Project Phase-II Report On

Smart Policing ML

Submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology

in

Computer Science and Engineering

By

JOSE F KAVALAKAT (U2003109)
KHALIL (U2003120)
LEVIN JOSEPH POOVAKULATH (U2003124)
MUHAMMED FARIZ P. A. (U2003138)

Under the guidance of
Ms. Dincy Paul

Department of Computer Science and Engineering
Rajagiri School of Engineering & Technology (Autonomous)
(Parent University: APJ Abdul Kalam Technological University)
Rajagiri Valley, Kakkanad, Kochi, 682039

April 2024

CERTIFICATE

*This is to certify that the project report entitled **"Smart Policing ML"** is a bonafide record of the work done by **Mr. JOSE F KAVALAKAT (U2003109), Mr. KHALIL (U2003120), Mr. LEVIN JOSEPH POOVAKULATH (U2003124), Mr. MUHAMMED FARIZ P. A. (U2003138)**, submitted to the Rajagiri School of Engineering & Technology (RSET) (Autonomous) in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology (B. Tech.) in Computer Science and Engineering during the academic year 2023-2024.*

Ms. Dincy Paul
Project Guide
Assistant Professor
Dept. of CSE
RSET

Dr. Tripti C.
Project Coordinator
Associate Professor
Dept. of CSE
RSET

Dr. Preetha K. G.
Head of the Department
Professor
Dept. of CSE
RSET

ACKNOWLEDGMENT

We wish to express our sincere gratitude towards **Dr. P. S. Sreejith**, Principal of RSET, and **Dr. Preetha K. G.**, Head of the Department of Computer Science and Engineering for providing us with the opportunity to undertake our project, "Smart Policing ML".

We are highly indebted to our project coordinator, **Dr. Tripti C.**, Associate Professor, Department of Computer Science and Engineering for her valuable support.

It is indeed our pleasure and a moment of satisfaction for us to express our sincere gratitude to our guide **Ms. Dincy Paul** for her patience and all the priceless advice and wisdom she has shared with us.

Last but not the least, we would like to express our sincere gratitude towards all other teachers and friends for their continuous support and constructive ideas.

JOSE F KAVALAKAT
KHALIL
LEVIN JOSEPH POOVAKULATH
MUHAMMED FARIZ P. A.

Abstract

The proposed system integrates YOLOv7 for visual object detection and Convolutional Neural Networks (CNNs) for sound detection, creating a comprehensive solution for real-time identification of critical incidents such as gun fire, choking, and screams. Deployed on specialized hardware equipped with cameras and audio sensors, the system utilizes YOLOv7 and a trained CNN to process visual and auditory data concurrently. The YOLOv7 model is adapted to accommodate the challenges of audio processing, ensuring accurate classification of auditory stimuli associated with potential threats. An alert system, triggered upon incident detection, utilizes the Twilio API to promptly notify predefined emergency contacts, conveying detailed information about the incident, including both visual and auditory elements. Integration with services like the Google Maps API provides precise location data, enhancing emergency response. The user interface facilitates real-time monitoring and system management, allowing administrators to adjust alert thresholds and oversee system performance. Security measures, encompassing both visual and auditory components, safeguard against unauthorized access and protect user privacy. Rigorous testing, covering visual and auditory detection capabilities, ensures the system's accuracy and reliability across diverse scenarios. Regular updates to the YOLOv7 model and the sound detection DNN contribute to ongoing optimization. Adherence to local laws and regulations governing surveillance and emergency alert systems underscores the system's responsible deployment, balancing technological innovation, ethical considerations, and privacy safeguards to positively contribute to public safety.

Contents

List of Abbreviations

GPU - Graphics processing unit

PC - Personal Computer

YOLO - You Only Look Once

V7 - Version 7

CNN - Convolutional Neural Networks

R-CNN- Regional Convolution Neural Network

VGG-Visual Geometry Group

CCTV-Closed Circuit Television

API - Application programming interface

COCO -Common Objects in Context

BoW - Bag of Words

IoT - Internet Of Things

ReLU-Rectified Linear Unit

FPS - frames per second

DCNN - Deep Convolutional Neural Networks

HRN - Hierarchical Recurrent Networks

FLAME- Fire Luminosity Airborne-based Machine learning Evaluation

GB-GigaByte

FFmpeg- Fast Forward Moving Picture Experts Group

JPEG - Joint Photographic Experts Group

OpenCV - Open Source Computer Vision Library

IOU - Intersection over Union

NMS - Non-Maximum Suppression

SMS - short message service

AP - Average Precision

AI - Artificial Intelligence

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The use of firearms in violent crimes has a major negative influence on public health, psychological well-being, and the economy. Each year, gun-related violence claims a large number of lives. Children who are exposed to high levels of violence in their communities or through the media are often victims of psychological trauma. Whether they are observers, offenders, or victims, children exposed to violence using weapons may suffer short- and long-term psychological consequences. Numerous studies indicate that the most common weapon used in a variety of crimes, including larceny, rape, break-ins, and robberies, is a portable gun. Additionally, there may be instances of unplanned fires that, if ignored, escalate into larger catastrophes. By spotting disruptive conduct early on and closely monitoring any suspicious activity, law enforcement authorities can take prompt action to reduce these crimes and abnormalities. Large datasets, faster GPUs, sophisticated machine learning algorithms, and improved computations have made it possible for us to efficiently set up PCs and create automated computer-based systems that can accurately distinguish and identify a variety of objects on a site. We have used a few computational vision techniques [2] and using deep learning to recognize a firearm from an image that has been taken. Only with regard to object detection and recognition in images, recent work in the fields of machine learning and deep learning, particularly with regard to deep neural networks, has demonstrated significant advancements. Object recognition and categorization serve as the foundation for any video surveillance program and are necessary for subsequent object tracking tasks.The use of advanced technologies such as YOLOv7 and Deep Neural Networks (DNNs) has allowed for the creation of a sophisticated system that can quickly identify important events like screams in real-time. This system combines visual object detection capabilities with specialized DNNs designed for analyzing sound, providing a complete solution for detecting and responding to potential threats. With cameras and audio sensors, the system can process visual and auditory data simultaneously to ensure fast and accurate detection of incidents. Modifications to the YOLOv7 model enable seamless integration of audio processing, improving the system's ability to identify sounds associated with critical events. With these advancements, the system is better equipped to handle and respond to emergency situations.[3]The system quickly notifies chosen emergency contacts using the Twilio API, giving them detailed information that includes both visual and auditory aspects of the incident. By using services like the Google Maps API, emergency response is improved with accurate location data. Administrators can easily control the system through a user-friendly interface, allowing them to monitor and adjust alert levels in real time to meet different needs. Strict security measures protect against unauthorized access and maintain user privacy, ensuring compliance with applicable regulations.

1.2 Problem Definition

There have been a number of effective automatic convolutional neural network (CNN) weapon identification systems proposed in the past. Unfortunately, these methods are sluggish to provide real-time detection, which is crucial for weapon detection systems, and have a substantial processing overhead. These models have a significant false negative rate because the poor quality and visibility of the surveillance cameras frequently allow them to miss the firearms.By decreasing false positives and negatives and delivering prompt detection responses, our goal is to increase the accuracy of anomaly detection in this instance.YOLO V7 and the DNN mode employs Single-shot object identification, which makes predictions about the location and existence of objects in an image based on a single pass through the input.

1.3 Scope and Motivation

Convolutional neural networks (CNNs) and deep learning approaches have made significant progress in image identification, classification, and segmentation over the past several years, and they are now being utilized in a variety of applications. Technology breakthroughs and the most recent, cutting-edge detection models, like YOLO, Faster

R-CNN, and VGG-16, have produced satisfactory results. The most frequent problems encountered during weapon detection are the complexity increase brought on by partial or total occlusion of gun deformation and information loss during transmission. Because these sensitive systems are connected to alarms or other devices, the rate of false-positive and false-negative results in problems for weapon detection systems as well. Due to the vital nature of weapon detection systems, real-time processing and quick reaction times are required; therefore, research must identify and apply methods that shorten the processing time of weapon detection models. The primary driver of the rise in crime rates worldwide is the increased use of handguns during violent incidents. A nation needs to maintain law and order in order to advance. A tranquil and secure atmosphere is necessary for both attracting investors and generating income from the tourism sector. Gun violence is a major contributing factor to the crime rate in many parts of the world. It primarily consists of nations where it is lawful to own a handgun. Despite this, Closed Circuit Television (CCTV) cameras are utilized for monitoring and surveillance purposes, such as during robberies, but they still need human oversight and involvement.[4] We require a system that is capable of automatically identifying these illicit actions. Real-time weapon detection remains a formidable task even with the most advanced deep learning algorithms, fast processing hardware, and sophisticated CCTV cameras available.

1.4 Objectives

- By precisely recognizing and detecting weapons in real-time, security and surveillance systems can be improved.
- Reduce false negatives and false positives in weapon detection systems.
- Combine deep learning techniques, such as CNNs, with innovative detection models like YOLO V7 and DNN model for accurate anomaly detection
- Detecting fires in their early stages to minimize damage and prevent escalation.
- Notifying building management systems, emergency services, or residents as soon as an anomaly is discovered
- To discuss and compare the different weapon detection systems used.

- To understand the different evaluation parameters used.

1.5 Challenges

Because weapons come in so many different forms, sizes, and kinds—from knives to explosives to guns—detecting weapons is a complex task. A model with strong generalization skills is needed to achieve successful identification across these several classes and their corresponding variations within each class. Furthermore, real-time processing is frequently necessary in security and surveillance applications to enable prompt threat response. But there is a major technical challenge in accomplishing high-speed detection without sacrificing accuracy, particularly when processing high-resolution video feeds. These detection systems must also function dependably in a variety of environmental circumstances, including shifting illumination, shifting weather, busy areas, and complicated backgrounds. Weapon detection systems constantly face the problem of maintaining accuracy while adjusting to these changing situations.

1.6 Assumptions

- Clear Object Visibility: The system assumes that the weapons of interest will be clearly visible within the captured images or video frames. This assumption may not hold in scenarios where weapons are partially obscured or concealed.

- Static or Slowly Changing Backgrounds: The system assumes that the background against which the weapons are detected remains relatively static or changes slowly. Rapidly changing backgrounds, such as in crowded public spaces, may pose challenges to accurate detection.

- Sufficient Lighting: The system assumes that there is sufficient lighting for the detection algorithms to effectively analyze the visual data. Low-light conditions or highly variable lighting can affect the accuracy of detection.

- Early Warning Capability: It is assumed that the system can detect fires in their early stages, allowing for timely alerts and intervention to prevent the fire from spreading and causing extensive damage.

1.7 Societal / Industrial Relevance

- Public Safety: Deployed in public spaces, transportation hubs, and crowded events to identify concealed firearms or dangerous objects, enhancing security measures.
- Law Enforcement: Assisting law enforcement agencies in detecting weapons during searches or in surveillance footage to prevent crime and ensure public safety.
- Customs Inspections: Screening baggage and cargo for illegal firearms or weapons in airports, seaports, and border crossings.
- Border Security: Monitoring and detecting illegal weapons trafficking at border checkpoints.
- School Security Systems: Integrating weapon detection systems into security protocols in educational institutions to prevent potential threats and ensure a safe environment for students and staff.
- Event Security: Utilizing weapon detection systems at corporate events, concerts, and private gatherings to maintain security measures.
- Facility Security: Deploying weapon detection in sensitive areas within corporate or government facilities to prevent unauthorized entry with weapons.
- Forest and Wildfire Monitoring: Fire detection systems can be used in forested areas to monitor for wildfires and provide early warning to authorities for rapid response and containment.
- Buildings: Fire detection systems are commonly installed in homes,schools,industries to provide early warning to the people in the event of a fire, allowing for timely evacuation and alerting emergency services.[5]
- Integration with Surveillance Systems: Integrating weapon,scream and fire detection capabilities into existing surveillance camera networks for real-time threat identification.[6]

1.8 Organization of the Report

The implementation of YOLOv7 for weapon detection is described in the paper, which covers a number of topics including the setup of hardware and software, dataset gathering and preparation, machine learning model integration, and model performance evaluation. The alert mechanism is implemented through the Twilio API. It talks on the problems and presumptions that come with weapon detection systems as well as their importance to society and industry. The paper has parts on system implementation, database design, user interface design, and implementation methodologies in addition to a Gantt chart for task scheduling. It also includes debates and outcomes, with an emphasis on the function of AI in surveillance, the capabilities of YOLOv7, dataset descriptions, methodology/algorithms, database design, user interface design, and implementation techniques. A synopsis of YOLOv7's developments and possible implications for surveillance applications round out the study..

Because early threat identification and reaction are now possible because to developments in fire and weapon detection systems, public safety and security have been greatly improved. Early fire detection, damage reduction, and escalation prevention are all made possible by the use of fire detection systems. These systems are frequently used in commercial, industrial, and residential contexts. They are also used in environmental monitoring to detect wildfires early. Conversely, weapon detection systems, which find use in law enforcement, border security, public areas, and events, tackle the problems brought about by gun-related violence and security threats. In order to increase accuracy and real-time processing capabilities, these systems make use of cutting-edge models like YOLO V7 and deep learning approaches like convolutional neural networks (CNNs). By integrating with surveillance and security systems, fire and weapon detection systems contribute to the overall safety and security of various environments, making them essential tools for crime prevention and public safety.

Chapter 2

Literature Survey

Section 2.1 to 2.5 briefs about the different literature available related to the project. Arrange them in chronological order with respect to year. The titles need not be exact copy of the paper, they can be condensed and written. More than five papers are also encouraged in this section. Increment the subsections accordingly if you add more titles.

2.1 Weapon Detection Using YOLO V3 for Smart Surveillance System

This model is applied to smart surveillance systems for weapon detection and can be implemented in high-end surveillance and security robots.The system can lock doors when a weapon is detected and share live footage with security personnel.The model is pretrained on datasets like COCO and Imagenet. The system records the location of weapon detection using geocoding and integrates this data into a database.The YOLO V3 model's performance was compared with other models like VGG-16 and BoW.The YOLO V3 model shows better performance and computational efficiency compared to YOLO V2.[7]

2.2 FireNet: A Specialized Lightweight Fire and Smoke Detection Model for Real-Time IoT Applications

Contents Proposes a lightweight model for fire and smoke detection, emphasizing real-time application in IoT systemFireNet has a total of 14 layers, including three convolution layers (each coupled with pooling and a dropout layer), and four dense layers (including a 'softmax' output layer). It uses Rectified Linear Unit (ReLU) as the activation function for all layers except the last, which uses Softmax. The first convolution layer takes a colored input image of size 64, and the number of features is doubled in each subsequent convolution layers.Dropout is used with both convolution and dense layers to improve network results and prevent overfitting. The dropout values are set at 0.5 for convolution layers and 0.2 for dense layers.FireNet is deployed on Raspberry Pi 3B, integrated with a smoke sensor and distinct fire alarms. The network is trained on a diverse dataset and tested on real-world fire and non-fire images similar to those captured by Raspberry Pi cameras. [8]

2.3 Efficient and Compact Convolutional Neural Network Architectures for Non-temporal Real-time Fire Detection

The goal is to enhance computational efficiency for real-time fire detection.Two compact CNN architectures, NasNet-A-OnFire and ShuffleNetV2-OnFire, are proposed. These architectures are designed to be less complex while maintaining high effectiveness. When implemented on low-powered devices like Nvidia Xavier-NX, the architectures achieve up to 49 fps for full-frame classification via ShuffleNetV2-OnFire, demonstrating suitability for various real-world applications.Notably, they achieve a classification speed up by a factor of 2.3× for binary classification and 1.3× for superpixel localization, with runtimes of 40 frames per second (fps) and 18 fps, respectively. [9]

2.4 Aerial Imagery Pile burn detection using Deep Learning: the FLAME dataset

Utilizes the FLAME dataset for aerial imagery analysis, applying deep learning for fire detection and segmentation. FLAME provides a dataset of aerial images of fires, captured by drones during a prescribed burn in an Arizona pine forest. For the binary classification, the Xception deep convolutional neural network (DCNN) is used. This model is trained to classify frames as either containing fire or not. For fire segmentation, the DCNN model predicts the label of each pixel in a frame, categorizing them as ”fire” or ”non-fire”. To enhance the model's efficiency, a dropout method is used to avoid overfitting, given the small number of ground truth data samples. [10]

2.5 ”DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction” by Zhang et al[1]

produces a Hierarchical Recurrent Networks (HRN) to predict crime occurrences using both spatial and temporal data.DeepCrime overcomes traditional failures by uncovering dynamic crime patterns and exploring the evolving inter-dependencies between crimes and other data in urban spaces. It captures crime dynamics using an attentive hierarchical recurrent network, which is a novel approach in the context of crime prediction.DeepCrime predicts crime occurrences of different categories in each region of a city by jointly embedding all spatial, temporal, and categorical signals into hidden representation vectors.

2.6 Summary and Gaps identified

Method	Advantages	Disadvantages	Gaps
Weapon Detection Using YOLO V3	- Real-time weapon detection - Integration with security systems - Automatic response capabilities - Geocoding for location tracking	- Dependency on training data quality - Potential false positives/negatives - Requires significant computational resources	- Limited adaptability to new weapons - Challenges in benign threat differentiation
FireNet for IoT Applications	- Lightweight and real-time suitable - Effective dropout strategy - Deployable on low-cost devices	- Limitations under different conditions - Dependency on training data quality	- Struggle with similar phenomena - Limited by IoT device sensors
Efficient and Compact CNN for Fire Detection	- High efficiency and speed - Suitable for low-powered devices - Reduced complexity	- Performance varies with environment - Limited by hardware capacity	- Adapting to varying fire characteristics - Reliance on real-time data accuracy
FLAME Dataset for Aerial Fire Detection	- Comprehensive detection using aerial imagery - Binary classification and segmentation - Enhanced model efficiency	- Quality and diversity of imagery - Processing high-resolution data	- Differentiating thermal anomalies - Limited by drone coverage
DeepCrime: HRN for Crime Prediction	- Spatial and temporal data analysis - Dynamic crime pattern capture - Adapts to evolving scenarios	- Reliance on historical data - Limited by input data quality	- Challenges in real-time processing - Ethical and privacy concerns

Table 2.1: Summary and gaps identified in current state of arts methods

Chapter 3

Requirements

3.1 Hardware and Software Requirements

3.1.1 Hardware Requirements

- Minimum 4GB RAM.
- Sufficient memory
- i3 processor system or higher

3.1.2 Software Requirements

- Windows 10 or higher operating system.
- Python3
- Yolo-v7

Chapter 4

System Architecture

Modern surveillance and emergency alert systems are complex constructions of separate parts united under a common task to protect the public. Preventing unauthorized access and preserving user privacy are key focuses. Comprehensive procedures are regularly put in place to test the system's robustness, including visual and auditory detection capabilities. Updating the YOLOv7 model and sound detection DNN on a regular basis is critically important, so that the system can rapidly adapt to changing patterns in an environment where everything changes. Strict compliance to local laws and regulations is paramount throughout the development, focusing responsibly on deployment. The balance between technological innovations, ethical concerns and privacy protections is therefore vital so that the system not only works very well but also complies with societal standards.

4.1 System Overview

The proposed comprehensive system incorporates sound detection capabilities alongside visual object detection using YOLOv7, enhancing its capacity to identify critical incidents such as gun fire, choking, and screams. Deployed on dedicated hardware with integrated microphones or audio sensors, the system utilizes Deep Neural Networks (DNNs) to process auditory data for real-time sound detection. This integration enables the system to respond not only to visual cues but also to acoustic signals associated with potential threats.

The YOLOv7 model, renowned for its effectiveness in real-time object detection, is extended to accommodate the unique challenges of audio processing. Training the DNN on labeled datasets containing sound patterns corresponding to the targeted incidents optimizes its ability to accurately identify and classify auditory stimuli.

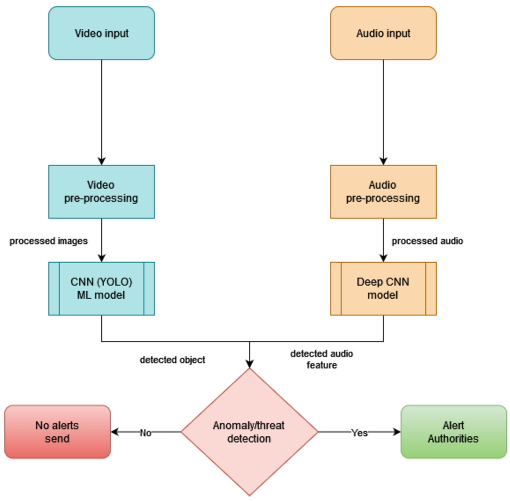


Figure 4.1: System architecture diagram

Upon detection of a critical incident, the alert system is triggered, initiating immediate notifications through the Twilio API to predefined emergency contacts. These notifications convey crucial details about the incident, including both visual and auditory information. The system's map integration, powered by services like the Google Maps API, enhances emergency response by providing accurate location data derived from both the visual and auditory sensors.

A user interface facilitates real-time monitoring of both visual and auditory feeds, incident logs, and system status. Administrators can use this interface to adjust alert thresholds, manage emergency contacts, and oversee the system's overall performance.

Security measures, including access controls and encryption, extend to the auditory components of the system to prevent unauthorized access and protect user privacy. Regular testing procedures encompass both visual and auditory detection capabilities, ensuring the system's accuracy and reliability across diverse scenarios. Periodic updates to the YOLOv7 model and the sound detection DNN contribute to ongoing optimization and adaptation to emerging patterns.

Adherence to local laws and regulations governing surveillance and emergency alert systems remains a core consideration, underscoring the system's responsible deployment. Striking a balance between technological innovation, ethical considerations, and privacy safeguards ensures that the system contributes positively to public safety while respecting societal norms.

4.2 Architectural Design

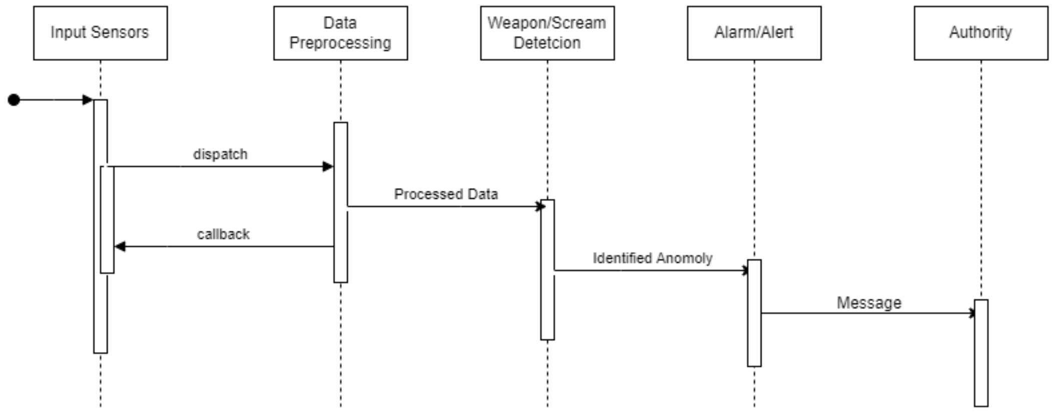


Figure 4.2: System sequence diagram

4.3 Module Division

4.3.1 Frame Conversion

- Video Processing Tool (e.g., FFmpeg): Use a video processing tool, such as FFmpeg,to extract individual frames from the video. Save each frame as an image file (e.g., JPEG). This process converts the video into a sequence of images.

- Read Video File Frame by Frame (OpenCV Library):

  Use an image processing library, such as OpenCV (cv2 module in Python), to read the video file frame by frame. OpenCV provides functions for handling video files and extracting frames.

4.3.2 Image Pre-processing

- Load Input Image:

  Start by loading the original input image.

- Resize to YOLO Input Size:

  Resize the image to the input size expected by the YOLO model (e.g., 416x416 pixels).

- Normalize Pixel Values:

  Ensure that pixel values are normalized to a specific range, often [0, 1] or [0, 255]

4.3.3 Image Classification

- Input:

  A single image is fed into the YOLO model.

- Neural Network Architecture:

  YOLO is modified to have only one bounding box prediction per grid cell (setting numboxes parameter to 1). The output layer produces a vector of class probabilities representing the likelihood of the input image belonging to different classes.

- Training:

  The model is trained on a dataset with images and their corresponding class labels. During training, the loss is computed based on the difference between predicted class probabilities and ground truth labels.

- Inference:

  During inference, a test image is passed through the trained YOLO model. The model outputs a vector of class probabilities. The class with the highest probability is assigned as the predicted class for the entire image.

4.3.4 Object detection

- In the object detection phase, the YOLO model divides the input image into a grid. Each grid cell is responsible for predicting bounding boxes and class probabilities for objects located within its spatial region.

- For each grid cell, the model predicts multiple bounding boxes along with their associated class probabilities and confidence scores.

- The bounding box predictions include four values: (x, y, w, h), representing the coordinates of the box's center, width, and height. These values are predicted relative to the dimensions of the grid cell.

- The chance that every bounding box belongs to a particular object class is shown by the class probabilities.Model confidence in the presence of an object inside the anticipated bounding box is expressed by the confidence score.

- The majority of the time, even though not all of them are significant, a single object in an image can have multiple grid box candidates for prediction.

- The user sets the threshold for its IOU selection to eliminate these grid boxes.

- Next, YOLO divides the intersection area by the union area to find each grid cell's IOU. and disregards the forecast that the grid cells' IOUs will be less than or equal to the threshold.

- Repetitive bounding box removal is accomplished by applying Non-Maximum Suppression (NMS). Among overlapping boxes, NMS chooses the box with the highest confidence score and suppresses the others.

- The final object detections are the remaining bounding boxes and the class probabilities associated with them.

- YOLO predicts bounding boxes and class probabilities for multiple objects in an image at once, completing object detection and identification in a single pass.

4.3.5 Audio sample pre-processing

- In order to handle multidimensional signals in a dataset, the suggested algorithm eliminates signals whose voices are softer than a specified threshold.

- A reference power level, a desired sample size, and audio signals are the phase's inputs. These are passed through a sequence of if else conditions.

- As outputs, signals and their labels are produced.

4.3.6 Feature extraction

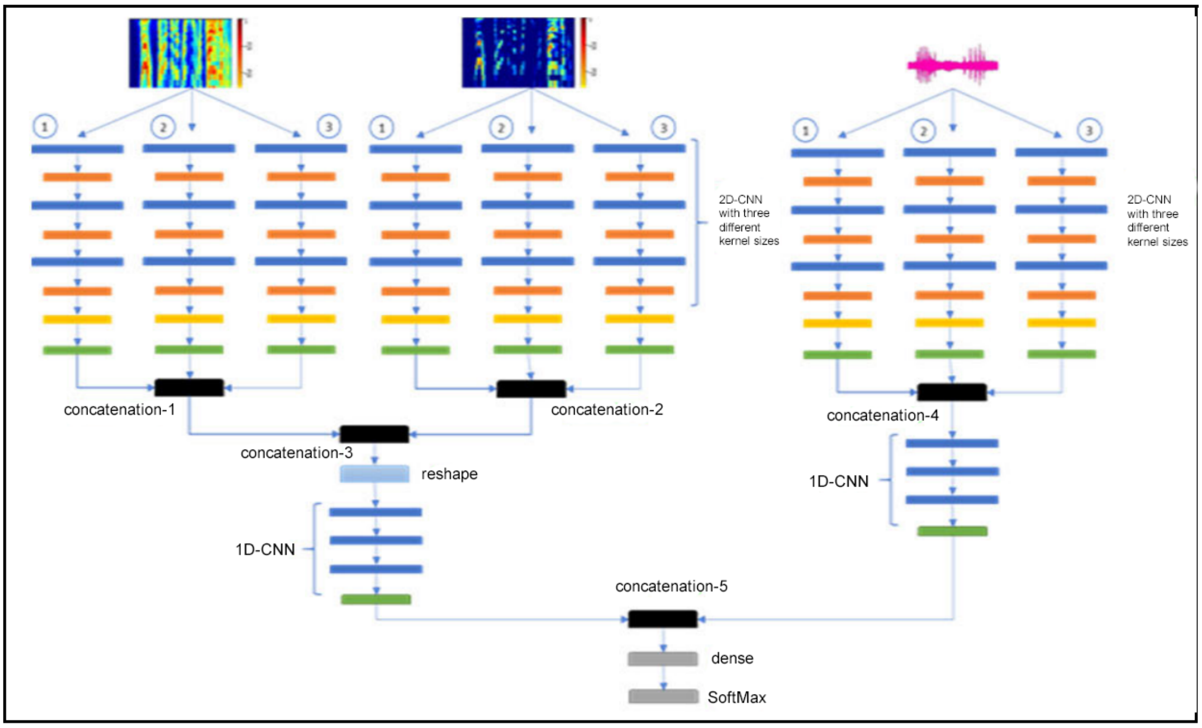
- In the study, time-domain signals, as well as Mel and log-Mel spectrograms and signal waveforms, are used as audio features.

- The deep convolution neural network uses the spectrogram features, which are taken out of the audio signals, as input data.

4.3.7 Deep learning model

- There is use of a deep multi-scale DNN.
- The model used three different kernel sizes to capture different input feature regions.
- Has two parallel branches with MEL and log-MEL as their respective inputs, as well as multiple 2D layers.
- Subsequently, every branch divides into three contiguous streams. Three global average pooling layers, three maximum pooling layers, one activation layer, and three convolution layers are present in each stream.
- For the first, second, and third streams, the convolution layers kernel sizes were 3×3, 5×5, and 7×7, with filters of order 8, 16, and 32, respectively.
- 1D feature maps are created by applying the global average pooling layer to 2D feature maps.
- Two merge layers are created when there are two parallel branches. To create the final feature maps, these two merge layers are further concatenated.
- Before being classified at the output layer, the extracted feature is passed to the dropout layer and fully connected (dense) layer.
- Signal waveforms are fed into a parallel 2D layer, which in turn feeds the waveforms to three parallel streams.
- The architecture of these three parallel streams is the same as that of the parallel streams for the 2D layer's spectrogram-taking branch.
- The feature map that results from concatenating the outputs of the parallel streams is then moved to the output layer and fully connected layer.

- Fig. shows the model with convolutional layers (blue bars) and maximum pooling layers (brown bars), respectively. The use of maximum pooling follows each convolution.
- Activation layers are represented by the yellow bar, and nonlinearity is increased by the rectified linear unit function.
- The global average pooling layers are shown by green bars. A feature map's dimensions can be globally converted using global average pooling.
- Merge layers, which are used for concatenation, are represented by the black bars.



4.3.8 Alert mechanism

- To alert the corresponding officials Twilio API is used.
- Twilio is a cloud communications platform that provides various APIs and services for sending messages, making phone calls, and handling other communication needs.
- The API supports sending text, images, videos, documents, and other types of media as part of the WhatsApp message.

- Developers can use the Twilio Python library or other Twilio SDKs in various programming languages to interact with the Twilio API and send WhatsApp messages and SMS.

- Installing the required hardware and software components and integration.(Levin): setting up the necessary hardware components, installing the required software components and frameworks like YOLOv7, along with any dependencies and libraries needed for machine learning.

- Establishing input systems,data sets for model.(Fariz,Jose): involves acquiring relevant data, cleaning and preprocessing it, and creating a structured dataset that can be used for training and testing the YOLOv7 model.

- Incorporate machine learning models,(training and testing).(Fariz,Khalil): includes training the model on the prepared datasets and testing its performance to ensure accurate and efficient crime detection.

- Integration of alert generation and alarming mechanisms.(Jose): This involves implementing features that can generate alerts or alarms when the crime detection model identifies suspicious activities or objects.

- Verifying and adjusting the model's hyper-parameters to improve the model performance.(Levin,Fariz): This involves implementing features that can generate alerts or alarms when the crime detection model identifies suspicious activities or objects.

4.4 Work Schedule - Gantt Chart

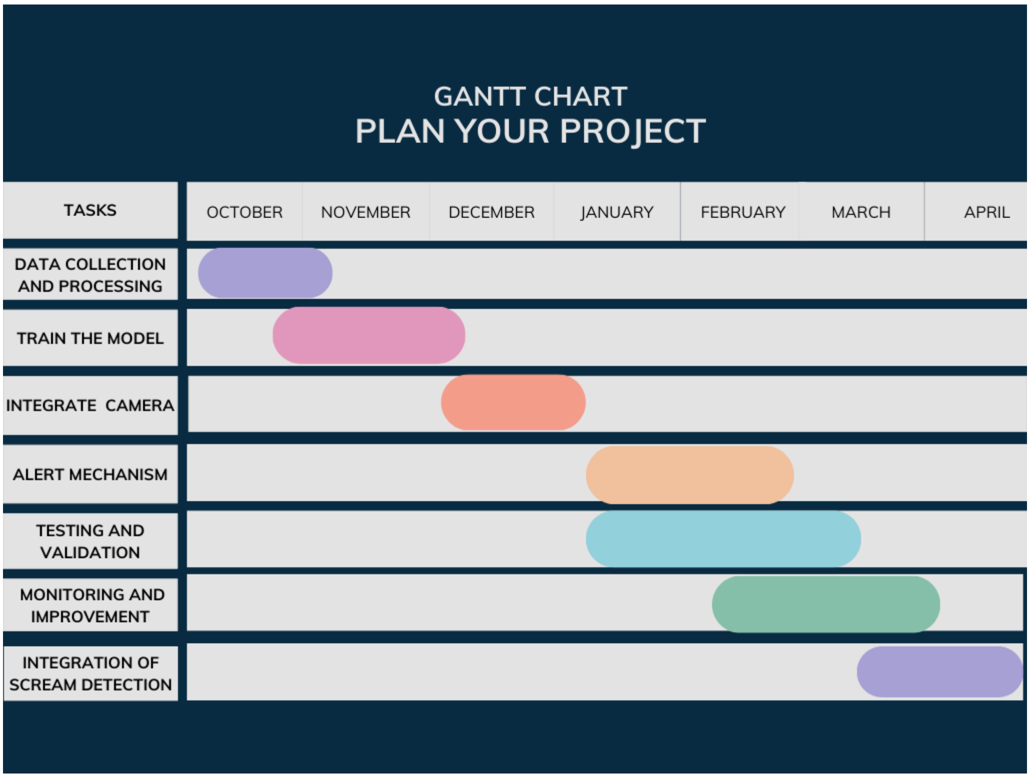


Figure 4.3: Gantt chart

YOLO model itself goes through an in-depth procedure – from training on labeled datasets to detection during actual quisition. Object detection involves advanced steps like grid cell predictions, bounding box calculations and confidence score evaluations leading to a set of detection improved upon.

The audio processing side of the functionality is no less complicated, threshold-based filtering and eventual usage of spectrograms to perform feature extraction. The deep learning model, which is the core of this approach, adopts a multi-scale DNN architecture with parallel branches , complicated convolutional layers and global average pooling for efficient feature extraction.

There is also an increased responsiveness of the system through integration with alert mechanism that uses Twilio API, a cloud communications platform. Hardware and software setup, Data creation stage of dataset modifying model training Groove integration – the work schedule is shown in a full-scale Gantt chart. The reiterative procedure for validating and refining hyperparameters allows the model to enhance its performance.

Chapter 5

System Implementation

For our project, we have identified three separate datasets to train the YOLOv7 model in order to detect fire, weapons, and human screams. The first dataset, called "Fire Dataset in YOLO Format," contains images specifically chosen to recognize instances of fire. The second dataset, "Weapon Detection YOLOv5," includes images showcasing various types of weapons like guns and knives. The third dataset, "The Human Screaming Detection Dataset," is a subset selected from a larger voice dataset focusing on human screams. The methodology and algorithms we will use will be discussed in upcoming sections, along with details on the user interface design and implementation strategies. These sections will offer a thorough overview of the project's application.

5.1 Datasets Identified

- "Fire Dataset in Yolo Format" consisting of a collection of images specifically curated for identifying various instances of fire. This dataset usually includes images of fire captured in different scenarios - indoor and outdoor and under different lighting conditions. It also contain images without fire as negative examples.Each image in the dataset should be annotated with bounding boxes around the fire regions.
- "Weapon detection YOLOv5" comprises a wide range of images featuring different types of weapons - guns and knives.The weapon images also had to be annotated with bounding boxes around each weapon in the image.
  Both the datasets were combined and formed a new dataset for training the YOLOv7 model to detect both fire and weapons involves combining these datasets, ensuring the model can simultaneously identify multiple types of objects.
- "The Human Screaming Detection Dataset" is a specialized subset curated from a broader voice dataset. The negative voice sample includes 70 voice samples,

each varying in duration from 1 to 10 seconds containg non-screaming voices. The positive voice sample includes 70 voice samples, each varying in duration from 1 to 10 seconds consisting of different scream voices.

5.2 Proposed Methodology/Algorithms

5.2.1 Video Module

- Import necessary libraries and modules. Set up configurations like weights for the YOLOv7 model, camera source, image size and confidence threshold and load the YOLOv7 model for the weapon,fire and violence classes.

- Pre-process the frame and pass the frame through the YOLOv7 model for object detection. Apply non-maximum suppression to filter out overlapping bounding boxes.

- Check if the detected object belongs to the classes of interest and if found

  Log the detection with a timestamp and location and send a WhatsApp message alert using Twilio and store the information in the database also(Firebase).

- Draw bounding boxes around detected objects on the frame.

5.2.2 Audio Module

- Import necessary libraries and modules.

- Define a function to extract features(MFCCs, chroma features, and mel-frequency cepstral coefficients) from audio files using librosa.

- Load scream and non-scream audio data,extract the features and labelthem(1 for scream, 0 for non-scream). Use LabelEncoder to convert them to one-hot encoding.

- Define a function to predict whether an audio file contains a scream or not using the trained model.

- Extract features from the audio file, preprocess them, and feed them into the model for prediction. Convert the prediction into a human-readable label using the label encoder.

- Define a function to record audio continuously for a specified duration and test the prediction using the microphone and save it to a file.

- If a scream is detected, log the detection with a timestamp and location, and send a WhatsApp alert.

- Run the audio recording and prediction process continuously to monitor for screams in real-time.

In conclusion after merging three different datasets, the YOLOv7 model has been successfully trained to recognize instances of fire, weapons, and human screams. The first dataset, known as the 'Fire Dataset in YOLO Format,' consists of images depicting fire in different situations, with bounding boxes marking the fire areas. The 'Weapon Detection YOLOv5' dataset contains images of guns and knives, similarly annotated with bounding boxes. The 'Human Screaming Detection Dataset' includes both positive and negative voice recordings of varying lengths. By combining these datasets, the YOLOv7 model is now able to detect multiple types of objects simultaneously.

Chapter 6

Results and Discussions

The increasing role of artificial intelligence in enhancing surveillance systems. Emphasize the need for real-time, accurate detection of potential threats like weapons and fire for public safety and security. YOLOv7 stands as a remarkable milestone in the evolution of object detection models, epitomizing the fusion of speed and accuracy in computer vision. YOLOv7 not only improved in accuracy and speed but also became more adaptable to different types of hardware, ranging from high-end GPUs to edge devices.The improvements in the underlying neural network architecture, better feature extraction and more efficient use of computational resources.YOLOv7 continues the tradition of the YOLO family in providing real-time object detection. Its ability to perform under varied lighting, angles, and backgrounds makes it ideal for surveillance systems where conditions are not always controlled or predictable. YOLOv7 achieves higher accuracy, reducing false positives and false negatives.

6.1 Overview

Dataset and Training: "Fire Dataset in Yolo Format" consisting of a collection of images specifically curated for identifying various instances of fire. This dataset usually includes images of fire captured in different scenarios - indoor and outdoor and under different lighting conditions. It also contain images without fire as negative examples.Each image in the dataset should be annotated with bounding boxes around the fire regions.
"Weapon detection YOLOv5"comprises a wide range of images featuring different types of weapons - guns and knives.The weapon images also had to be annotated with bounding boxes around each weapon in the image. Training the YOLOv7 model to detect both fire and weapons involves combining these datasets, ensuring the model can simultaneously identify multiple types of objects.

Model Training and Fine-Tuning: Training the YOLOv7 model on this dataset would focus on accurate and reliable fire and weapon identification, with particular attention to minimizing false positives and false negatives. Due to the differing nature of the objects (fire and weapons), careful tuning of the model's hyperparameters.

Real-Time Detection Capabilities: YOLOv7's architecture is designed for high-speed processing, allowing it to analyze video frames in real-time. The model can be optimized for performance on different types of hardware, from high-end GPUs to more constrained environments like edge devices or mobile phones.YOLOv7's advanced algorithms help reduce false alarms, a key factor in real-time applications.While speed is crucial, it must not come at the cost of detection accuracy.

6.2 Testing

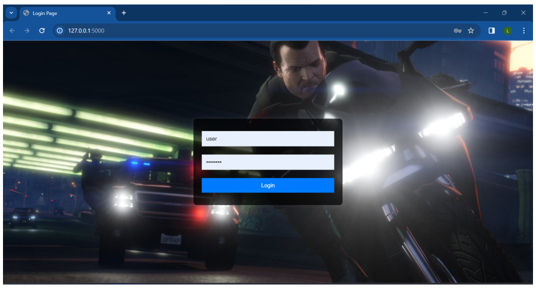


Figure 6.1: Login page

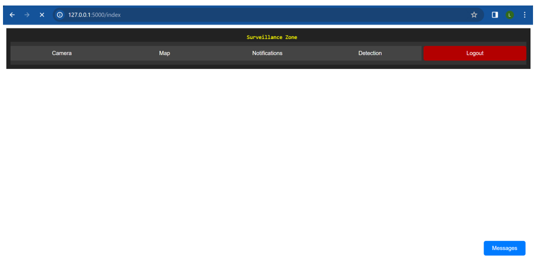


Figure 6.2: Dashboard

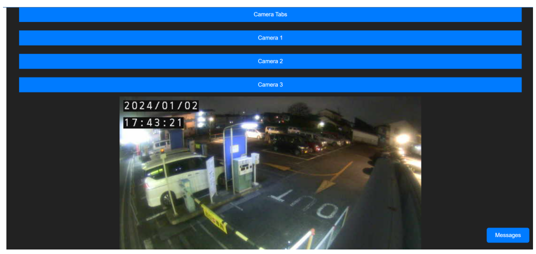


Figure 6.3: Camera page

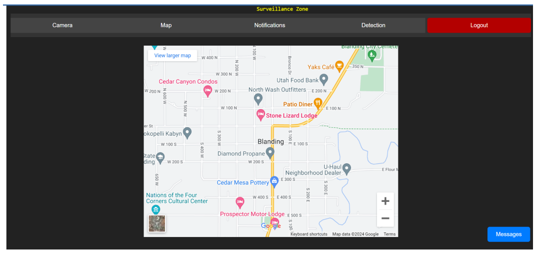


Figure 6.4: Maps

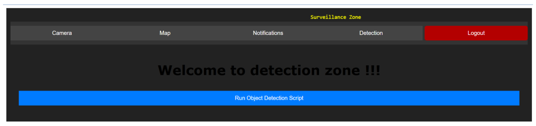


Figure 6.5: Detection page

Figure 6.6: Detection process

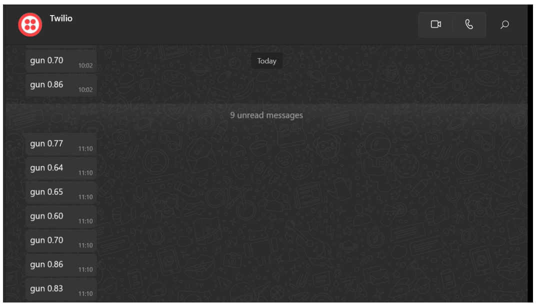


Figure 6.7: Alert generation page

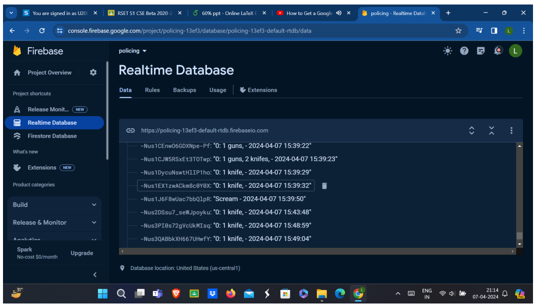


Figure 6.8: Firebase page

6.3 Quantitative Results

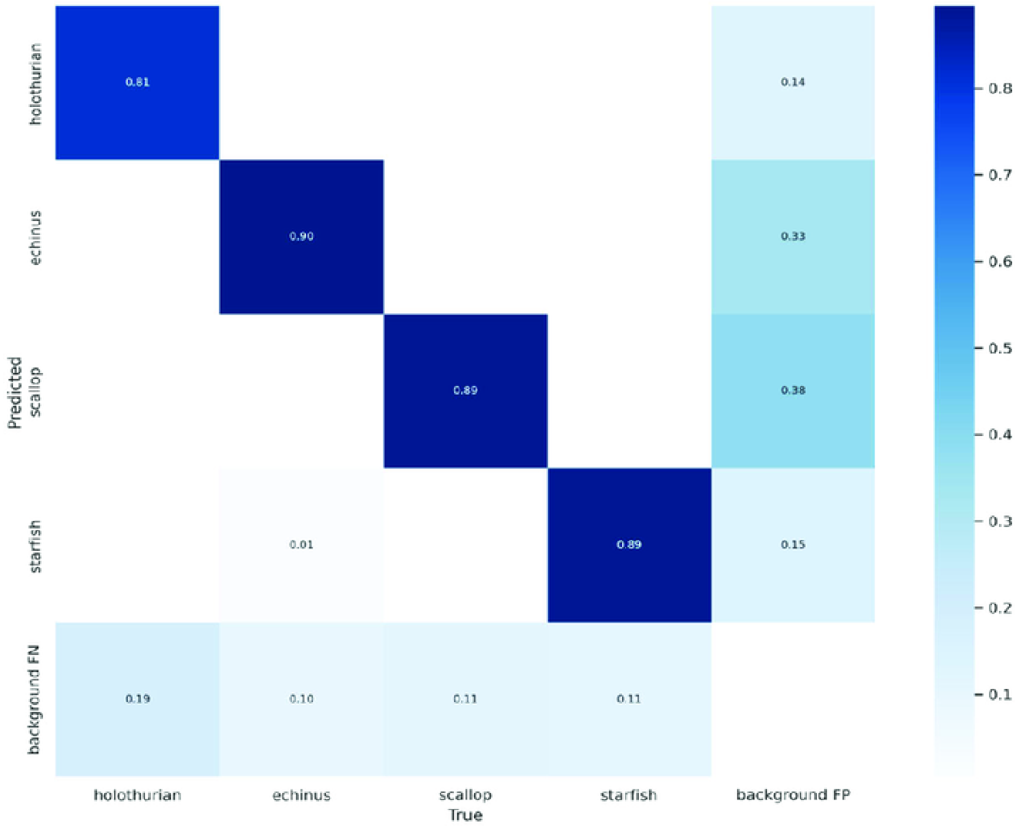


Figure 6.9: Confusion matrix for yolov7

6.4 Graphical analysis of Yolov7

The graph is plotting Average Precision (AP) against inference time measured in milliseconds on a single NVIDIA V100 GPU, batch size 1.

AP(verical axis):A higher AP suggests that the model is correctly identifying and localizing the objects with high confidence.

Inference Time (Horizontal Axis): The further to the left a model is on the graph, the faster it is at processing images.

The annotation claims that YOLOv7 is ”120percent faster,” presumably in comparison to a baseline or previous version.

Trade off: YOLOv7’s position in the graph suggests it has managed to strike an effective balance, offering both high speed and accuracy.

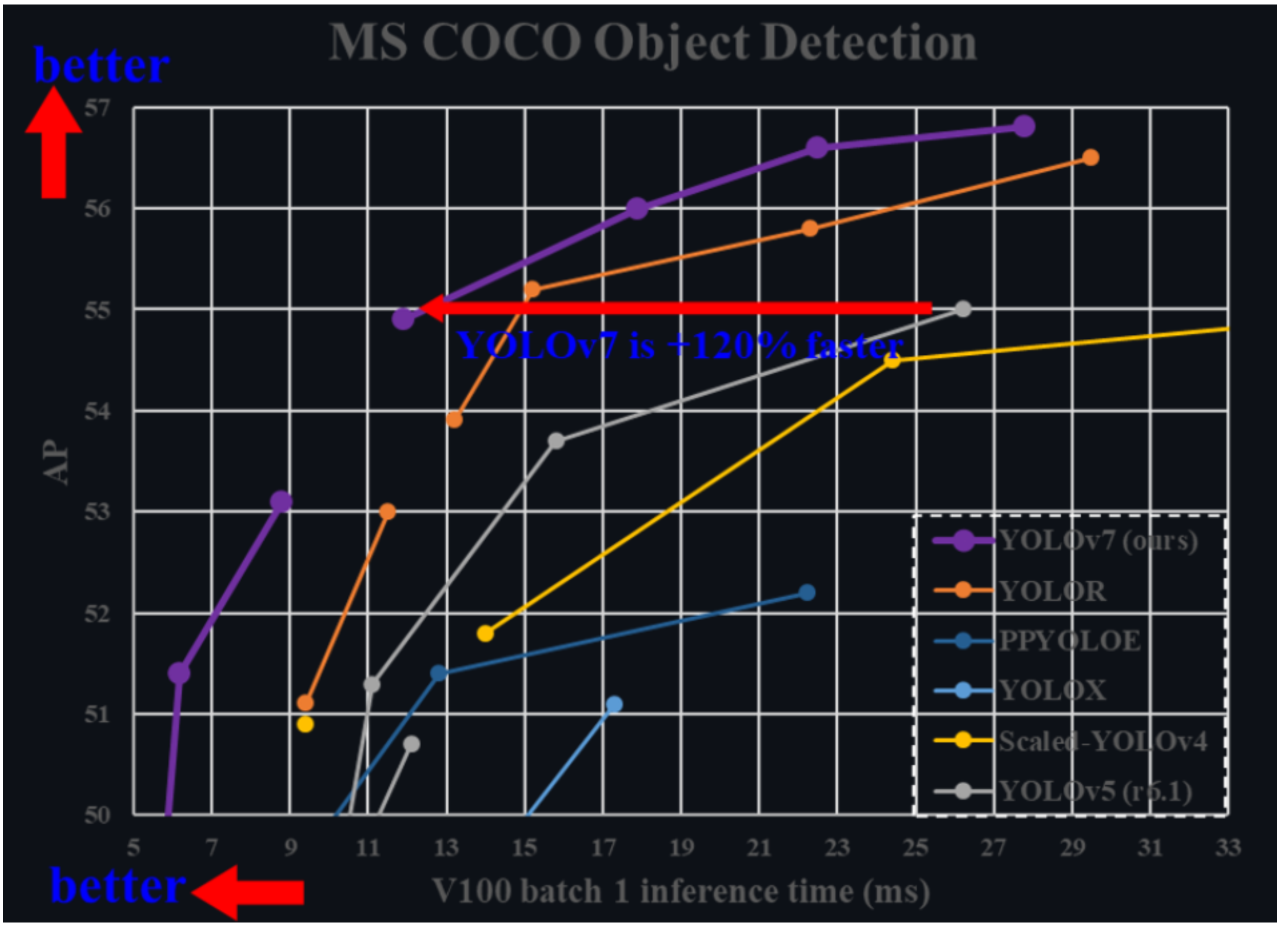


Figure 6.10: Graphical analysis with other yolo models

6.5 Scream module

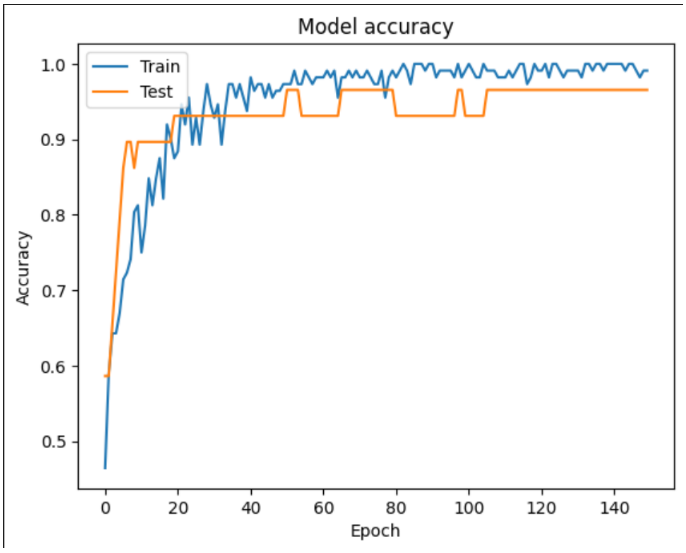


Figure 6.11: Model accuracy

X-axis represents the number of epochs and Y-axis represents the model's accuracy. Yellow line is the model's performance on training data and blue is the model's performance on the test data. A significant gap between training and test accuracy suggests overfitting.

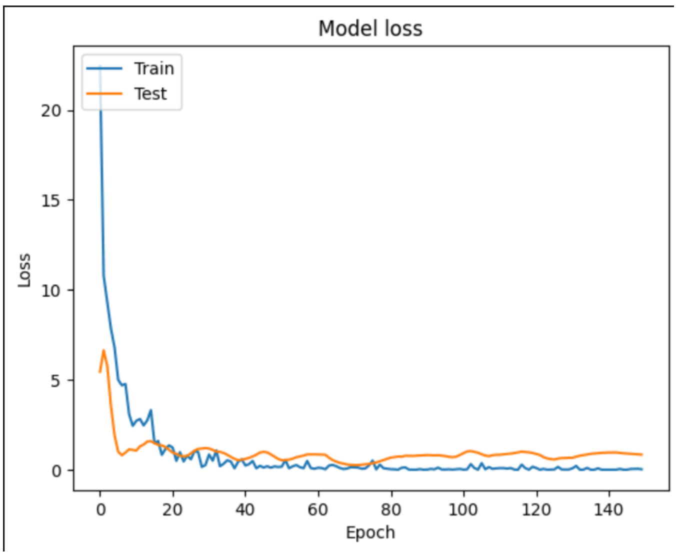


Figure 6.12: Model Loss

X-axis represents the number of epochs and Y-axis represents the training loss. The loss curve should steadily decrease over epochs, signifying that the model is progressively

learning to minimize prediction errors.
Balanced Decrease in Loss and Increase in Accuracy indicates that the model is learning effectively and generalizing well. Loss Decrease Slowing Down While Accuracy Plateaus might suggest the model has reached a point of diminishing returns. Early stopping (ending training before overfitting occurs) could be a good strategy. Loss Stagnation or Increase with No Accuracy Improvement points towards underfitting or hyperparameter tuning issues. Consider increasing model complexity (more layers/neurons) or adjusting the learning rate.

6.6 Discussion

The integration of AI in surveillance has enabled the automated detection of potential threats, such as unauthorized intrusions, unattended packages, suspicious behavior, and specifically, weapons and fires.the ability to quickly identify a threat can be the difference between a contained incident and a potential disaster. Real-time detection allows for immediate action, potentially saving lives and preventing property damage.
Enhancing Neural Network Architecture: YOLOv7 features an optimized neural network structure that boosts its object detection performance.
Advancing Feature Extraction: The model is capable of better understanding and extracting relevant features from input data, which is crucial for accurately identifying diverse objects.
Efficient Use of Computational Resources: YOLOv7 is designed to be more computationally efficient, making it deployable on various hardware platforms, from high-end GPUs to edge devices. Fire Dataset: Comprising images of fire in multiple scenarios, this dataset helps the model learn to recognize fires in different environments and under various lighting conditions.
Weapon Detection Dataset: Containing diverse images of weapons, such as guns and knives, the dataset is annotated to help the model learn to identify and localize weapons within an image.
The different natures of fire (dynamic and amorphous) and weapons (static and rigid), hyperparameters must be carefully adjusted to accommodate these differences.
Quantitative Results: Metrics such as the confusion matrix provide a clear picture of the

model's performance across different classes.
Graphical Analysis: Graphs comparing YOLOv7 to other models in terms of AP and inference time offer visual evidence of its superior performance.
Trade-Off Analysis: The model's positioning on the graph reflects its ability to deliver both high-speed and high-accuracy detection, addressing the typical trade-off between these two aspects. In audio part the small gap between training and testing accuracy is good.It's important to aim for high accuracy in AI systems, but it's also crucial to balance this with other metrics like precision and recall, depending on the specific use case.A very rapid increase could suggest overfitting.It is possible that the model is not complex enough to accurately capture the data patterns, or there may be issues with hyperparameters such as a learning rate that is too high.A decreasing loss but plateauing accuracy might require early stopping to prevent overfitting.

6.7 Summary

YOLOv7 emerges as a groundbreaking development in object detection models, seamlessly blending speed and accuracy—a quintessential demand of modern computer vision applications. Yolov7 improvements over its predecessors in both accuracy and processing velocity.YOLOv7 marks a notable advancement in the series, demonstrating substantial improvements over its predecessors in both accuracy and processing velocity.Its success includes more effective feature extraction and a more judicious use of computational resources, ensuring that the model remains both powerful and efficient. The convergence of these datasets enables YOLOv7 to adeptly identify multiple object types simultaneously, a crucial feature for multifaceted threat detection.Training the YOLOv7 model on these datasets is a focused endeavor, particularly attuned to achieving precise and reliable identification of both fire and weapons. This process involves a careful calibration of the model's hyperparameters to suit the distinct nature of the objects it's intended to detect. The model is engineered for swift processing, allowing for the immediate analysis of video frames—a pivotal feature for real-time surveillance and rapid threat response. Testing of YOLOv7 involvesthe creation of confusion matrices and graphical analyses against other models. Average Precision (AP) against inference time reveals a model that is not only faster but also more accurate. This delicate balance between speed and accuracy is the hallmark of YOLOv7's design, affirming its status as a superior choice for safety-critical surveillance applications.Moreover, by incorporating a scream detection module into surveillance systems, it enhances the system's ability to identify distress or emergency situations based on sound cues. The performance of the scream detection module is accurately analyzed through graphical representation, highlighting its effectiveness in recognizing important audio signals. In summary, the integration of YOLOv7 and additional modules in surveillance systems marks a significant step forward in utilizing AI for public safety and security. These systems combine real-time object detection with effective scream detection features, providing authorities with proactive measures to mitigate threats and improve situational awareness and emergency response. In audio detection model the accuracy of the model, as shown by its performance in both training and testing, gives us a good idea of how well it can correctly classify audio samples. Over time, as the model goes through training epochs, we see improvements in both training and testing accuracy. This shows us that the model is learning and applying patterns from the data effectively.The training loss is a way to see how well the model's predictions match the real labels during training. It's best if the training loss goes down consistently over time, showing that the model is getting better at reducing prediction errors and performing well.When a model experiences a balanced reduction in loss and an improvement in accuracy, it shows that the learning process is effective and the model is able to generalize well. However, if the decrease in loss slows down while accuracy reaches a plateau, it could mean that the model is not improving as much. In such cases, strategies like early stopping or adjusting hyperparameters need to be considered to avoid overfitting and to enhance performance.

Chapter 7

Conclusions & Future Scope

To sum up, combining YOLOv7 and DNNs for identifying incidents in real-time represents a big step forward in proactive threat detection and emergency response systems.The system for detecting human screams is a groundbreaking use of machine learning to improve public safety and emergency response. Through continued research and development, as well as partnerships across different industries, this system has the ability to save lives and minimize the effects of emergencies in communities around the globe. The integration of top-notch visual object detection with high-level sound recognition allows for a strong system to quickly identify important events such as screams, gunshots, and choking incidents. This holistic approach not only improves awareness of the situation but also gives authorities and first responders the ability to effectively address potential threats. Additionally, the alert system effortlessly connects to the Twilio API for quick sharing of incident alerts with emergency contacts, giving them important details from visual and auditory data streams. Integrating services such as the Google Maps API also improves emergency response by providing accurate location information for targeted interventions. These collaborations between advanced technologies and emergency protocols show a proactive stance on public safety.

The successful implementation of YOLOv7 involves various stages, including the installation of hardware and software components, dataset acquisition and preprocessing, machine learning model training and testing, integration of alert mechanisms, and fine-tuning of model hyperparameters. These steps ensure that the model is capable of accurately and efficiently detecting threats in real-world scenarios.[11]

The YOLOv7 model's ability to concurrently recognize fire and weapons is dependent on the datasets used for training, including the Weapon Detection YOLOv5 and the Fire Dataset in YOLO format. To guarantee accurate threat detection, the training procedure places a strong emphasis on reducing false positives and false negatives.[12]

Testing of YOLOv7 involves the creation of confusion matrices and graphical analyses, which demonstrate the model's high accuracy and speed compared to other models. The model is well-suited for safety-critical surveillance applications where real-time threat identification is crucial since it can balance speed and accuracy.[13]

Our system uses a carefully selected dataset that includes both instances of screams and background noise. We use feature extraction techniques to identify important acoustic features from audio recordings, which help the model recognize patterns that indicate screams. We have created a neural network architecture specifically for classifying audio samples as either screams or non-screams. This model has been trained on a wide range of data and fine-tuned to ensure accurate and reliable scream detection. The system can quickly detect screams in real-time, allowing for immediate responses to emergencies. It continuously monitors audio inputs to identify and flag scream instances, enabling timely interventions.When a scream is detected, the system sends alerts to nearby authorities or emergency responders using communication channels like WhatsApp or SMS. This ensures that relevant personnel are notified promptly for a swift response to the incident location.

As integrated systems advance through testing and optimization, they offer great potential to revolutionize emergency response methods. These solutions aim to find a balance between innovation and accountability by following strict privacy regulations and ethical standards. The successful implementation of these integrated systems could bring about a new era of security and resilience, enhancing community safety through timely detection and response mechanisms.

- Refining Model Training: Further improving the model by training it with diverse datasets and adjusting its settings could enhance its accuracy and adaptability in real-world situations.[14]

- Enhanced Alert Systems: Integrating the system with advanced alert mechanisms using AI-based decision-making algorithms may enhance the accuracy and reliability of alerts, reducing false alarms and improving response times.

- Multi-Sensor Integration: Combining data from various sensors, like infrared cameras and motion detectors, could provide a more comprehensive understanding of the environment, enabling better threat detection and response.[15]

- Contextual Analysis: Including contextual information such as historical data and weather conditions could help the system differentiate between normal and suspicious activities, reducing false positives.[16]

- Deployment on Edge Devices: Optimizing the model for use on edge devices like IoT devices or drones could expand its potential applications to remote or resource-constrained areas.[17]

- User Feedback and Iterative Improvement: Seeking input from users, emergency personnel, and community members can help make ongoing improvements to the system.

References

[1] C. Huang, J. Zhang, Y. Zheng, and N. V. Chawla, "Deepcrime: Attentive hierarchical recurrent networks for crime prediction," in *Proceedings of the 27th ACM international conference on information and knowledge management*, 2018, pp. 1423–1432.

[2] J. Zhao, X. Zhang, J. Yan, X. Qiu, X. Yao, Y. Tian, Y. Zhu, and W. Cao, "A wheat spike detection method in uav images based on improved yolov5," *Remote Sensing*, vol. 13, no. 16, p. 3095, 2021.

[3] I. Martinez-Alpiste, G. Golcarenarenji, Q. Wang, and J. M. Alcaraz-Calero, "Search and rescue operation using uavs: A case study," *Expert Systems with Applications*, vol. 178, p. 114937, 2021.

[4] Y. Hu, J. Zhan, G. Zhou, A. Chen, W. Cai, K. Guo, Y. Hu, and L. Li, "Fast forest fire smoke detection using mvmnet," *Knowledge-Based Systems*, vol. 241, p. 108219, 2022.

[5] L. Zhang and G. Wang, "Design and implementation of automatic fire alarm system based on wireless sensor networks," in *Proceedings. The 2009 International Symposium on Information Processing (ISIP 2009)*. Citeseer, 2009, p. 410.

[6] P. Foggia, A. Saggese, and M. Vento, "Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion," *IEEE TRANSACTIONS on circuits and systems for video technology*, vol. 25, no. 9, pp. 1545–1556, 2015.

[7] P. B. Narejo, Sanam, "Weapon detection using yolo v3 for smart surveillance system," *Mathematical Problems in Engineering*, 2021.

[8] "Firenet: A specialized lightweight fire smoke detection model for real-time iot applications," https://github.com/arpit-jadon/FireNet-LightWeight-Network-for-Fire-Detection, 2019.

[9] "Efficient and compact convolutional neural network architectures for non-temporal real-time fire detection," https://github.com/NeelBhowmik/efficient-compact-fire-detection-cnn, 2020.

[10] "Aerial imagery pile burn detection using deep learning: the flame dataset," https://github.com/AlirezaShamsoshoara/Fire-Detection-UAV-Aerial-Image-Classification-Segmentation-UnmannedAerialVehicle, 2020.

[11] S. U. Khan, I. U. Haq, S. Rho, S. W. Baik, and M. Y. Lee, "Cover the violence: A novel deep-learning-based approach towards violence-detection in movies," *Applied Sciences*, vol. 9, no. 22, p. 4963, 2019.

[12] "Bowfire: Detection of fire in still images by integrating pixel color and texture analysis," https://github.com/Lukeli0425/Fire-Detection, 2015.

[13] N. Kurek, L. Darzi, and J. Maa, "A worldwide perspective provides insights into why a us surgeon general annual report on firearm injuries is needed in america," *Current Trauma Reports*, vol. 6, 03 2020.

[14] T. Senst, V. Eiselein, A. Kuhn, and T. Sikora, "Crowd violence detection using global motion-compensated lagrangian features and scale-sensitive video-level representation," *IEEE transactions on information forensics and security*, vol. 12, no. 12, pp. 2945–2956, 2017.

[15] K. Singh, S. Rajora, D. K. Vishwakarma, G. Tripathi, S. Kumar, and G. S. Walia, "Crowd anomaly detection using aggregation of ensembles of fine-tuned convnets," *Neurocomputing*, vol. 371, pp. 188–198, 2020.

[16] S. Accattoli, P. Sernani, N. Falcionelli, D. N. Mekuria, and A. F. Dragoni, "Violence detection in videos by combining 3d convolutional neural networks and support vector machines," *Applied Artificial Intelligence*, vol. 34, no. 4, pp. 329–344, 2020.

[17] M. Cheng, K. Cai, and M. Li, "Rwf-2000: an open large scale video database for violence detection," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 4183–4190.

Appendix A: Presentation

Final Presentation

Smart Policing ML

Muhammed Fariz P A (U2003138)
Jose F Kavalakat (U2003109)
Levin Joseph Poovakulath (U2003124)
Khalil (U2003120)

GUIDE: Ms. Dincy Paul

Contents

Problem Definition

- In recent years, there has been a rising number of criminal activities in the world and it has been difficult to identify each cases through normal means efficiently.

Project Objective

- Weapon detection using YoLo which includes knife and guns sends alert message through an application.
- Fire detection using YoLo including fire and other fire related hazards and sends alert through an application.
- Violence detection using YoLo including aggressive behavior and sends alert through an application.
- Scream detection uses TensorFlow to implement and train a neural network model that learns to classify audio files as containing screams or not, based on extracted features. This model is then used for real-time prediction of scream presence in recorded audio.
- Detected anomalies have to be stored in Firebase in a real-time basis.
- An application which is given to both police and fire department which gives alert to their activities.

Novelty of idea and Scope of implementation

- Integration of audio and video data of anomaly to provide a reliable and complete model.
- Not compromising the speed and accuracy percentage in the detections.
- Use of modern software API like Twilio for alert generation and firebase for database.
- Provide better resource allocation and an effortless system to fight crimes.

Literature Survey

- "DeepCrime: A Deep Learning Framework for Crime Prediction" by Mahmood et al.(2018) presents a deep learning approach for crime prediction using CCTV footage.CNNs and RNNs are used to analyze video data and predict potential criminal activities.
- CNN extracts relevant features from the frames and are used for image and video analysis while RNN captures patterns over time and used to model the sequential nature of video frames.
- Advantages:

  A novel moethod in the field of computer vision and crime prevention.

  Potential for automation of crime prediction.
- Disadvantages :

  lack of transparency in predictions.

  high false positives and negatives.

Literature Survey

- N. Kurek, L. A. Darzi and J. Maa, "A Worldwide perspective into why a US surgeon general annual report on firearm injuries is needed in America," using deep learning combining faster R-CNN with VGG-16.
- VGG-16 uses dropout regularization in the fully-connected layer and applies ReLU activation to all the convolutional layers in addition to five max-pooling layers and one linear Softmax output layer.
- Advantages:
  - Uniform structure with only 3x3convolutions.
  - High test accuracy of 92.7percent.
- Disadvantages :
  - Deep architecture so training VGG-16 model is very slow.
  - The VGG-16 model is large.

Literature Survey

- A. Farhadi and J. Redmon,"YOLOv3: An incremental improvement," in Proc. Comput. Vis. Pattern Recognit., 2018, pp. 1–6.
- YOLOv3 is one of the fastest and most accurate deep learning-based object detectors.
- The Darknet-53 CNN backbone is used as feature extractor.
- Advantages:
  - Faster.
  - Higher Accuracy.
- Disadvantages:
  - Worse when dealing with small images.

Literature Survey

- "DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction" by Zhang et al produces a Hierarchical Recurrent Networks (HRN) to predict crime occurrences using both spatial and temporal data.
- Hierarchical Recognition Network (HRN) is a deep learning architecture that utilizes a hierarchical structure to analyze and recognize complex patterns in data.
- Advantages:
  capture complex patterns in the data, improving prediction accuracy.
  adapt to evolving crime patterns.
- Disadvantages :
  large amounts of labeled data for training.
  expensive and require specialized hardware.

Comparison

Paper	Method	Accuracy percent
Paper-1	CNN and RNN	80
Paper-2	VGG-16	90
Paper-3	YoLov3	98
Paper-4	HRN	62

Table: Comparison table

Methodology

Video Module

- Import necessary libraries and modules. Set up configurations like weights for the YOLOv7 model, camera source, image size and confidence threshold and load the YOLOv7 model for the weapon,fire and violence classes.
- Pre-process the frame and pass the frame through the YOLOv7 model for object detection. Apply non-maximum suppression to filter out overlapping bounding boxes.
- Check if the detected object belongs to the classes of interest and if found Log the detection with a timestamp and location and send a WhatsApp message alert using Twilio and store the information in the database also(Firebase).
- Draw bounding boxes around detected objects on the frame.

Methodology

Audio Module

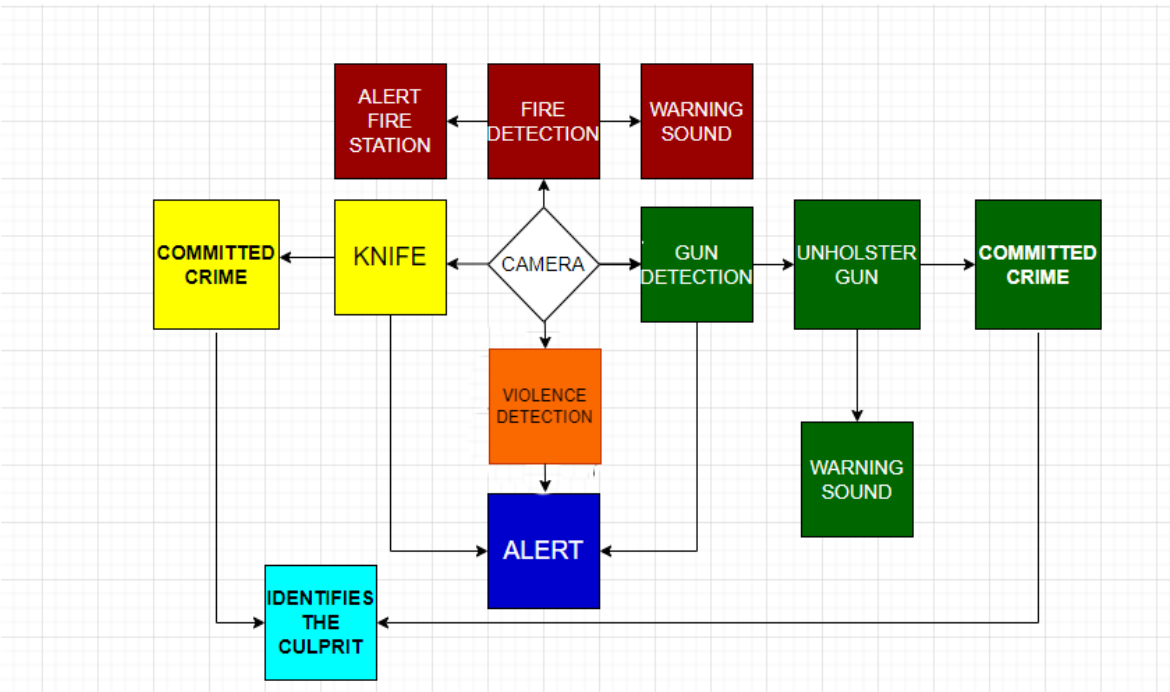
- Import necessary libraries and modules.
- Define a function to extract features(MFCCs, chroma features, and mel-frequency cepstral coefficients) from audio files using librosa.
- Load scream and non-scream audio data,extract the features and labelthem(1 for scream, 0 for non-scream). Use LabelEncoder to convert them to one-hot encoding.
- Define a function to predict whether an audio file contains a scream or not using the trained model.
- Extract features from the audio file, preprocess them, and feed them into the model for prediction. Convert the prediction into a human-readable label using the label encoder.

Methodology

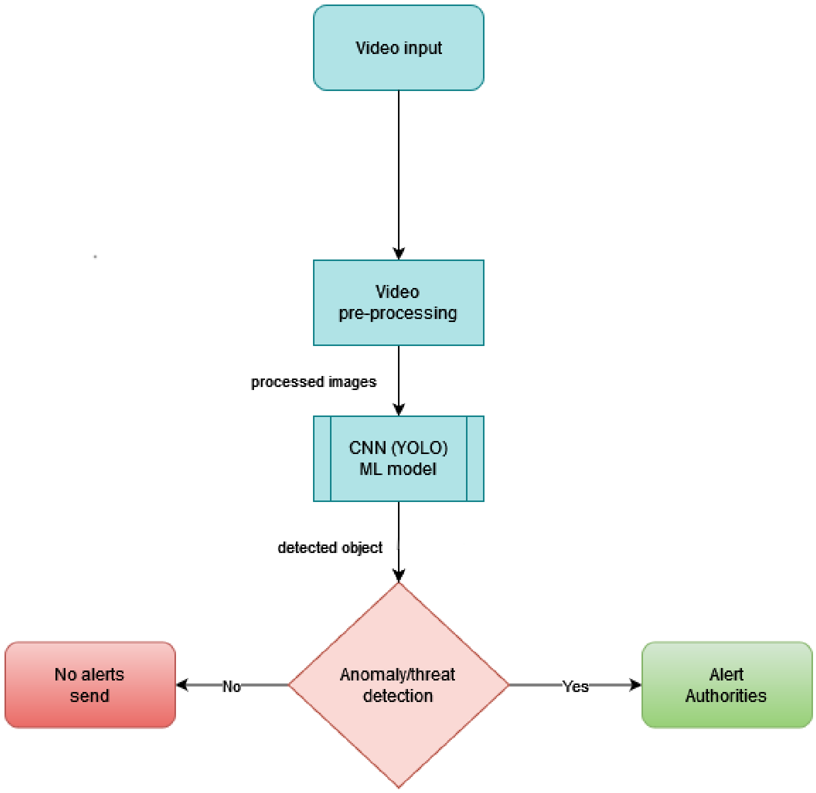
Audio Module

- Define a function to record audio continuously for a specified duration and test the prediction using the microphone and save it to a file.
- If a scream is detected, log the detection with a timestamp and location, and send a WhatsApp alert.
- Run the audio recording and prediction process continuously to monitor for screams in real-time.

System Architecture Diagram



Architecture Diagram



Results

- Integration of scream and video anomaly.
- Connecting to database for anomaly detection.
- Twilio integration for the detected scream anomaly.

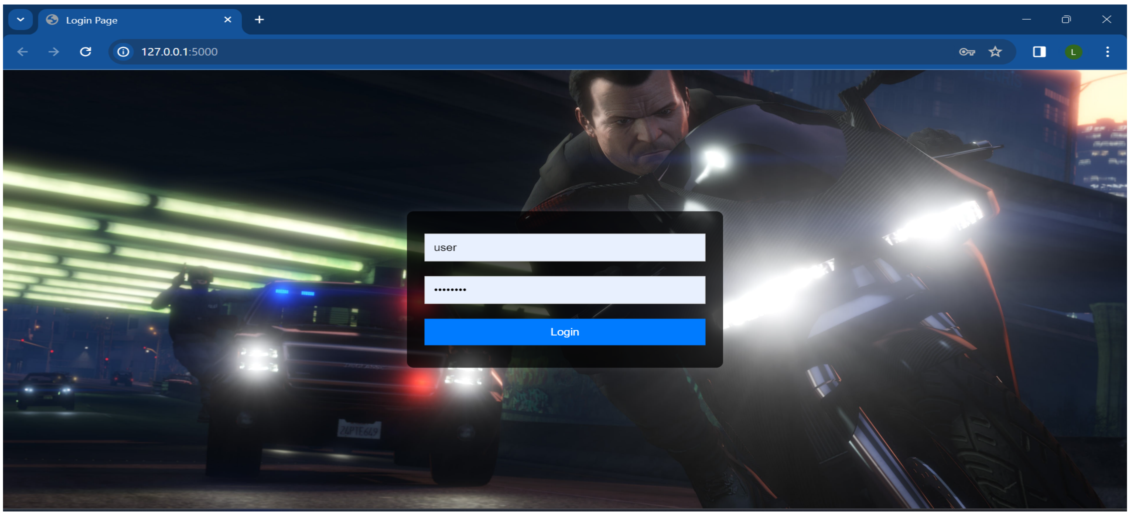
Future Scope

- Refining Model Training: training with diverse datasets and adjusting settings can boost its accuracy and adaptability in real-world scenarios.
- Enhanced Alert Systems: integrating advanced alert mechanisms using AI-based algorithms can enhance reliability, and reducing false alarms.
- Multi-Sensor Integration: combining data from various sensors enhances threat detection and response.
- Contextual Analysis: incorporating contextual information like historical data and weather conditions aids in distinguishing normal and suspicious activities.
- Deployment on Edge Devices: optimizing YOLOv7 for edge devices expands its application potential to remote areas.

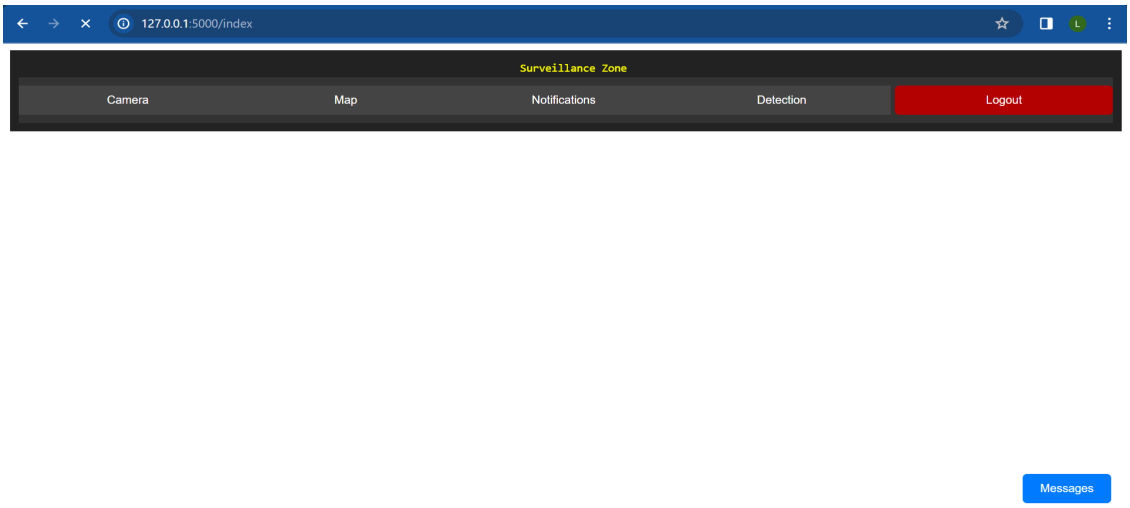
Task Distribution

- Installing the required hardware and software components and integration.(Jose,Khalil)
- Creation and integration of front-end modules.(Khalil,Levin)
- Establishing input systems,data sets for model.(Fariz,Jose)
- Incorporate machine learning models,(training and testing).(Fariz,Levin).
- Addition and integration of scream detection with the existing system.(Levin,Khalil)
- Integration of alert generation and alarming mechanisms.(Jose,Khalil)
- Database connectivity in the project.(Fariz,Levin)
- Verifying and adjusting the model's hyper-parameters to improve the model performance.(Levin,Fariz)

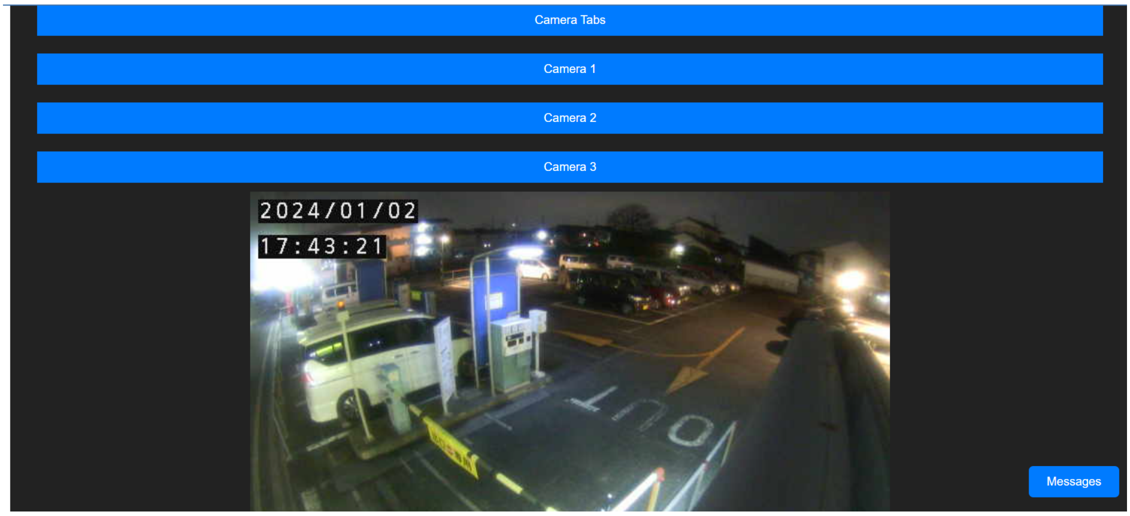
SCREENSHOTS



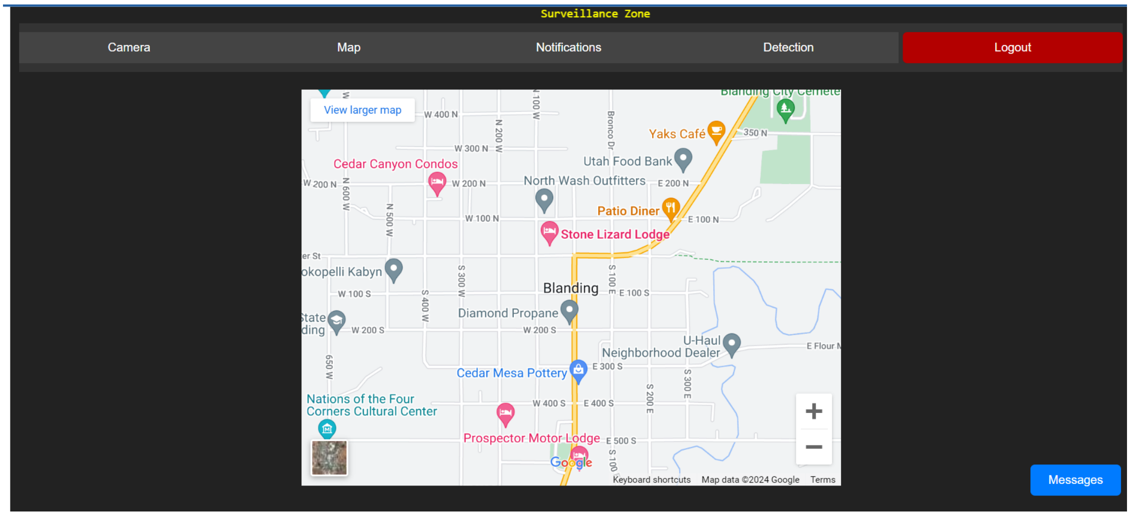
SCREENSHOTS



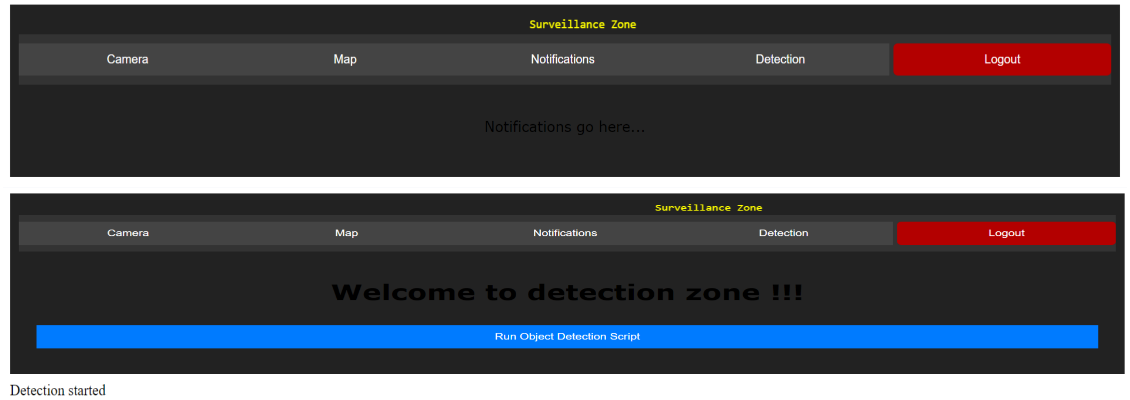
SCREENSHOTS



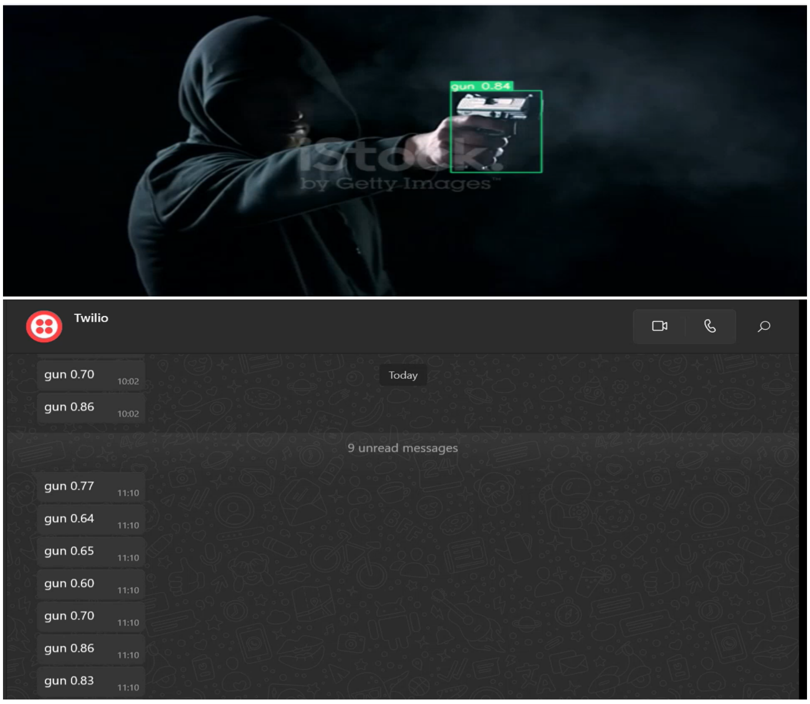
SCREENSHOTS



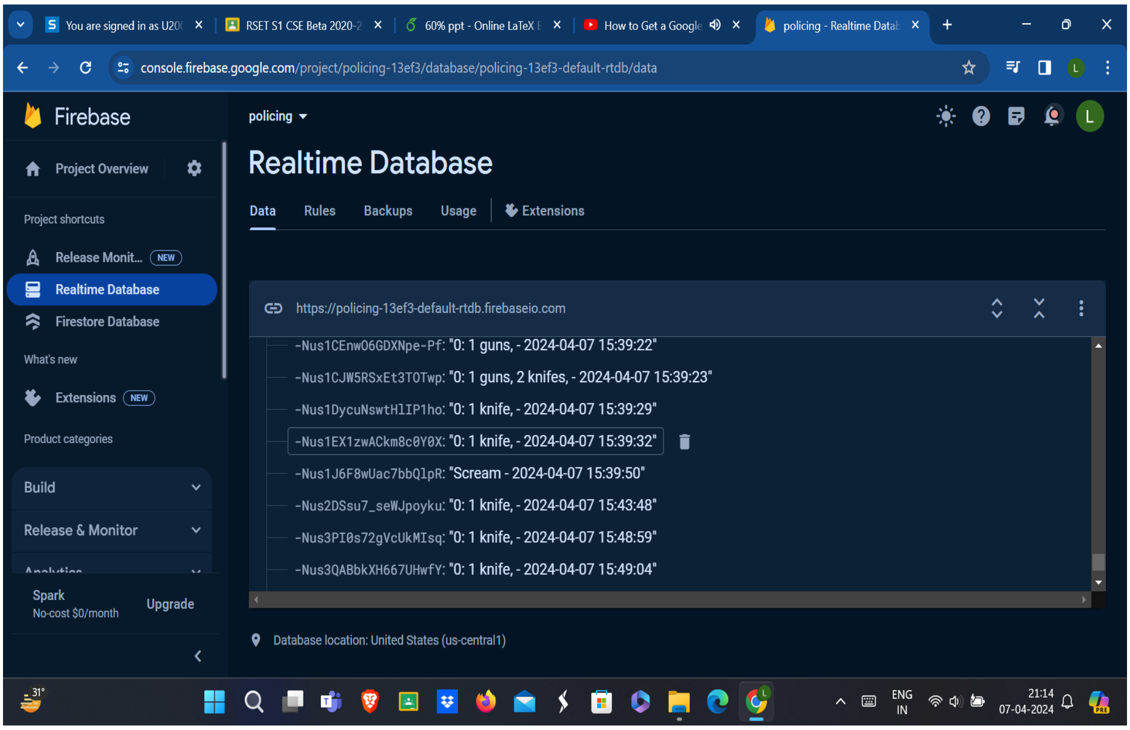
SCREENSHOTS



SCREENSHOTS

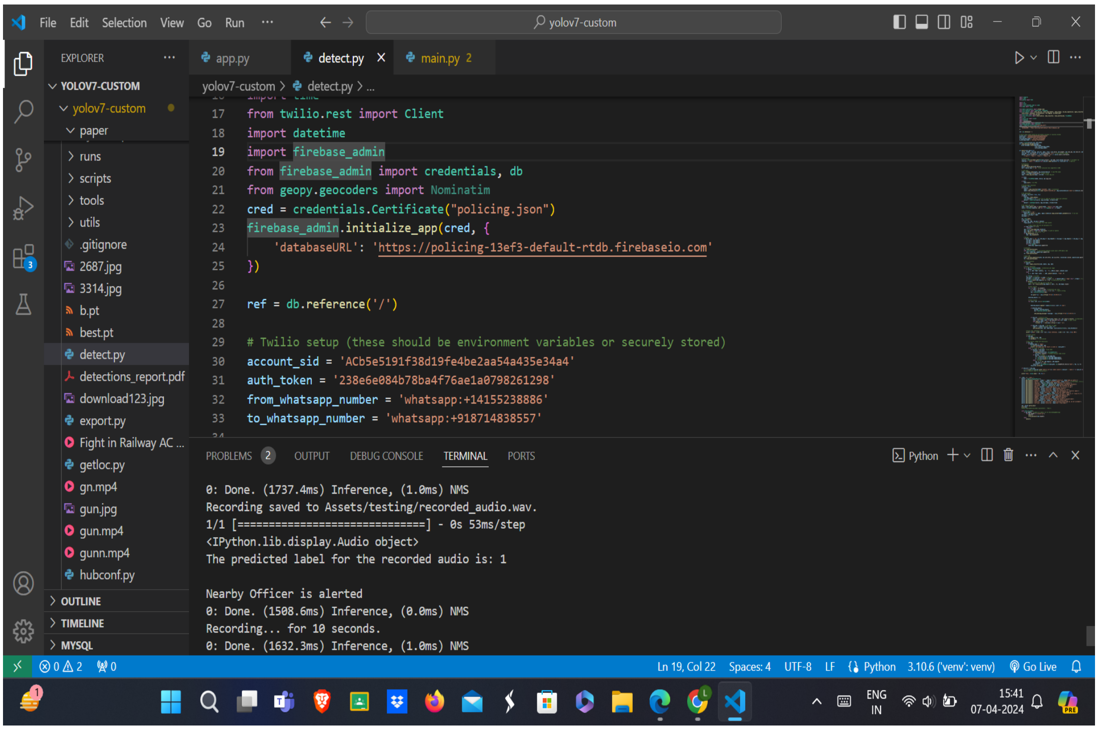


SCREENSHOTS



SCREENSHOTS

SCREENSHOTS



Conclusion

- The proposed model is expected to detect weapons,violence,fire and scream with the speed that can be used with alarm-based systems in applications of surveillance.
- It aims to the leverage the benefits of latest model YOLO-v7 combined with Tensorflow-keras model that has very effective results and speed.
- The detected anomalies are stored in a Real-time basis in Firebase.
- It also aims to alert the authorities correspondingly on detection of anomaly.

Reference I

1. "DeepCrime: A Deep Learning Framework for Crime Prediction," by Mahmood et al.(2018) in Proceedings of the IEEE International Conference on Advances in Computer Vision and Pattern Recognition, 2018, pp. 45-51.
2. N. Kurek, L. A. Darzi and J. Maa, "A Worldwide perspective provides insights into why a US surgeon general annual report on firearm injuries is needed in America," Current Trauma Reports, vol. 6, pp. 36–43, 2020.
3. A. Farhadi and J. Redmon, "YOLOv3: An incremental improvement," in Proc. Comput. Vis. Pattern Recognit., 2018, pp. 1–6.

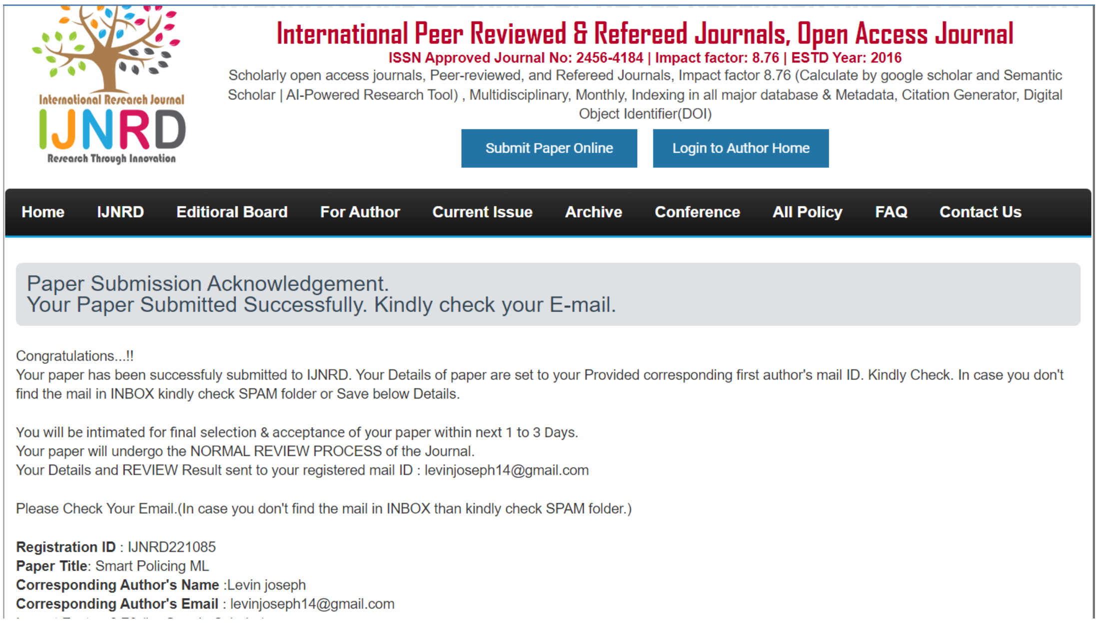
Smart Policing ML

Reference II

4. Deep convolutional neural networks for the classification and detection of human vocal exclamations of panic in subway systems. (2023). IEEE Journals and Magazine — IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/10148981
5. Y. Yang et al., "Deep Reinforcement Learning for Multi-Target Multi-Camera Tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 3, pp. 678-691, 2020.
6. K. Muhammad, "AI-Assisted Edge Vision for Violence Detection in IoT-Based Industrial Surveillance Networks," IEEE Transactions on Industrial Informatics, vol. 18, no. 8, pp. 5359-5370, Aug. 2022.

Status of paper publication

- Our paper is communicated to the journal IJNRD.



Thank You

Appendix B: Vision, Mission, Programme Outcomes and Course Outcomes

Vision, Mission, Programme Outcomes and Course Outcomes

Institute Vision

To evolve into a premier technological institution, moulding eminent professionals with creative minds, innovative ideas and sound practical skill, and to shape a future where technology works for the enrichment of mankind.

Institute Mission

To impart state-of-the-art knowledge to individuals in various technological disciplines and to inculcate in them a high degree of social consciousness and human values, thereby enabling them to face the challenges of life with courage and conviction.

Department Vision

To become a centre of excellence in Computer Science and Engineering, moulding professionals catering to the research and professional needs of national and international organizations.

Department Mission

To inspire and nurture students, with up-to-date knowledge in Computer Science and Engineering, ethics, team spirit, leadership abilities, innovation and creativity to come out with solutions meeting societal needs.

Programme Outcomes (PO)

Engineering Graduates will be able to:

1. Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. Conduct investigations of complex problems: Use research-based knowledge including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. Modern Tool Usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. Individual and Team work: Function effectively as an individual, and as a member or leader in teams, and in multidisciplinary settings.

10. Communication: Communicate effectively with the engineering community and with society at large. Be able to comprehend and write effective reports documentation. Make effective presentations, and give and receive clear instructions.

11. Project management and finance: Demonstrate knowledge and understanding of engineering and management principles and apply these to one's own work, as a member and leader in a team. Manage projects in multidisciplinary environments.

12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and lifelong learning in the broadest context of technological change.

Programme Specific Outcomes (PSO)

A graduate of the Computer Science and Engineering Program will demonstrate:

PSO1: Computer Science Specific Skills
The ability to identify, analyze and design solutions for complex engineering problems in multidisciplinary areas by understanding the core principles and concepts of computer science and thereby engage in national grand challenges.

PSO2: Programming and Software Development Skills
The ability to acquire programming efficiency by designing algorithms and applying standard practices in software project development to deliver quality software products meeting the demands of the industry.

PSO3: Professional Skills
The ability to apply the fundamentals of computer science in competitive research and to develop innovative products to meet the societal needs thereby evolving as an eminent researcher and entrepreneur.

Course Outcomes (CO)

Course Outcome 1: Model and solve real world problems by applying knowledge across domains (Cognitive knowledge level: Apply).
Course Outcome 2: Develop products, processes or technologies for sustainable and socially relevant applications (Cognitive knowledge level: Apply).
Course Outcome 3: Function effectively as an individual and as a leader in diverse teams and to comprehend and execute designated tasks (Cognitive knowledge level: Apply).
Course Outcome 4: Plan and execute tasks utilizing available resources within timelines, following ethical and professional norms (Cognitive knowledge level: Apply).
Course Outcome 5: Identify technology/research gaps and propose innovative/creative solutions (Cognitive knowledge level: Analyze).
Course Outcome 6: Organize and communicate technical and scientific findings effectively in written and oral forms (Cognitive knowledge level: Apply).

Appendix C: CO-PO-PSO Mapping

CO-PO AND CO-PSO MAPPING

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	2	2	2	1	2	2	2	1	1	1	1	2	3		
CO2	2	2	2		1	3	3	1	1		1	1		2	
CO3									3	2	2	1			3
CO4					2			3	2	2	3	2			3
CO5	2	3	3	1	2							1	3		
CO6					2			2	2	3	1	1			3

3/2/1: high/medium/low

JUSTIFICATIONS FOR CO-PO MAPPING & CO-PSO MAPPING

MAPPING	LOW/MEDIUM/ HIGH	JUSTIFICATION
100003/ CS722U.1-PO1	M	Knowledge in the area of technology for project development using various tools results in better modeling.
100003/ CS722U.1-PO2	M	Knowledge acquired in the selected area of project development can be used to identify, formulate, review

		research literature, and analyze complex engineering problems reaching substantiated conclusions.
100003/ CS722U.1-PO3	M	Can use the acquired knowledge in designing solutions to complex problems.
100003/ CS722U.1-PO4	M	Can use the acquired knowledge in designing solutions to complex problems.
100003/ CS722U.1-PO5	H	Students are able to interpret, improve and redefine technical aspects for design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
100003/ CS722U.1-PO6	M	Students are able to interpret, improve and redefine technical aspects by applying contextual knowledge to assess societal, health and consequential responsibilities relevant to professional engineering practices.
100003/ CS722U.1-PO7	M	Project development based on societal and environmental context solution identification is the need for sustainable development.
100003/ CS722U.1-PO8	L	Project development should be based on professional ethics and responsibilities.
100003/ CS722U.1-PO9	L	Project development using a systematic approach based on well defined principles will result in teamwork.

100003/ CS722U.1-PO10	M	Project brings technological changes in society.
100003/ CS722U.1-PO11	H	Acquiring knowledge for project development gathers skills in design, analysis, development and implementation of algorithms.
100003/ CS722U.1-PO12	H	Knowledge for project development contributes engineering skills in computing & information gatherings.
100003/ CS722U.2-PO1	H	Knowledge acquired for project development will also include systematic planning, developing, testing and implementation in computer science solutions in various domains.
100003/ CS722U.2-PO2	H	Project design and development using a systematic approach brings knowledge in mathematics and engineering fundamentals.
100003/ CS722U.2-PO3	H	Identifying, formulating and analyzing the project results in a systematic approach.
100003/ CS722U.2-PO5	H	Systematic approach is the tip for solving complex problems in various domains.
100003/ CS722U.2-PO6	H	Systematic approach in the technical and design aspects provide valid conclusions.

100003/ CS722U.2-PO7	H	Systematic approach in the technical and design aspects demonstrate the knowledge of sustainable development.
100003/ CS722U.2-PO8	M	Identification and justification of technical aspects of project development demonstrates the need for sustainable development.
100003/ CS722U.2-PO9	H	Apply professional ethics and responsibilities in engineering practice of development.
100003/ CS722U.2-PO11	H	Systematic approach also includes effective reporting and documentation which gives clear instructions.
100003/ CS722U.2-PO12	M	Project development using a systematic approach based on well defined principles will result in better teamwork.
100003/ CS722U.3-PO9	H	Project development as a team brings the ability to engage in independent and lifelong learning.
100003/ CS722U.3-PO10	H	Identification, formulation and justification in technical aspects will be based on acquiring skills in design and development of algorithms.
100003/ CS722U.3-PO11	H	Identification, formulation and justification in technical aspects provides the betterment of life in various domains.
100003/ CS722U.3-PO12	H	Students are able to interpret, improve and redefine technical aspects with mathematics, science and

		engineering fundamentals for the solutions of complex problems.
100003/ CS722U.4-PO5	H	Students are able to interpret, improve and redefine technical aspects with identification formulation and analysis of complex problems.
100003/ CS722U.4-PO8	H	Students are able to interpret, improve and redefine technical aspects to meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
100003/ CS722U.4-PO9	H	Students are able to interpret, improve and redefine technical aspects for design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
100003/ CS722U.4-PO10	H	Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools for better products.
100003/ CS722U.4-PO11	M	Students are able to interpret, improve and redefine technical aspects by applying contextual knowledge to assess societal, health and consequential responsibilities relevant to professional engineering practices.
100003/ CS722U.4-PO12	H	Students are able to interpret, improve and redefine technical aspects for demonstrating the knowledge of, and need for sustainable development.

100003/ CS722U.5-PO1	H	Students are able to interpret, improve and redefine technical aspects, apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
100003/ CS722U.5-PO2	M	Students are able to interpret, improve and redefine technical aspects, communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
100003/ CS722U.5-PO3	H	Students are able to interpret, improve and redefine technical aspects to demonstrate knowledge and understanding of the engineering and management principle in multidisciplinary environments.
100003/ CS722U.5-PO4	H	Students are able to interpret, improve and redefine technical aspects, recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.
100003/ CS722U.5-PO5	M	Students are able to interpret, improve and redefine technical aspects in acquiring skills to design, analyze and develop algorithms and implement those using high-level programming languages.
100003/ CS722U.5-PO12	M	Students are able to interpret, improve and redefine technical aspects and contribute their engineering skills in

		computing and information engineering domains like network design and administration, database design and knowledge engineering.
100003/ CS722U.6-PO5	M	Students are able to interpret, improve and redefine technical aspects and develop strong skills in systematic planning, developing, testing, implementing and providing IT solutions for different domains which helps in the betterment of life.
100003/ CS722U.6-PO8	H	Students will be able to associate with a team as an effective team player for the development of technical projects by applying the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
100003/ CS722U.6-PO9	H	Students will be able to associate with a team as an effective team player to Identify, formulate, review research literature, and analyze complex engineering problems
100003/ CS722U.6-PO10	M	Students will be able to associate with a team as an effective team player for designing solutions to complex engineering problems and design system components.
100003/ CS722U.6-PO11	M	Students will be able to associate with a team as an effective team player, use research-based knowledge and research methods including design of experiments, analysis and interpretation of data.

100003/ CS722U.6-PO12	H	Students will be able to associate with a team as an effective team player, applying ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
100003/ CS722U.1-PSO1	H	Students are able to develop Computer Science Specific Skills by modeling and solving problems.
100003/ CS722U.2-PSO2	M	Developing products, processes or technologies for sustainable and socially relevant applications can promote Programming and Software Development Skills.
100003/ CS722U.3-PSO3	H	Working in a team can result in the effective development of Professional Skills.
100003/ CS722U.4-PSO3	H	Planning and scheduling can result in the effective development of Professional Skills.
100003/ CS722U.5-PSO1	H	Students are able to develop Computer Science Specific Skills by creating innovative solutions to problems.
100003/ CS722U.6-PSO3	H	Organizing and communicating technical and scientific findings can help in the effective development of Professional Skills.