories and concepts from computer science and artificial intelligence. These theories help our system "see," "hear," and "think" like a human in order to identify danger and send quick alerts. The main supporting theories are:

1. Computer Vision

Computer vision is a field of AI that allows computers to understand and interpret visual information from the world, such as images and videos. In our project, computer vision is used to analyze live CCTV footage and detect different objects and activities. For example, the system can identify if a person is holding a weapon, if there is a fire in the area, or if animals have entered a restricted zone. This helps the system watch the surroundings, just like a human security guard.

2. Object Detection (YOLOv8)

Object detection is the process of identifying and locating objects in an image or video. YOLO (You Only Look Once) is a popular and fast object detection model. We are using YOLOv8, which is the latest and most accurate version. It helps detect multiple objects in real time with high speed and precision. In our project, YOLOv8 will detect People,Cars and vehicles.weapons, fire, animals,accident, crowds,etc This model allows the system to quickly react when any threat appears on the screen.

3. Convolutional Neural Networks (CNN)

CNNs are a type of deep learning model specially designed for image and video data. They are the core of modern computer vision systems. YOLOv8 itself is built using CNNs. These networks learn how to recognize patterns, shapes, and features in an image. For example, a CNN can learn what a gun or a fire looks like by training on many example images. Once trained, the model can detect these items in live video feeds with high accuracy.

4. Audio Classification

Besides video, our system also listens to sound using microphones connected to the surveillance system. Audio classification is a process that helps the AI recognize differ-

ent types of sounds. We convert sounds into visual forms like spectrograms (which look like colorful graphs of sound) and use machine learning to identify them. The system can detect unusual or threatening sounds such as Loud crashes e.g., car crash, Human screams, Gunshots This makes the system smarter, as it doesn't rely on video alone—it can also detect danger through sound.

5. Anomaly Detection

Anomaly detection is about finding things that are different from normal behavior. In our system, anomalies include unusual movements (like someone running in a normally quiet area), large sudden crowds, or strange sounds at odd times. AI models are trained to understand what is "normal" and flag anything that seems out of the ordinary. This helps identify problems even if the exact object is not known in advance.

6. Automation and Alert Systems

This theory focuses on creating systems that can work without needing constant human control. In our project, when the AI system detects a threat, it automatically sends alerts to the appropriate people (such as nearby security guards or police officers) using SMS, email, or alarms. It also shows the location of the incident using GPS on the dashboard. This helps save time and ensures that help arrives quickly.

CHAPTER 3

LITERATURE REVIEW

1. Traditional CCTV Surveillance Systems

Traditional surveillance systems are still widely used in public spaces, commercial buildings, and even residential areas. These systems consist of CCTV cameras that record video footage continuously. However, the role of these systems is mostly passive , they store video, but do not provide intelligent analysis or automatic alerts. Human security personnel must monitor multiple screens to spot any suspicious or dangerous activities. This method is highly inefficient for several reasons:

-Humans can lose focus or become fatigued, especially during long shifts or when monitoring multiple cameras at once.

-Critical incidents may go undetected in real time, and are only noticed later during video review.

-There's no proactive prevention of crime or emergency; response is often too late.

-These limitations highlight the need to shift from traditional passive surveillance to active, intelligent monitoring systems.

2. AI-Powered Video Surveillance

To overcome human limitations, many researchers and industries have started developing AI-based surveillance systems. These systems use deep learning models, such as YOLO and SSD , for real-time object detection and classification. They can identify humans, vehicles, bags, weapons, and suspicious movement. detect intrusions, loitering, or unauthorized access, Analyze large video feeds 24/7 without getting tired. Such AI systems offer faster, more accurate, and more consistent monitoring than humans. However, they also have several weaknesses:

They are mostly implemented in urban areas with high security budgets. Many are designed to detect only a single type of threat (e.g., weapon detection, fire detection). No support for sound — they cannot respond to gunshots, arguments, screams, etc. In most cases, they don't automatically notify authorities or nearby responders. This shows that while AI video surveillance is a step forward, it's still limited and needs to

be expanded in scope and intelligence.

3. Audio-Based Threat Detection

Audio analysis using AI is an emerging field in surveillance. Certain dangerous or abnormal events are better detected through sound — such as: Gunshots,Screams or calls for help, Breaking glass,Fights or crashes. AI models can be trained on sound datasets like Google AudioSet, UrbanSound8K, and ESC-50 to recognize such sounds accurately. These models can classify environmental sounds and identify anomalies. While this technique is promising, most real-world surveillance systems do not use sound, or they treat it as a secondary source. Also, background noise, multiple overlapping sounds, or poor audio quality can reduce accuracy. The lack of integration between sound and video limits the system's ability to make accurate real-time judgments.

4. Lack of Multi-Modal Integration

In today's world, threats can come in many forms — visual, audio, or both. Unfortunately, most surveillance systems use only one input type (video or audio), which means they often miss important context. For example:

A gun may not be seen on camera, but the sound of the shot could be heard.

Someone could scream for help off-camera, but there's no video evidence.

A multi-modal system, which combines video and audio analysis in real-time, would be much more powerful and reliable. However, such systems are rare. Most are experimental or used only in high-budget smart cities. They are not widely deployed, especially in rural or resource-limited areas. In addition, very few systems are equipped to automatically send alerts (via SMS, email, or call) to emergency services or local authorities. This delay in response can make the difference between saving lives or not.

5. Need for an Advanced Unified Surveillance System Considering all the above points, it is clear that there is a strong need for a more complete, unified surveillance solution. This system should:

-Analyze both video and audio inputs simultaneously.

-Detect a wide range of events: intrusion, violence, fire, fights, screaming, weapons, etc.

-Be flexible enough to work in both urban and rural areas (e.g., detect whether a region

is crowded, open, forested, etc.).

-Include GPS-based zone awareness to identify the location of incidents.

-Automatically trigger alerts (via SMS, email, mobile notification) to police, ambulance, fire departments, or local security personnel.

-Use edge AI or lightweight models so it can run even on low-resource devices in remote areas.

-Such a system would be multi-functional, scalable, and truly AI-driven, offering real-time protection and faster emergency response. It fills the gaps left by current solutions and can make public surveillance far more effective and inclusive.

CHAPTER 4

METHODOLOGY

Our project involves training models, testing them, and improving accuracy which needs many repeated cycles.

Features like object detection, sound classification, and dashboard alerts will likely be developed in phases, not all at once.

We may add new types of events (e.g. gunshots, fire, animals) later as our dataset grows which suits the iterative style.

For the development of SafeZone AI, we will follow the Iterative Software Development Model. This method is suitable because our project involves working with AI models, which need to be trained, tested, and improved many times. First, we will collect data and build a basic version of the system using object and sound detection models. Then, we will test the system with real videos and sounds, found mistakes, and make improvements. We will repeate this process in several small steps, each time making the system better. This approach will allow us to add new features, fix errors quickly, and increase the accuracy of our detections. Since we will work with live video, sound, and alerts, the iterative method will help us develop each part one by one and test it before moving to the next step.

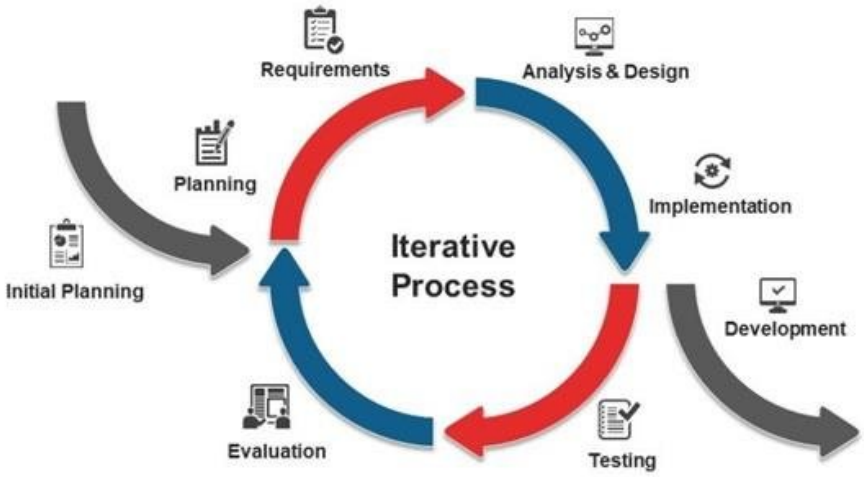

Figure 4.1: Iterative Development Model

CHAPTER 5

TOOLS AND TECHNOLOGY USED

1. Programming Language and Code Editor

i. Python: Python is the core programming language used to develop the entire system. It supports:

-AI model development (object and audio classification)

-Real-time video and audio processing

-Integration with various libraries and APIs

-Backend logic for the dashboard and alert system

- VS stdio and Jupyter Notebook

2. Machine Learning and Deep Learning Frameworks

i. YOLOv8 (Ultralytics): Used for real-time object detection from live CCTV video feeds. YOLOv8 is powerful and lightweight, making it suitable for detecting multiple objects such as: People.Fire and smoke, Weapons, Vehicles, Animals.

ii. OpenCV: For Video frame capturing and processing

iii. TensorFlow / PyTorch: These are popular deep learning frameworks used to train and deploy CNN models for tasks such as:

-Audio classification (e.g., detecting gunshots, screams, crashes)

-Anomaly detection: PyTorch offers flexibility for model development, while TensorFlow has better production deployment options.

iv. FASTAPI,Websockets

-FASTAPI For Backend intergration

-Websockets for real time communication

v. Librosa, PANNs, YAMNet: They a are library specifically designed for audio processing. They are used to:

-Load and analyze audio signals

-Convert audio files into mel-spectrograms or MFCCs using Keras.

-Normalize and extract meaningful data from sounds for training

3. User Interface

-Use for displaying live CCTV feed, alerts, GPS maps, and dashboard UI:

-HTML + CSS + JavaScript (core structure, design, interactivity)

-Bootstrap or Tailwind CSS (responsive design)

-React.js (for a dynamic, modern UI)

-Mapbox or Leaflet.js (for GPS location maps)

4. Data Labeling Tools

i. Roboflow / LabelImg: These are visual annotation tools used to prepare the dataset for object detection training. We can:

-Manually draw bounding boxes on people, weapons, fire, etc.

-Export data in YOLO format for training models

ii. Audacity / Audino: -Used for editing and labeling audio files. These tools help to:

-Trim and clean raw audio data (e.g., remove background noise)

-Annotate and label specific sounds (e.g., scream, explosion)

-Create a structured audio dataset for supervised learning

5. Database / Storage

i. SQLite / CSV Files: Used for logging and storing detected events, including Timestamp of detection, Type of event (e.g., fire, intrusion, scream), GPS coordinates or location name, Status (alert sent or not)

ii. Google Drive / Local Storage: Google Drive also serves as a cloud backup solution for portability and also used to store:

-Trained models (YOLO weights, audio classifiers)

-Video/audio datasets

-Event backup files

6. Alerting Tools

i. Twilio / SMTP Email: These services allow the system to send real-time alerts. For

example:

-SMS notifications to police or emergency responders

-Email alerts with details of detected events and possible threats

-Notification Sound / Buzzer Module: Local hardware like a speaker or buzzer can be triggered when a serious threat is detected (e.g., fire or weapon). It serves as a local alarm system, a warning for nearby people to take action.

7. Visualization and Mapping

i. Folium / OpenStreetMap: Used for visualizing GPS locations of incidents on a map. It allows:

-Plotting of detection events based on location

-Displaying nearby police stations or hospitals

ii. Matplotlib / Seaborn: Used for data visualization and performance evaluation. With these, we can plot model accuracy and loss curves, show detection frequency by time, location, or type, generate confusion matrices for classification models.

8. Hardware (If Applicable)

i. CCTV Cameras is used to capture real-time video footage for analysis and USB webcams can be used for prototyping or indoor testing.

ii. Microphone is used for capturing live sound to detect events like Gunshots, Screams, Crashes or alarms. External mics or mic arrays improve range and clarity in open areas.

iii. Raspberry Pi / Jetson Nano (Optional): These are edge computing devices that allow to run AI models locally (without cloud dependency), lower latency for real-time detection and is ideal for deployment in remote areas with limited internet access.

8. Datasets

- people, cars, animals, etc. https://cocodataset.org/
- crime surveillance videos https://webpages.uncc.edu/cchen62/dataset.html
- dense crowds and human posture https://www.crowdhuman.org/
- fire/smoke https://cchen62.org/alexnath/fire-detection-dataset
- Huge labeled dataset with 2 million audio clips https://research.google.com/audioset/
- gunshots, sirens, crashes https://urbansounddataset.weebly.com/urbansound8k.html

- Surveillance videos with unusual activity

https://www.cse.cuhk.edu.hk/leojia/projects/detectabnormal/dataset.html

- urban/rural surveillance anomalies https://zenodo.org/StevenLiuWen/ano$_pred_cvpr$2018

CHAPTER 6

SYSTEM DESIGN

6.1 Proposed System Architecture

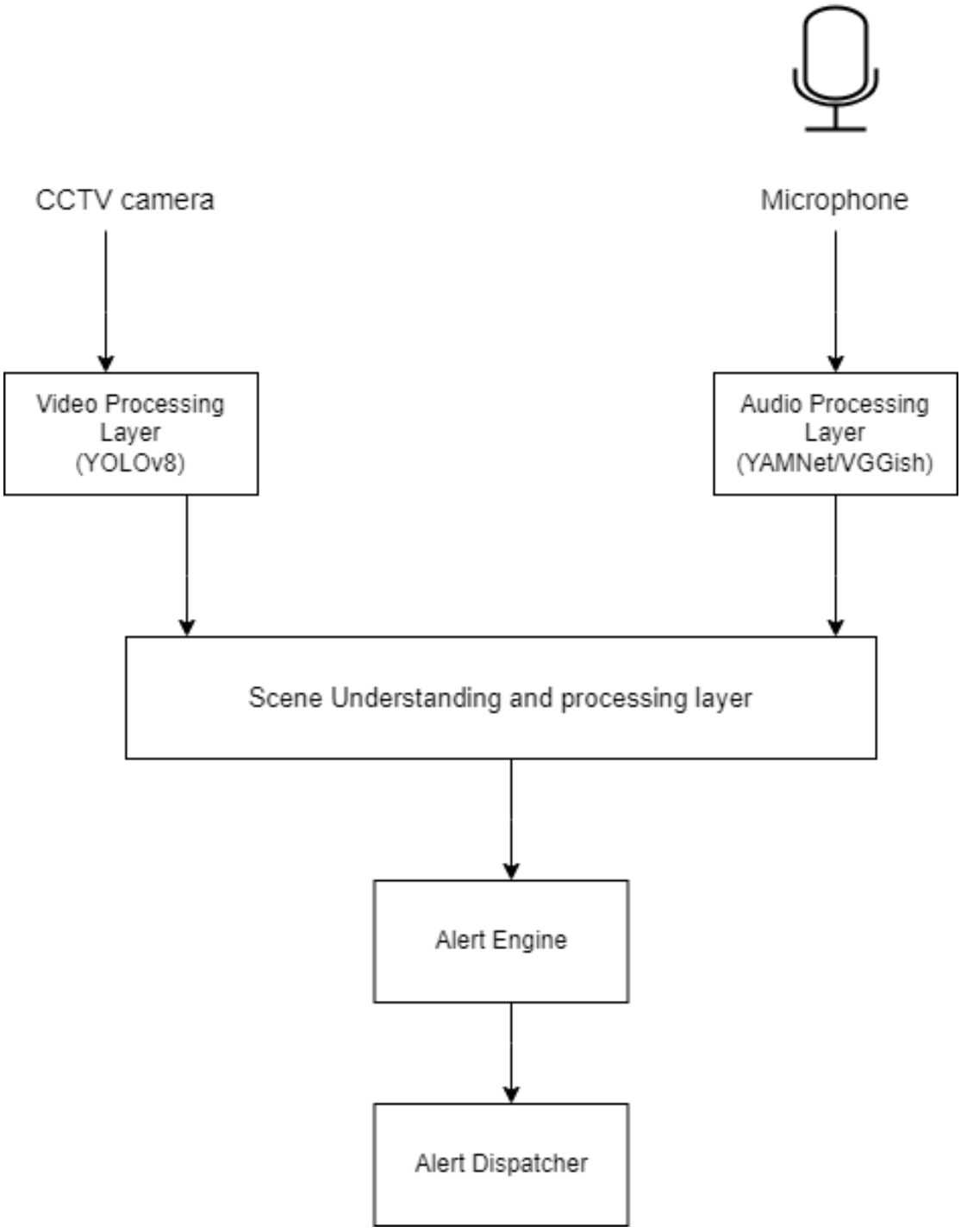

Figure 6.1: Proposed System Architecture

6.2 Use Case Diagram

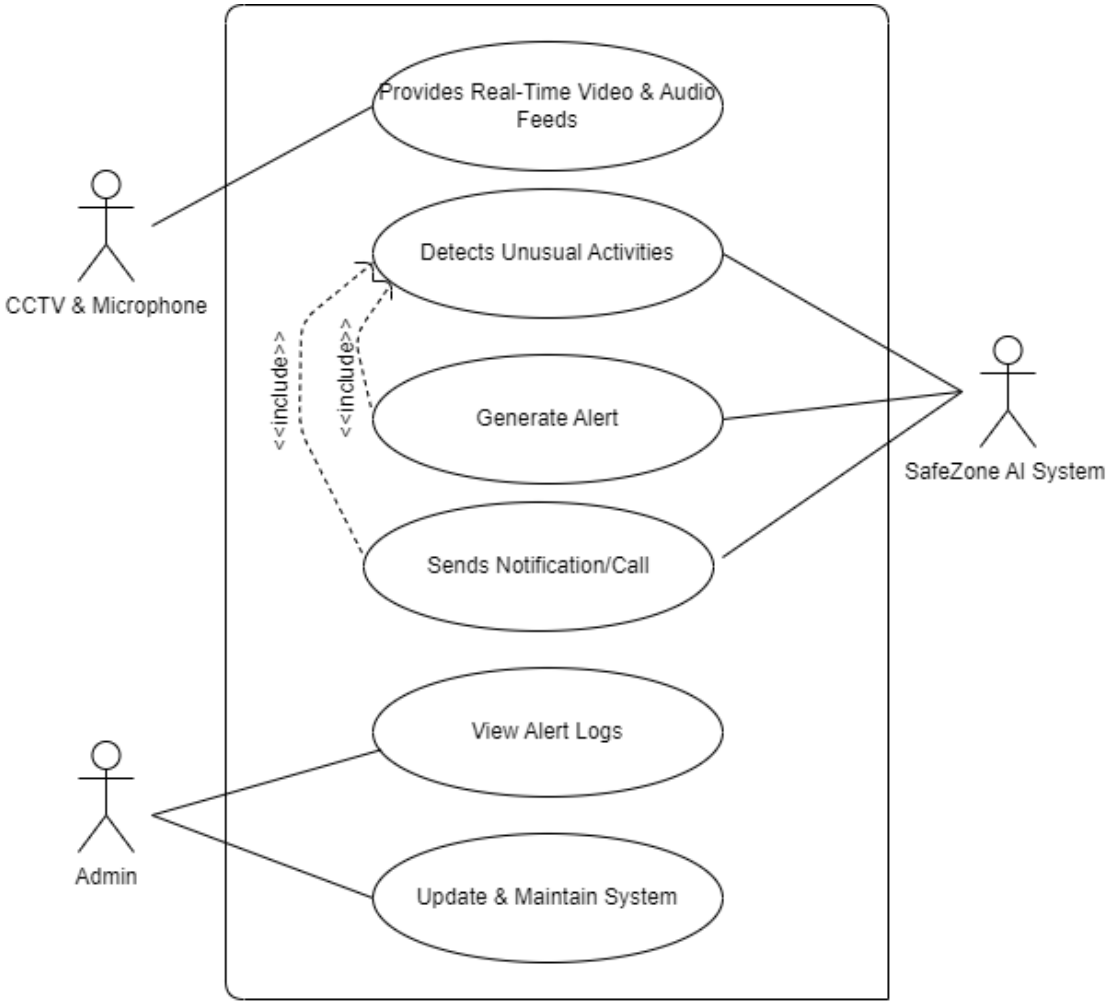

Figure 6.2: Use case Diagram

6.3 Sequence Diagram

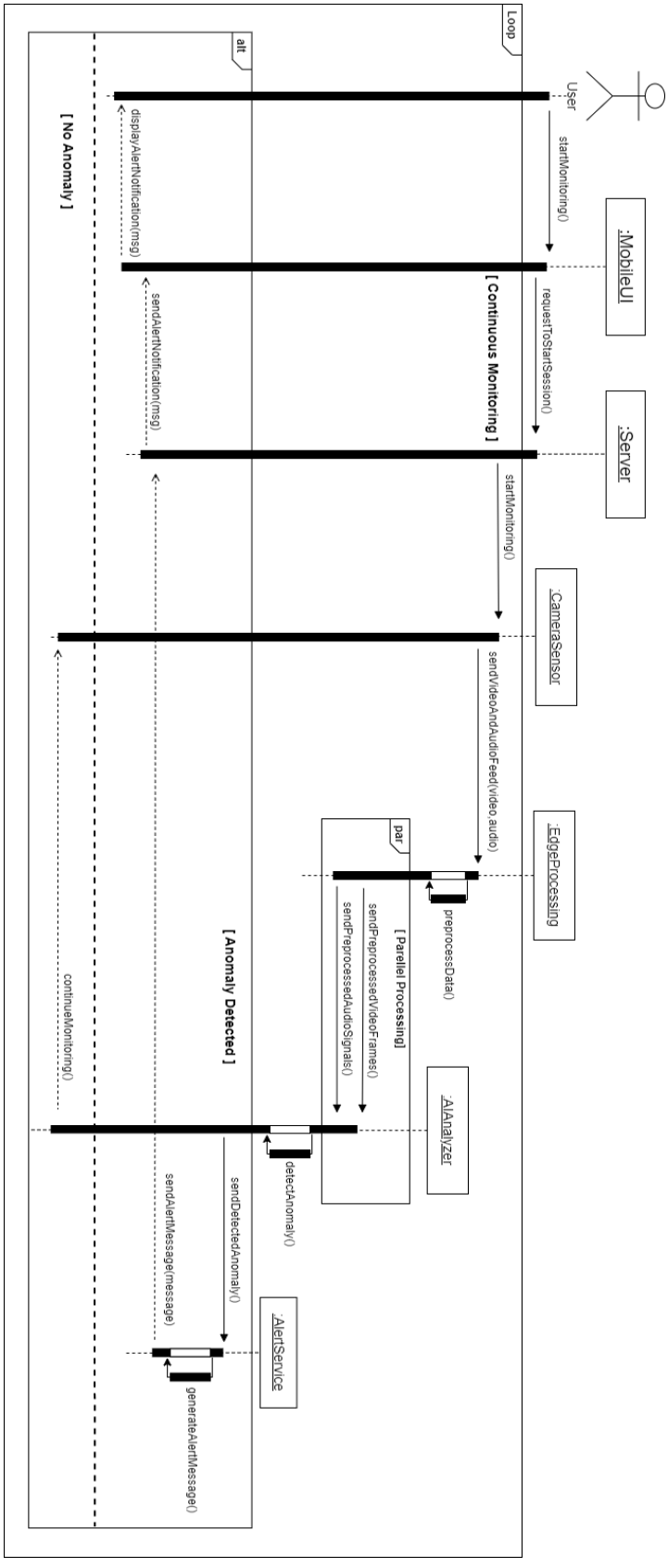

Figure 6.3: Sequence Diagram

6.4 Gantt Chart

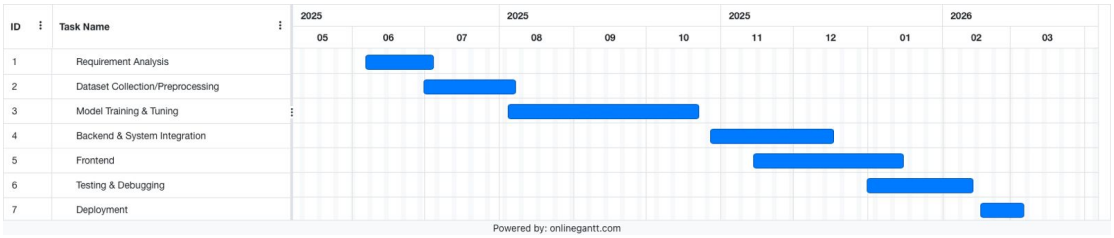

Figure 6.4: Gantt Chart

CHAPTER 7

EXPECTED RESULTS

7.1 Expected Result1

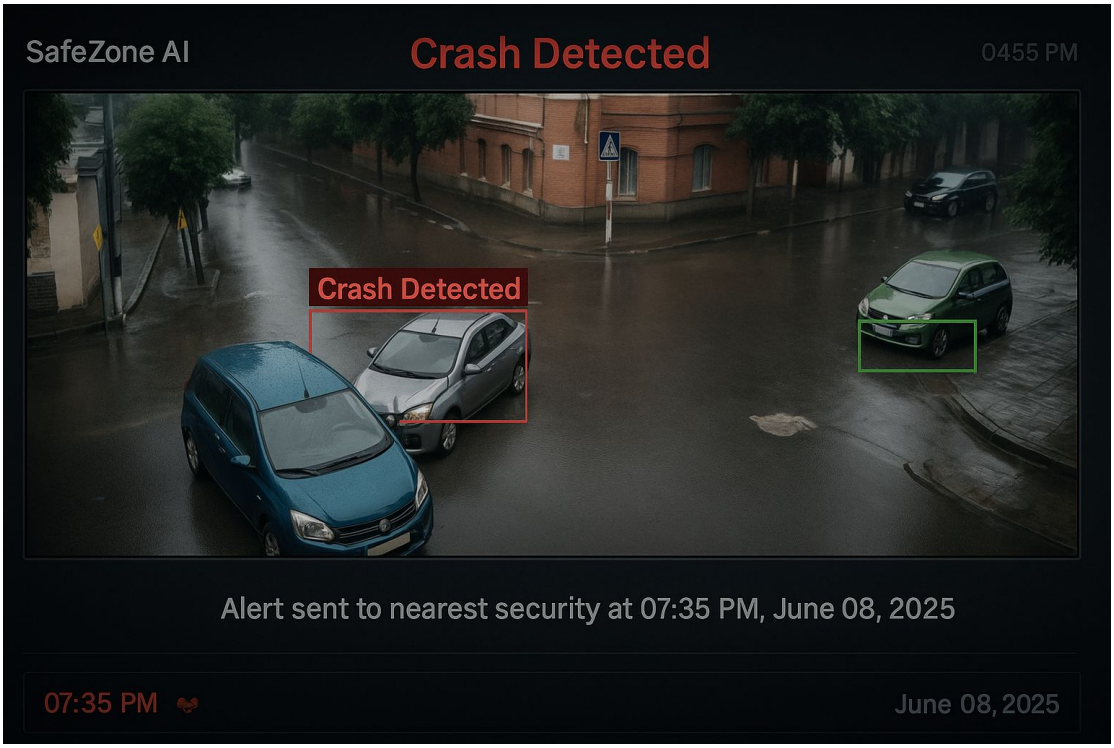

Figure 7.1: Expected Result1

7.2 Expected Result2

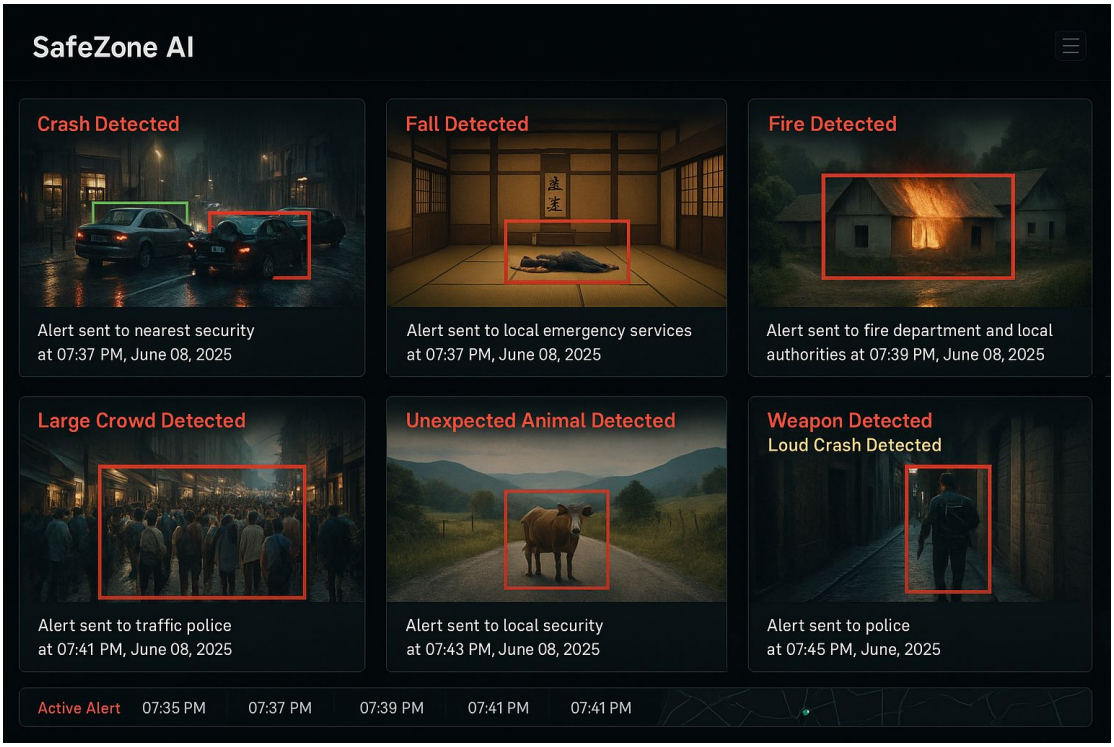

Figure 7.2: Expected Result2

REFERENCES

1. Ribeiro, M., Gonçalves, P., and Silva, F. (2018). Anomaly Detection in Video Surveillance: A Review DOI: 10.3390/app8112346 (Covers approaches to detecting anomalies in public surveillance.)

2. Sultani, W., Chen, C. and Shah, M. (2018). Real-World Anomaly Detection in Surveillance Videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

3. UCF-Crime: A Large-Scale Anomalous Video Detection Dataset Waqas Sultani, Chen Chen, Mubarak Shah. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

4. ESC-50 Dataset: Environmental Sound Classification Karol J. Piczak Proceedings of the International Conference on Machine Learning (ICML), 2015.

5. Palanisamy, K., Thirumurugan, P., and Sundararajan, M. (2019). Smart Video Surveillance System Using IoT (IEEE Conference Proceedings) (Useful for IoT + smart alert systems.)

6. Hershey, S., Chaudhuri, S., Ellis, D.P., et al. (2017). CNN Architectures for Large-Scale Audio Classification https://arxiv.org/abs/1609.09430 (For sound detection and classification using CNNs.)