A Project Report on

Crime Detection Using Enhanced AI Techniques

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

BACHELOR OF ENGINEERING

In

Computer Science & Engineering Artificial Intelligence

& Machine Learning

By

Devesh Sali: 21106016

Harshal Deshmukh: 22206008

Jeet Manjrekar: 21106061

Sakshi Rajeshirke : 22206002

Under the Guidance of

Prof. Sayali Badhan

Department of Computer Science & Engineering

(Artificial Intelligence & Machine Learning)

A.P. SHAH INSTITUTE OF TECHNOLOGY, THANE

UNIVERSITY OF MUMBAI

2024-2025

A. P. SHAH INSTITUTE OF TECHNOLOGY

CERTIFICATE

This is to certify that the project entitled "Crop Recommendation System Using Machine Learning and IOT" is a bonafide work of **Devesh Sali(21106016), Harshal Deshmukh (22206008), Jeet Manjrekar(21106061), Sakshi Rajeshirke(22206002)** submitted to the University of Mumbai in partial fulfilment of the requirement for the award of the degree of **Bachelor of Engineering** in **Computer Science & Engineering (Artificial Intelligence & Machine Learning**).

Prof. Sayali Badhan
Guide

Prof. Sayali Badhan
Project Co-ordinator

Dr. Jaya Gupta
Head of Department

Dr. Uttam Kolekar
Principal

A. P. SHAH INSTITUTE OF TECHNOLOGY

Project Report Approval for B.E.

This project report entitled ***(Crime Control using Deep Learning Techniques)*** by ***(Devesh Sali, Harshal Deshmukh, Jeet Manjrekar, Sakshi Rajeshirke)*** is approved for the degree of ***Bachelor of Engineering*** in ***Computer Science & Engineering (Artificial Intelligence & Machine Learning)***, ***2024-25***.

Examiner Name	Signature
1.________________	______________
2.__________________	_______________

Date:

Place:

Declaration

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Devesh Sali 21106016

Harshal Deshmukh 22206008

Jeet Manjrekar 21106061

Sakshi Rajeshirke 22206002

Date:

Abstract

In today's world, ensuring safety and security in both public and private spaces is a growing concern due to rising crime rates. Intelligent surveillance technologies offer new opportunities for crime control through early detection and intervention. This paper proposes an integrated system that combines scream detection and weapon detection to proactively identify potential criminal activity. The scream detection module uses advanced machine learning techniques to analyze real-time audio, identifying distress signals like human screams. By distinguishing between normal background noise and distress sounds, it enables the system to flag incidents that require immediate attention, particularly in places like public transportation, parks, schools, or residential areas.

At the same time, the weapon detection module leverages image and video processing techniques to identify visual cues of dangerous objects such as firearms, knives, and other weapons. Deep learning algorithms trained on vast weapon datasets ensure accurate detection in surveillance footage or live video streams. Once a weapon is identified, the system automatically raises an alert to notify law enforcement or security personnel. The combination of scream and weapon detection enhances traditional surveillance by providing dual-mode detection, improving response times and reducing the dependency on manual monitoring. Deploying such systems in public spaces, commercial areas, and homes can significantly reduce crime rates, creating safer environments for all.

Keywords: Crime control, scream detection, weapon detection, safety and security, intelligent surveillance, real-time detection, machine learning, audio analysis, image processing, deep learning, public safety, crime prevention.

CONTENTS

LIST OF FIGURES

ABBREVAION

AI	*Artificial Intelligence*
ML	*Machine Learning*
CCTV	*Closed Circuit Television*
OpenCV	*Open Source Computer Vision*
ANN	*Artificial Neural Network*
CNN	*Convolution Neural Network*

Chapter 1

Introduction

In an era of rising crime rates, ensuring public safety has become a pressing global priority. Traditional surveillance systems, while effective in monitoring environments, often fall short in terms of real-time threat detection and response. These systems typically rely on human operators to monitor video feeds, which is not only time-consuming but also prone to errors that can delay critical responses. This limitation necessitates the development of more intelligent systems that leverage emerging technologies like artificial intelligence (AI) and computer vision. The proposed crime prevention system integrates two advanced detection modules—scream detection and weapon detection—to address these gaps and enhance situational awareness. The scream detection module uses Artificial Neural Networks (ANN) to process real-time audio signals, distinguishing distressing sounds, such as screams, from ambient noise, offering an early indication of emergencies like assaults or violent crime. Meanwhile, the weapon detection module utilizes OpenCV and MediaPipe to analyze video feeds and identify dangerous objects, such as firearms or knives, thereby enhancing the system's ability to spot threats visually.

The integration of audio-based and visual threat detection provides a multi-layered approach to security. By combining these two technologies, the system can address a broader spectrum of potential threats, including both audible distress signals and the presence of weapons in public spaces, schools, and high-security zones. The AI-powered system enhances the accuracy of threat detection while reducing reliance on human monitoring, which is often slow and error-prone. One of the system's key features is its real-time alert mechanism, which notifies law enforcement or security personnel immediately when a potential threat is detected. This swift notification ensures that responses can be initiated without delay, providing a significant advantage in emergency situations. Additionally, the system's real-time analysis capabilities minimize the time between threat identification and intervention, potentially preventing crimes before they escalate further.

The proposed system is also designed to be scalable and adaptable to a variety of environments, making it suitable for deployment in diverse settings ranging from urban centers to private institutions. It can effectively operate in environments with varying background noise or lighting conditions, thanks to its machine learning models that are continually optimized for better performance. However, as with any surveillance technology, the system must address potential privacy concerns. The ethical considerations surrounding surveillance, particularly regarding the protection of personal data and privacy rights, are significant. This system is designed with privacy safeguards in place, ensuring that data is processed responsibly and in compliance with regulations. Overall, the AI-driven surveillance framework represents a significant advancement in crime prevention technology, offering a more effective, accurate, and timely response to security threats compared to traditional manual methods.

The ability to detect potential threats quickly and accurately has the potential to transform how law enforcement and security agencies approach public safety. As the system evolves, its scalability and adaptability will allow it to be implemented in a wide range of environments, from schools and hospitals to urban centers and high-security zones. Moreover, the ongoing development of AI and machine learning technologies promises even greater potential for enhancing the system's capabilities, making it an essential tool in the fight against rising crime rates. As we move forward into an increasingly technology-driven world, intelligent surveillance systems like the one proposed in this paper will play a crucial role in ensuring the safety and security of our communities.

Chapter 2

Literature Survey/ Existing system

HISTORY

The evolution of intelligent surveillance systems for crime control is rooted in the broader development of security technologies that began gaining traction in the mid-20th century. Early crime control efforts primarily relied on human intervention, with law enforcement officers patrolling areas and responding to incidents based on citizen reports. The introduction of Closed-Circuit Television (CCTV) marked a significant milestone, allowing for continuous monitoring of public and private spaces and providing valuable post-incident evidence. However, these systems were heavily dependent on human operators, often resulting in delayed responses and susceptibility to human error.

As urban areas expanded in the late 20th and early 21st centuries, the limitations of manual surveillance became increasingly apparent. The rise of digital technologies and enhanced computational capabilities facilitated the automation of surveillance systems. Motion detection mechanisms were integrated with CCTV setups to alert operators of unusual activity, although these systems lacked the intelligence to interpret complex scenarios or distinguish between harmless and threatening events..

The emergence of artificial intelligence (AI) and machine learning in the 2010s brought a transformative shift to the field of surveillance. Deep learning models such as Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) were applied to pattern recognition tasks, enabling capabilities like facial recognition and anomaly detection. This era marked the beginning of intelligent surveillance, where systems could learn from data and adapt over time.

In recent years, the development of advanced surveillance systems has significantly progressed with the refinement of ANN-based audio analysis, enabling real-time recognition of distress signals such as human screams. These capabilities have proven especially valuable in public environments like parks, streets, schools, and transit hubs, where audio cues may be the first indicators of potential threats. The continued convergence of AI, deep learning, and real-time monitoring technologies is positioning intelligent surveillance as a critical component of modern crime prevention and public safety infrastructure.

LITERATURE REVIEW

Integrated violence and weapon detection plays a vital role in enhancing the security and surveillance systems in today's increasingly unsafe environments. In the study titled "Integrated Violence and Weapon Detection Using Deep Learning" by V. Yadav et al., the authors address the limitations of existing violence detection systems that primarily rely on identifying aggressive body movements, which often fail in recognizing isolated but critical threats such as a single gunshot. To overcome this, a novel hybrid approach is introduced that combines weapon detection with violence recognition. This integration significantly reduces false positives and improves the overall system accuracy. The methodology utilizes convolutional neural networks (CNNs) for spatial feature extraction and long short-term memory (LSTM) networks for learning temporal relations, supported by transformer-based models like Vision Transformer for more refined feature representation and improved surveillance accuracy. [1]

In the paper "A Comparative Analysis of Weapons Detection Using Various Deep Learning Techniques" by S. Tamboli et al., the authors highlight the critical importance of object detection in crime scene analysis, where accuracy and speed are essential for timely and effective investigation. They emphasize the need for meticulous manual annotation of images to train deep learning models effectively. These annotated images provide ground truth data essential for training various object detection algorithms. By comparing different deep learning techniques, the study identifies their strengths and weaknesses in weapon detection and evaluates them based on parameters such as accuracy, speed, and efficiency, making a significant contribution to real-time threat detection systems. [2]

In "BERSting at the Screams: Recognition of Shouted and Distressed Speech from Smartphone Recording" by A. Sharma and S. Kaul, the focus is on distinguishing between various types of shouted speech to recognize distress signals. The study addresses a gap in existing emotion recognition systems which often fail to detect distressed speech when audio is recorded in non-ideal conditions. A unique dataset of emotionally charged speech recorded by 96 actors using smartphones was created, incorporating real-world obstacles like varied distances and obstructions. This dataset was then used to train paralinguistic models capable of detecting distressed speech, contributing significantly to real-time audio-based threat detection systems. [3]

The work titled "Real-Time Weapons Detection System using Computer Vision" by P. Nale et al. underscores the need for automated, intelligent surveillance systems to complement CCTV setups. Traditional CCTV monitoring is limited by human error and fatigue, which can lead to delayed or missed threat identification. This paper presents a deep learning-based approach to real-time weapon detection using computer vision techniques. Despite advancements in surveillance technologies, challenges remain due to image obstructions, varying angles, and lighting conditions. The paper proposes a cost-effective and efficient model that aims to reduce false positives and operate effectively with minimal human intervention. [4]

In "Artificial Neural Network using Image Processing for Digital Forensics Crime Scene Object Detection," D. Devasenapathy et al. explore how digital forensics can be enhanced using Artificial Neural Networks (ANNs) for accurate object detection in crime scene imagery. By leveraging various preprocessing techniques like enhancement, segmentation, and feature extraction, this study focuses on increasing the accuracy of weapon identification. The implementation spans across multiple domains including airport security and medical diagnostics, offering a multidisciplinary benefit of the model, and further establishes the importance of ANN in forensic sciences. [5]

"Automatic Weapon Detection using Deep Learning" by P. Akshaya et al. presents a YOLOv8-based deep learning solution for automatic weapon detection in CCTV footage. The authors explain the need for faster and more accurate surveillance systems, particularly in public places such as schools, hospitals, and museums. This model uses an extensive dataset sourced from YouTube robbery videos, GitHub repositories, CCTV footage, and movie scenes to train the detection algorithm. A sliding window technique is employed for feature extraction, and the YOLOv8 model is praised for its speed and accuracy compared to traditional models, making it highly applicable in real-time scenarios. [6]

In the study "Weapon Detection from Surveillance Images using Deep Learning" by A. Goenka and K. Sitara, the emphasis is on reducing false positives in firearm detection using a dataset-driven approach guided by a deep CNN classifier. The problem is reframed to focus on minimizing errors in weapon identification, especially in high-density areas or locations prone to crime. By evaluating different CNN architectures and implementing region proposal techniques, this study provides valuable insights into designing more reliable surveillance systems. [7]

"Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications" by A. Kiran et al. explores the use of Mask RCNN for handgun detection in surveillance footage. The authors highlight the limitations of traditional surveillance systems and propose deep learning methods as a superior alternative. Gaussian deblurring is applied to enhance image quality, especially in blurry footage, to improve weapon detection accuracy. The paper emphasizes the necessity of equipping current surveillance systems with AI-based detectors to enable quicker response times in potential threat scenarios. [8]

In "Identification of Illicit Activities & Scream Detection using Computer Vision & Deep Learning" by R. Mathur et al., the authors focus on creating an automated system that reduces manual labor and increases detection accuracy in surveillance systems. By using deep learning algorithms to process vast volumes of video data, the proposed solution can recognize abnormal activities such as scream-based distress signals. The paper advocates for intelligent surveillance systems to proactively identify and prevent crimes in real-time. [9]

The work "Real-Time Handgun Detection in Surveillance Videos based on Deep Learning Approach" by P. T et al. addresses the challenge of detecting handguns in real-world, unconstrained environments through video surveillance. Highlighting the issue of human oversight in traditional monitoring systems, the authors propose a deep learning-based automatic detection system capable of identifying handguns of varying sizes and orientations. The study aims to improve public safety by enhancing the ability of surveillance systems to detect threats efficiently and autonomously. [10]

"An Automatic Detection of Military Weapons Using Multi-Level CNN Architecture" by D. Pahuja and S. Jain introduces a multi-level CNN approach for detecting military weapons in images. Utilizing three convolutional layers and trained on a dataset of 2700 images, the system achieves 97.7% classification accuracy. The model was built using TensorFlow and Keras and provides a scalable solution for military and defense-related applications in image-based threat detection. [11]

Lastly, "Danger Detection for Women and Child Using Audio Classification and Deep Learning" by M. Ashikuzzaman et al. addresses the growing concern of safety for women and children by leveraging audio-based deep learning techniques. The paper critiques existing mobile apps and sensor-based systems for requiring physical interaction during emergencies. Instead, it proposes an automatic audio classification system that detects screams as a sign of danger. Using deep neural networks, the system effectively distinguishes between normal and distress audio, thus offering a non-intrusive and intelligent solution to improve safety in public and private settings. [12]

Chapter 3

Limitation of Existing system

The current crime detection and surveillance systems, primarily relying on traditional methods like Closed-Circuit Television (CCTV) and manual monitoring, face numerous limitations that hinder their effectiveness in maintaining public safety. Despite the widespread adoption of these systems, their dependency on human intervention and lack of advanced analytical capabilities results in several critical challenges:

- Heavy Reliance on Human Oversight: Traditional CCTV systems require constant human monitoring, which leads to fatigue and reduced attention spans among security personnel. As a result, incidents are more likely to be missed, and responses may be delayed.
- Inefficient Handling of High Data Volumes: Surveillance systems generate vast amounts of audio and video data, making it impractical for human operators to analyze everything in real time. This often leads to overlooked events and undetected security threats.
- Limited Analytical Capabilities: Conventional systems are not equipped to autonomously detect abnormal activities such as screams, weapon appearances, or physical altercations. They rely on manual intervention to interpret and respond to potential threats.

- High False Positive and Negative Rates: Existing automated detection systems frequently encounter issues with high false alarm rates due to environmental noise, varying lighting conditions, and crowded scenes. Conversely, real incidents might be missed due to occlusions or poor video quality.
- Inability to Integrate Audio and Visual Data: Most current surveillance systems treat audio and video streams separately, missing valuable opportunities to correlate both data types for improved threat detection. This lack of integration results in incomplete situational awareness.
- Delayed Response Times: In the absence of real-time processing and automated alerts, there is often a significant delay between incident detection and response, which can compromise public safety and the effectiveness of intervention measures.

Chapter 4

Problem Statement and Objective

4.1 Problem Statement

Rising crime rates in public spaces have become a growing concern, posing significant threats to safety and security. Traditional surveillance methods, such as CCTV cameras and manual monitoring, rely heavily on human oversight, often leading to inefficiencies in threat detection and response. Security personnel face challenges in processing large volumes of surveillance data, which results in delayed responses and overlooked incidents. Furthermore, conventional systems lack the capability for real-time threat detection, making it difficult to recognize distress signals such as screams or visually identify dangerous objects like weapons. The absence of advanced analytical capabilities in traditional methods highlights the need for a more intelligent and efficient system for real-time crime detection.

To address these limitations, an AI-powered surveillance system must integrate both audio and visual analysis to enhance security measures. Scream detection, implemented using Artificial Neural Networks (ANN), enables the system to identify high-pitched distress sounds while effectively filtering background noise. Additionally, weapon and fight detection leverage computer vision techniques using OpenCV and MediaPipe to recognize weapons and physical altercations in surveillance footage. However, real-world challenges such as background noise, crowd density, varying lighting conditions, and occlusions can impact detection accuracy. Technical hurdles also arise in ensuring low false positive rates, maintaining high precision in complex environments, and achieving real-time processing capabilities.

Given these challenges, an advanced AI-powered system is essential to enhance public safety by enabling faster response times and improving crime prevention measures. By incorporating real-time audio and visual analysis, such a system can significantly reduce human dependency while providing a robust, efficient, and scalable solution to modern security concerns.

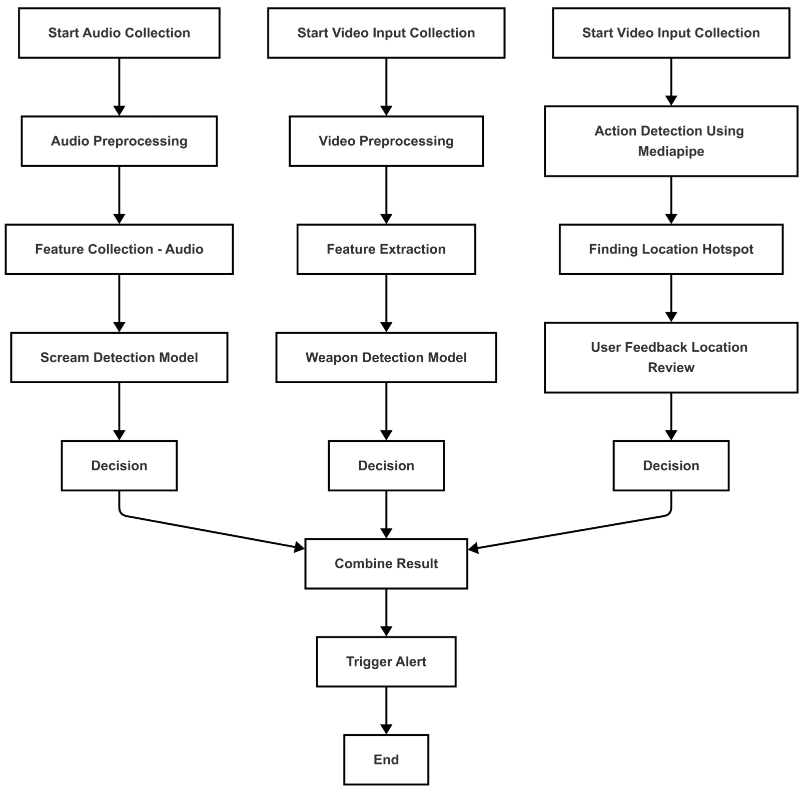
4.2 Objective

- Develop an AI-powered Crime Detection System: Build an intelligent system that autonomously identifies potential threats in real-time.
- Implement Scream Detection: Use Artificial Neural Networks (ANN) to accurately detect distress signals while filtering out background noise.
- Perform Weapon and Fight Detection: Leverage OpenCV and MediaPipe for computer vision-based identification of dangerous objects and violent altercations.
- Ensure Real-time Processing: Achieve high-speed analysis and detection without human intervention.
- Enhance Detection Accuracy: Minimize false positives and ensure high precision in diverse environments.
- Overcome Environmental Challenges: Address issues like background noise, crowd density, varying lighting conditions, and occlusions.
- Enable Faster Emergency Response: Support quicker decision-making and response through timely threat identification.
- Improve Public Safety: Strengthen crime prevention capabilities in urban environments with an advanced surveillance system.

Chapter 5

Proposed System

Block diagram:



Description of Block Diagram:

Multimodal Threat Detection and Alert System

This system is designed to proactively detect and respond to violent or emergency situations in real-time using a fusion of audio, video, and geolocation data. It leverages AI-powered models for scream detection, weapon detection, and action analysis to ensure a high level of public safety and surveillance accuracy. The core idea is to process and analyze environmental signals to trigger alerts that help authorities respond swiftly.

1. Start Audio Collection

The system begins by continuously capturing environmental audio signals using high-quality microphones. This component focuses on gathering natural human sounds, especially those that may signal distress, such as screaming, shouting, or loud cries. These real-time sounds form the foundation for the scream detection pipeline.

2. Audio Preprocessing

Once the audio is collected, it is passed through a preprocessing module that enhances signal quality. This step includes:

- Noise reduction to eliminate background disturbances
- Segmentation to divide the audio into manageable time frames
- Normalization to bring sound amplitude to a common scale

This ensures the model is not biased by environmental noise and is only analyzing meaningful sound patterns.

3. Feature Collection – Audio

After cleaning, relevant audio features are extracted. Common features include:

- MFCCs (Mel Frequency Cepstral Coefficients)
- Spectral Centroid, Zero-Crossing Rate
- Pitch and Energy Levels

These features help in differentiating screams from other noises like talking or music, forming the input to the scream detection model.

4. Scream Detection Model

The processed audio features are input to a pre-trained deep learning model (e.g., CNN, RNN, or LSTM) that identifies the presence of a scream or distress sound. The model outputs a decision with a confidence score, indicating whether a threatening sound is detected.

5. Start Video Input Collection (Weapon Detection Path)

At the same time, video feeds are captured using surveillance cameras or CCTV systems. These video streams are focused on visually detecting threatening objects such as weapons (guns, knives, etc.) through object detection models.

6. Video Preprocessing

The raw video frames undergo preprocessing such as:

- Frame resizing
- Color space conversion
- Frame skipping to reduce redundancy

These ensure consistency in input and reduce computational load for real-time processing.

7. Feature Extraction (Weapon Detection)

Important features like object shapes, edges, and bounding boxes are extracted from video frames. These features are passed to the detection model to identify the presence of lethal objects.

8. Weapon Detection Model

A state-of-the-art YOLOv8 or CNN-based weapon detection model is used here. It scans each frame and predicts the presence and location of any weapons. If a weapon is detected, it passes a positive decision along with the detection confidence to the decision fusion module.

9. Start Video Input Collection (Action Detection Path)

Another video input pipeline begins simultaneously for behavioral or action-based threat detection. This focuses on identifying unusual, aggressive, or suspicious actions through pose tracking and body movement analysis.

10. Action Detection using MediaPipe

Using MediaPipe, the system tracks human skeletons (joints and limbs) in real-time to detect abnormal or violent motions such as:

- Hitting
- Rapid movements
- Unusual gestures or chases

MediaPipe's lightweight and efficient architecture enables this to run in real-time with high accuracy.

11. Finding Location Hotspot

Once a suspicious action is identified, the system performs location mapping using GPS tags, IP tracking, or camera metadata. This helps to pinpoint the location of the detected incident, and associate it with previous incident zones.

12. User Feedback Location Review

To minimize false positives, user feedback from trusted personnel or past incident data is considered. This module reviews location credibility and updates hotspot maps, enhancing decision-making for future detections.

13. Decision Modules

Each of the three pipelines (audio, video-weapon, video-action) makes an independent decision:

- Audio Decision: Was a scream detected?
- Video Decision: Was a weapon detected?
- Action Decision: Was a suspicious activity observed?

Each decision includes a confidence score and timestamp

14. Combine Result

The decisions from all three sources are fused together in the combine result module. This uses majority voting, threshold logic, or weighted scoring to determine if the situation requires emergency response.

Example fusion logic:

- If any two out of three modalities are positive → Trigger Alert
- If all three are positive → High priority alert
- If only one is positive but confidence > 90% → Send for manual review

15. Trigger Alert

Once a positive threat is identified, the system triggers an alert.
The alert includes:

- Audio/Video snippets
- Timestamp
- Location Coordinates
- Type of threat (Weapon/Scream/Action)
- Alerts can be sent via:
- SMS or Email to security personnel
- Notification to a control dashboard
- API trigger to emergency systems

16. End

After the alert is triggered, the system completes its loop and resets for continued monitoring. All data is logged for audit and retraining purposes.

Methodology

The system collects real-time audio and video input to detect potential threats. Audio input undergoes preprocessing and feature extraction to detect screams using a deep learning model. Simultaneously, video input is used for two tasks: weapon detection through a YOLO-based model and action detection using MediaPipe for suspicious behavior. Location hotspots are identified and verified through user feedback. Each module makes a decision independently, and these results are combined. If a threat is confirmed, an alert is triggered and sent to authorities for quick action.

Chapter 6

Experimental Setup

6.1 Hardware Setup

To ensure smooth execution of the project, it is essential that the computer system used meets certain hardware requirements. These requirements will depend on the complexity of the tasks involved, especially when dealing with resource-intensive software like machine learning frameworks, image and video processing, and virtual environments. Below are the recommended specifications for optimal performance.

Processor:

A multi-core processor, such as Intel Core i3 or AMD Ryzen 3 (or equivalent), is the minimum recommended. For more complex projects, an Intel Core 5 or AMD Ryzen 5 is ideal, offering better data processing and analysis speeds.

RAM:

A minimum of 8 GB is required, but 16 GB is strongly recommended for tasks like deep learning, handling large datasets, and running virtual environments efficiently.

Storage:
At least 256 GB of storage is needed, preferably on a Solid-State Drive (SSD) for faster file access and system boot times.

Operating System:
Compatible OS options include Windows 10/11, macOS 10.15 or higher, or Linux (Ubuntu 20.04+), with Linux being preferred for Python-based AI development due to package management ease.

Monitor:
A HD (1920x720) display is essential for clarity when coding or analyzing data. A higher resolution display may benefit image or video quality.
Other Peripherals:Essential peripherals include a keyboard, mouse, and high-speed internet. Depending on project needs, external hardware such as Arduino boards, Raspberry Pi, or cloud storage may also be required.

6.2 Software Setup

The software setup for the project involves multiple tools, libraries, and frameworks that are essential for development, testing, and deployment. Ensuring proper installation and configuration of each component is crucial to avoid issues during execution. Below is a detailed breakdown of the required software and their purposes:

- Operating System:

  The project is compatible with modern operating systems like Windows, macOS, and Linux (preferably Ubuntu), all of which support the tools and frameworks necessary for development.

- Primary Development Tools:

  Python 3.x:Python is the primary language for the project due to its simplicity and the availability of powerful libraries for AI, ML, and data analysis. Python 3.x is recommended over Python 2.x for continued support and improved features.

  IDE/Code Editor:Suitable development environments include Visual Studio Code (lightweight and extensible), PyCharm (Python-focused with advanced features), and Jupyter Notebook (ideal for data visualization and inline documentation).

  Libraries and Packages:

  TensorFlow/Keras:Used for developing and training machine learning and deep learning models, offering tools for neural network construction and efficient data handling.

  OpenCV:Essential for image processing and computer vision tasks like object detection, real-time image analysis, and facial recognition.

  NumPy & Pandas:NumPy supports fast numerical operations on arrays and matrices, while Pandas is used for structured data manipulation and preprocessing.

  Matplotlib & Seaborn:These libraries are used for generating plots and visual representations to analyze datasets and model performance.

  scikit-learn:Offers implementations of various traditional machine learning algorithms for classification, regression, and clustering.

- Additional Tools:

  VMware/VirtualBox:Used to create isolated virtual environments for tasks such as network simulation, sandbox testing, and cross-platform validation.

  XAMPP/WAMP:Provides a local server environment to develop and test web applications with Apache, MySQL, and PHP/Perl support.

- Version Control:

  Git:Git enables efficient version control and team collaboration. Code can be pushed to remote repositories on GitHub or GitLab for backup, review, and deployment.

- Collaboration and Project Management:

  Installation and Configuration:All tools and packages should be correctly installed and configured. It is recommended to use Python virtual environments (like venv or conda) to manage dependencies and avoid conflicts.

6.3 Implementation

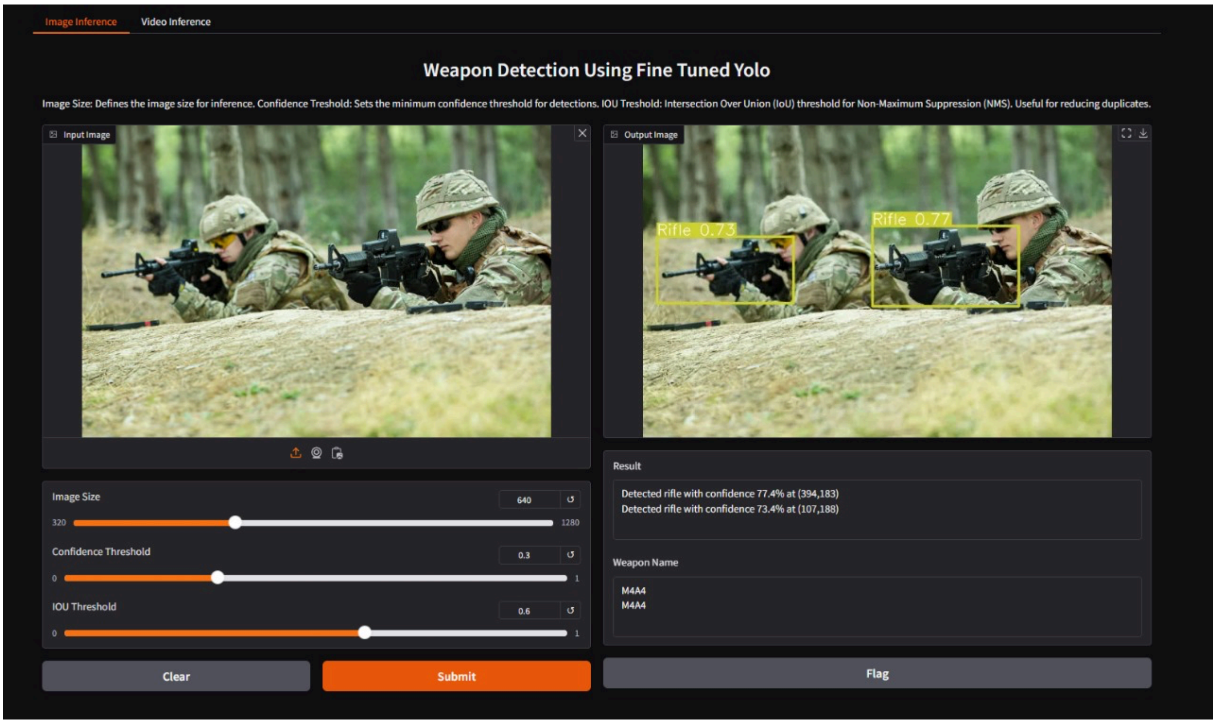


Fig 1

This interface demonstrates the potential of real-time weapon detection through an optimally tuned YOLO model. The system scans images and video streams to detect potential threats such as firearms like pistols, knives, rifles, and other weapons. The input image is displayed in the left pane, and the output with detected objects in the form of bounding boxes in the right pane. Each detection has a confidence score, which helps in ensuring accuracy. Users can adjust parameters like image size, confidence thresholds, and Intersection Over Union (IoU) thresholds to improve detection efficiency. The system is also capable of real-time inference through a camera interface, and it is therefore a handy tool for security-related applications. Results can also be tagged for further analysis or verification by security agents

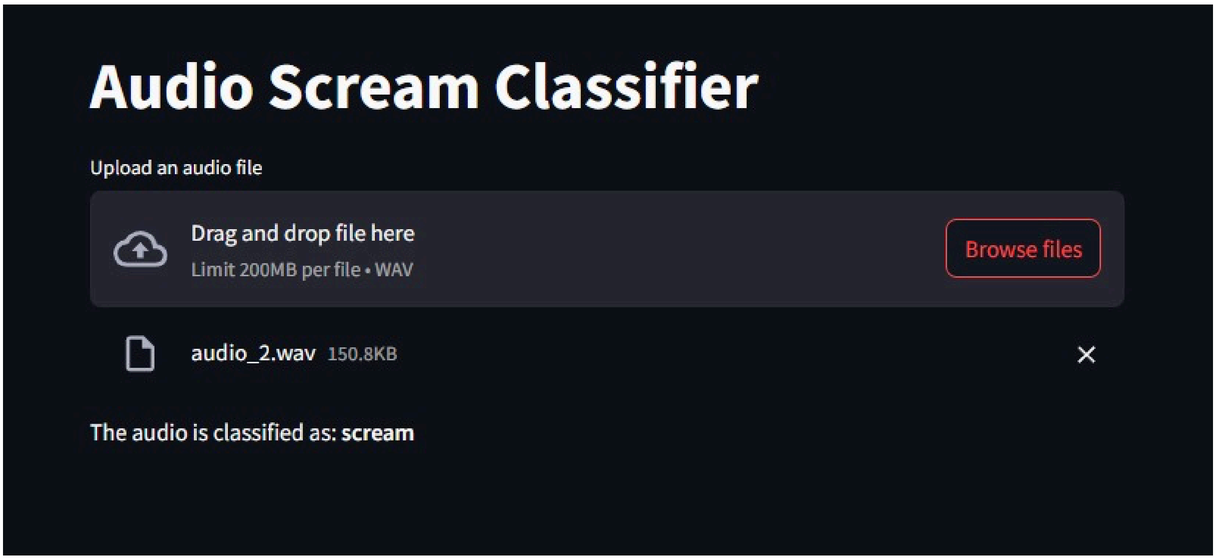


Fig 2

The Audio Scream Classifier uses an Artificial Neural Network (ANN) to identify scream sounds through processing of uploaded audio tracks. Audio is processed to filter out noise and extract salient features such as MFCCs and spectral features. These are fed into a trained ANN to decide whether or not the sound is a scream. When the sound is identified as a scream, the system can activate alerts for emergency response. The technology improves public safety through real-time scream detection in security and surveillance systems, which results in enhanced threat detection and more efficient emergency response.

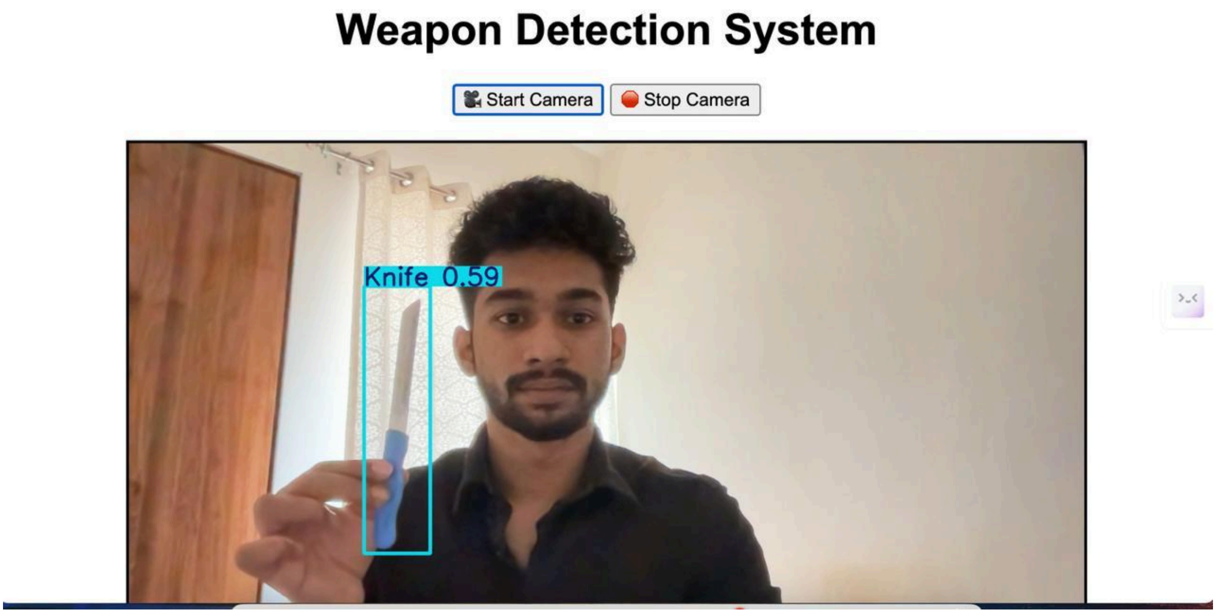


Fig 3

The picture given is of a weapon detection system with a computer vision model that identifies objects like grenades, knives, and pistols in real time. The user interface is a camera stream where the detected objects are outlined with bounding boxes and annotated with their corresponding names and an associated confidence value. In this case, the system identified a knife, gave it a confidence value of 0.59, and ringed it with a blue bounding box. The system can process video inputs and identify suspected threats with a trained deep-learning model. These systems can be utilized in security surveillance, law enforcement, and public threat detection, thus improving security measures by creating real time alerts. The fact that it is able to detect several weapons gives it a high potential as an automated security device.

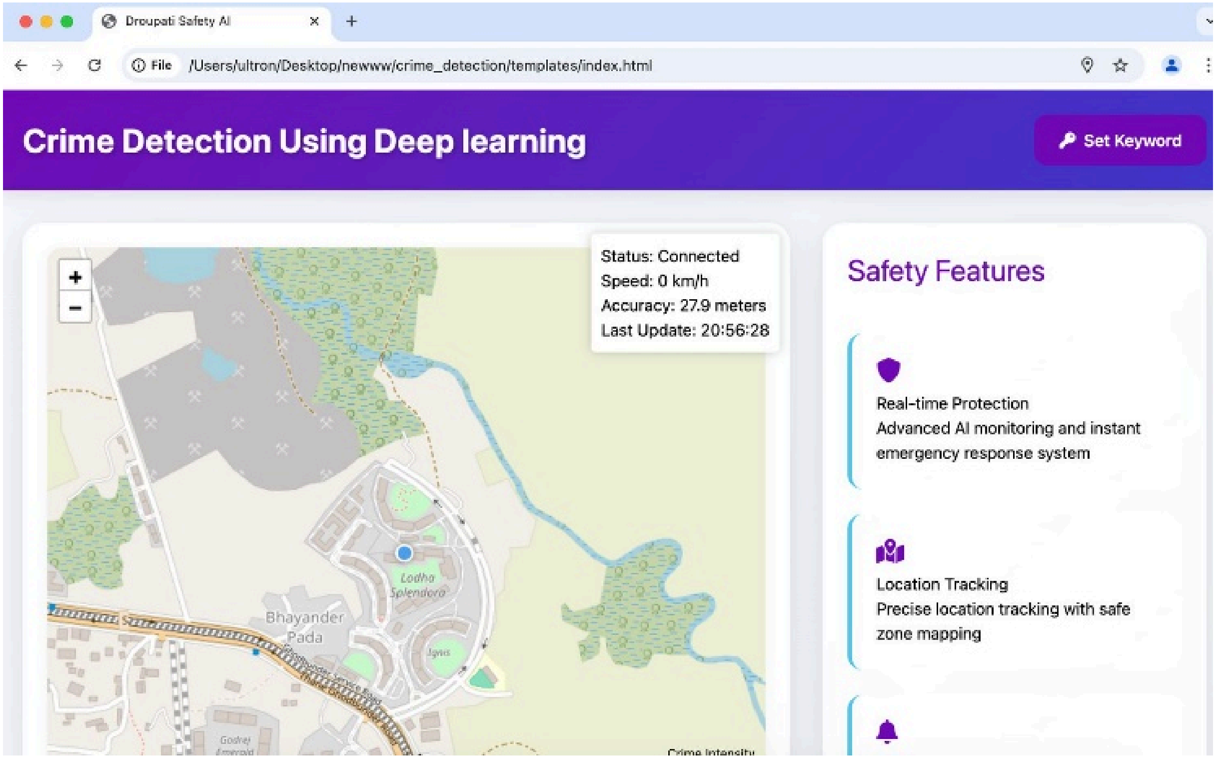


Fig 4

This is the primary home page of the UI, redirected after displaying the alert page. It indicates the user's present location with the Geocoders API or IP address for correct tracking of location. The interface offers real-time status updates regarding speed, precision, and timestamp of the last update. Also, the system incorporates security functions like real-time AI monitoring and emergency response functionality. The visual map facilitates individuals in understanding where they are with marked potential risks and safety points. The "Set Keyword" button facilitates adjusting alerts in response to predefined triggers for ease of use and security. The ease of use offers a smooth experience for users, increasing the effectiveness of crime detection and response.

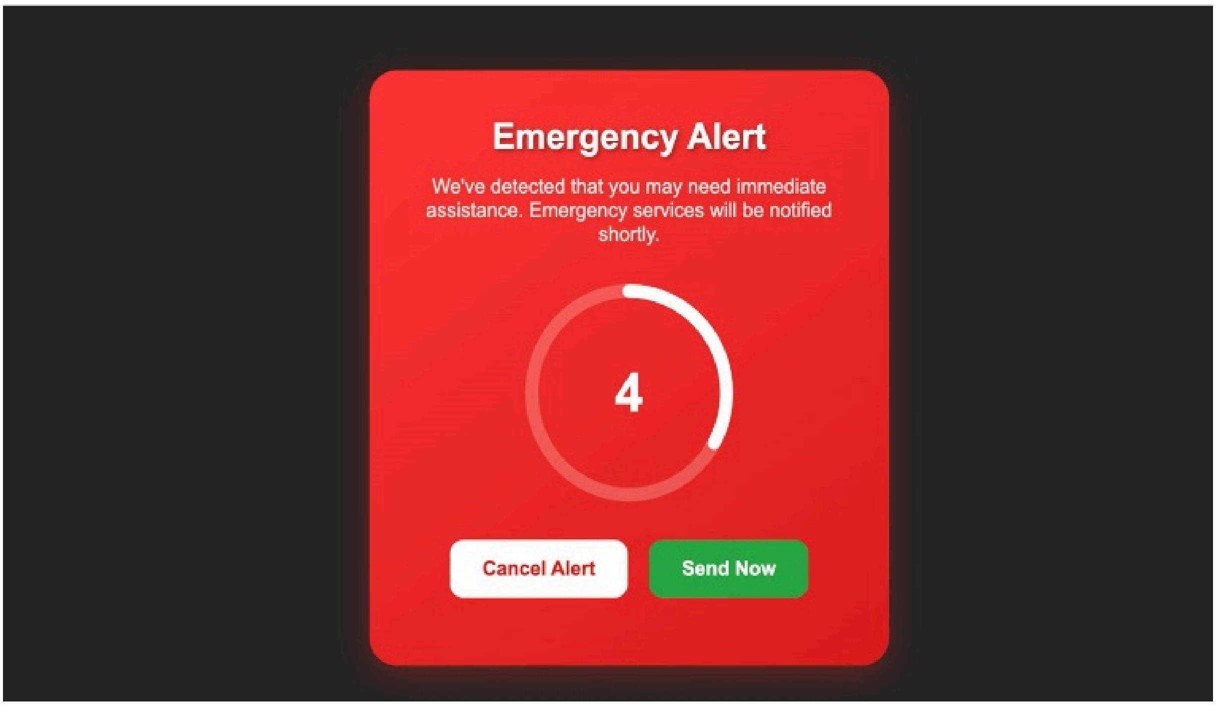
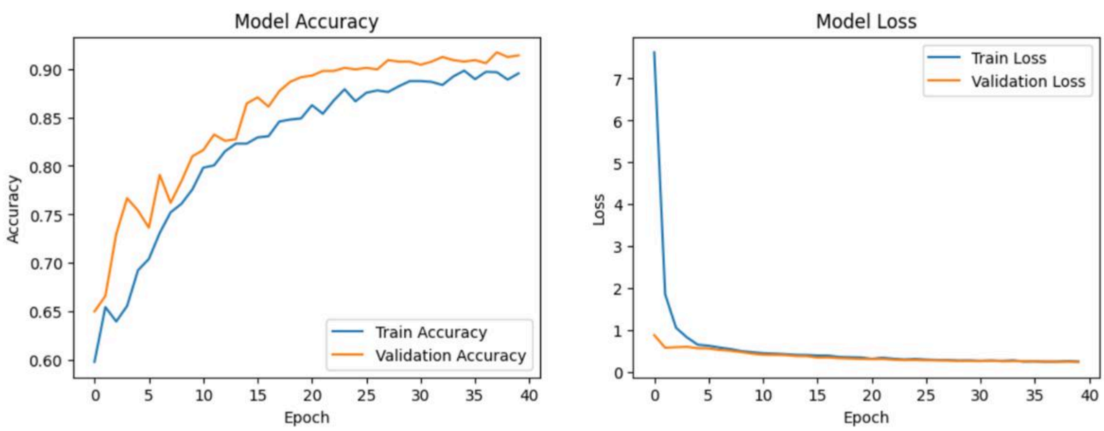


Fig 5

The emergency alert system identifies irregularities in CCTV video, including violence or weapon displays, such as knives or guns. With artificial intelligence and deep learning techniques, it examines real-time video streams and sends an alert when a potential threat is identified. There is an option for a countdown for the cancellation of false alerts or to alert emergency services directly. In the absence of action, authorities are alerted automatically, enabling swift action and increased public safety




```
[19]
# Evaluate the model on the training data
loss, accuracy = model.evaluate(X, y)
print(f"Training accuracy: {accuracy:.4f}")

98/98 ━━━━━━━━━━━━━━━━━━━━ 1s 6ms/step - accuracy: 0.9244 - loss: 0.2052
Training accuracy: 0.9239
```

Fig 6

The Model Performance Metrics table indicates significant evaluation measures for the trained model, measuring performance on training and validation sets. Final Accuracy is 92.44 (0.9244) on training and approximately 91 on the validation set, indicating very good learning capacity with minimal overfitting. The Final Loss measures are 0.2052 for training and approximately 0.20 for validation, indicating good convergence. The Training Trend indicates a consistent increase in accuracy, which plateaus at 92, whereas validation accuracy indicates a similar trend, sometimes higher than training accuracy, thereby indicating a well-generalized model. The Loss Trend indicates a steep decline in loss in the early epochs before plateauing, with validation loss following training loss very closely, thereby indicating the consistency of the model.

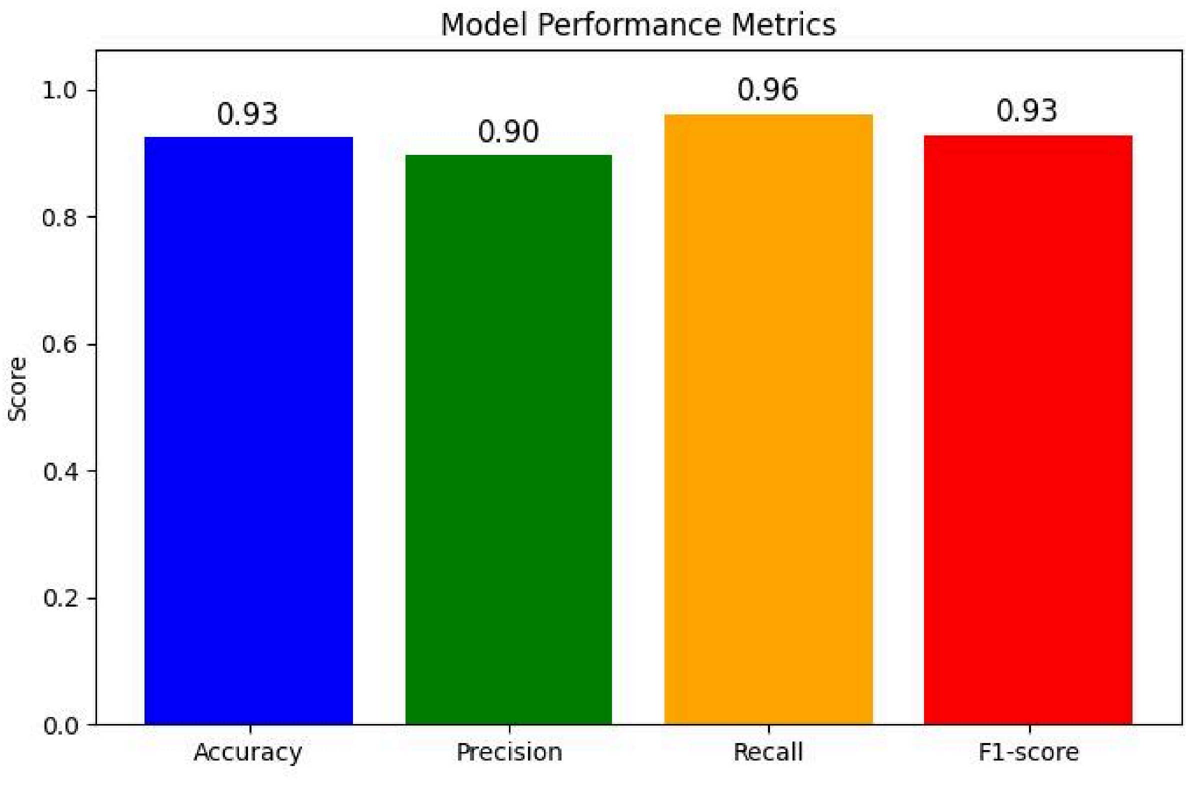


Fig 7

1) Evaluation Metrics After rigorous training and testing, we achieved the follow ing performance scores: • Accuracy: 0.93– The model correctly classified 93% of the input audio samples, indicating strong overall reliability. • Precision: 0.90– Precision reflects how many of the detected screams were actual screams. A score of 90% means that the model minimizes

false positives effectively. • Recall: 0.96– The model successfully identified 96% of all actual screams in the dataset, demonstrating its ability to detect distress signals efficiently. • F1-score: 0.93– The F1-score provides a balance between precision and recall, confirming that our model effectively identifies distress sounds while controlling false alarms. 2) Interpretation of Results The results show that the suggested model is very suitable for real-time scream classification. The high recall value ensures the identification of most distress sounds, thus making it reliable for use in applications that need to detect emergencies. Additionally, the high precision value minimizes the occurrence of false alarms while efficiently detecting necessary sounds. From these measurements, it is safe to conclude that our scream detector is capable of distinguishing between distress calls and normal background noise, making it a viable tool for surveillance and security applications.

Chapter 7

Results and Discussion

The implementation of the real-time Weapon and Scream Detection System using deep learning models and user-friendly interfaces demonstrated highly encouraging results. By leveraging a fine-tuned YOLO-based object detection model and an ANN-based audio classifier, the system effectively identified potential security threats from both visual and audio data sources with impressive accuracy and reliability.

The weapon detection interface processes both image and video streams, marking firearms (e.g., pistols, rifles), knives, grenades, and other weapons with bounding boxes and confidence scores. The system allowed parameter customization including image resolution, confidence threshold, and IoU threshold, offering flexibility in detection precision. Real-time inference via webcam and the ability to tag results for further analysis make this solution highly suitable for security applications such as surveillance systems in public places, transport hubs, and critical infrastructure.

Similarly, the Audio Scream Classifier, built using an Artificial Neural Network (ANN), efficiently classified uploaded audio files. The system extracted MFCCs and spectral features to filter noise and identify scream patterns. When a scream was detected, real-time alerts could be triggered, enhancing situational awareness and enabling rapid emergency responses in areas such as public spaces, smart buildings, or residential monitoring systems.

The performance metrics of both subsystems further validate the system's effectiveness. The weapon detection model achieved a training accuracy of 92.44% and a validation accuracy of approximately 91%, showing a well-generalized model with minimal overfitting. Loss curves indicated proper convergence with training and validation loss values around 0.205 and 0.20, respectively.

For the scream detection model, the system achieved:

• Accuracy: 0.93 – The model correctly classified 93% of the input audio samples, indicating strong overall reliability.

• Precision: 0.90 – Precision reflects how many of the detected screams were actual screams. A score of 90% means that the model minimizes false positives effectively.

• Recall: 0.96 – The model successfully identified 96% of all actual screams in the dataset, demonstrating its ability to detect distress signals efficiently.

• F1-score: 0.93 – The F1-score provides a balance between precision and recall, confirming that the model effectively identifies distress sounds while controlling false alarms.

These values reflect the model's ability to correctly detect distress calls with minimal false positives, making it a practical tool for real-world deployment.Moreover, the emergency alert system complements these models by monitoring CCTV streams and triggering alerts for weapon displays or violent actions. Users have the option to cancel false positives via countdown timers, ensuring practical usability and avoiding unnecessary escalations. The system also integrates geolocation services to map threat zones in real time, improving situational response strategies.

Despite the promising results, some limitations were observed, such as occasional false detections due to camera lighting, background noise, or occlusions. These challenges suggest the need for improved environmental normalization and additional training with diverse datasets. Nevertheless, the successful integration of YOLO and ANN with real-time interfaces confirms the potential of combining computer vision and audio analytics for intelligent surveillance and threat detection.

Chapter 8

Conclusion and Future Work

In conclusion, the developed Crop Recommendation System using ML and IoT bridges the gap between traditional farming practices and modern technology, empowering farmers with data-driven insights. The system effectively collects real-time environmental data, preprocesses it, extracts meaningful features, and uses machine learning algorithms to recommend the most suitable crops. The integration of IoT sensors, OpenWeather API, and a rainfall prediction model ensures recommendations are dynamic and adaptive to changing conditions, enhancing agricultural productivity.

The project not only optimizes crop selection but also minimizes input costs and promotes sustainable farming practices. By leveraging real-time monitoring and predictive analytics, farmers can make timely decisions, improve resource management, and mitigate the impact of unpredictable weather events.

For future work, the system can be enhanced in several ways:

Model Optimization: Experiment with advanced ML algorithms like XGBoost or LSTM models for improved accuracy and long-term weather pattern recognition.

Expanded Sensor Integration: Add NPK sensors for direct nutrient level monitoring, enhancing fertilizer recommendations.

Multi-Language Support: Enable multilingual interfaces to cater to diverse farming communities.

Scalability & Cloud Integration: Shift to a cloud-based architecture for better scalability, allowing data aggregation from multiple farms to train more generalized models.

By addressing these enhancements, the system can evolve into a comprehensive smart agriculture platform, capable of driving global agricultural transformation and ensuring food security through intelligent, sustainable farming practices.

References

Research paper

[1] V. Yadav, S. Kumar, A. Goyal, S. Bhatla, G. Sikka and A. Kaur, "Integrated Violence and Weapon Detection Using Deep Learning," 2024 First International Conference on Pioneering Developments in Computer Science & Digital Technologies (IC2SDT), Delhi, India, 2024, pp. 563-568, doi: 10.1109/IC2SDT62152.2024.10696591

[2] S. Tamboli, K. Jagadale, S. Mandavkar, N. Katkade and T. S. Ruprah, "A Comparative Analysis of Weapons Detection Using Various Deep Learning Techniques," 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2023, pp. 1141-1147, doi: 10.1109/ICOEI56765.2023.10125710..

[3] "BERSting at the Screams: Recognition of Shouted and Distressed Speech from Smartphone Recordings," 2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), Cambridge, MA, USA, 2023, pp. 1-5, doi: 10.1109/ACIIW59127.2023.10388214. [4] S. E. C. Osman, H. Jantan, M. T. Miskon, and W. A. K. W. Chek, “A comparative study of video coding standard performance via local area network,” in International Conference on Soft Computing in Data Science. Springer, 2015, pp. 189–197.

[4] P. Nale, S. Gite and D. Dharrao, "Real-Time Weapons Detection System using Computer Vision," 2023 Third International Conference on Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 2023, pp. 1-6, doi: 10.1109/STCR59085.2023.10396960.

[5] D. Devasenapathy, M. Raja, R. K. Dwibedi, N. Vinoth, T. Jayasudha and V. D. Ganesh, "Artificial Neural Network using Image Processing for Digital Forensics Crime Scene Object Detection," 2023 2nd International Conference on Edge Computing and Applications (ICECAA), Namakkal, India, 2023, pp. 652-656, doi: 10.1109/ICECAA58104.2023.10212302.

[6] P. Akshaya, P. B. Reddy, P. Panuganti, P. Gurusai and A. Subhahan, "Automatic weapon detection using Deep Learning," 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 2023, pp. 1-7, doi: 10.1109/RMKMATE59243.2023.10369889.

[7] A. Goenka and K. Sitara, "Weapon Detection from Surveillance Images using Deep Learning," 2022 3rd International Conference for Emerging Technology (INCET), Belgaum, India, 2022, pp. 1-6, doi: 10.1109/INCET54531.2022.9824281.

[8] A. Kiran, P. Purushotham and D. D. Priya, "Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications," 2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC), Bhubaneswar, India, 2022, pp. 1-5, doi: 10.1109/ASSIC55218.2022.10088403.

[9] R. Mathur, T. Chintala and D. Rajeswari, "Identification of Illicit Activities & Scream Detection using Computer Vision & Deep Learning," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2022, pp. 1243-1250, doi: 10.1109/ICICCS53718.2022.9787991.

[10] P. T, R. Thangaraj, P. P, U. R. M and B. Vadivelu, "Real-Time Handgun Detection in Surveillance Videos based on Deep Learning Approach," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2022, pp. 689-693, doi: 10.1109/ICAAIC53929.2022.9793288.

[11] D. Pahuja and S. Jain, "An Automatic Detection of Military Weapons Using multi-Level CNN Architecture," 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2022, pp. 1-4, doi: 10.1109/ICRITO56286.2022.9964878.

[12] M. Ashikuzzaman, A. A. Fime, A. Aziz and T. Tasnima, "Danger Detection for Women and Child Using Audio Classification and Deep Learning," 2021 5th International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, 2021, pp. 1-6, doi: 10.1109/EICT54103.2021.973360.

URL

[1]https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10696591&isnumber=10
69595
[2]https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10125710&isnumber
=1012528
[3]https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10388214&isnumber
=1038807
[4]https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10396960&isnumber
=1039684
[5]https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10212302&isnumber
=1021209
[6]https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10369889&isnumber
=1036860
[7]
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9824281&isnumber=98
23966
[8]https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10088403&isnumber
=1008828
[9]
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9787991&isnumber=97879
85
[10]https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9793288&isnumber
=9792586
[11]
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9964878&isnumber=99644
06
[12]https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9733601&isnumber=973343
0

Appendix

Publication

Siubmitted To Journel

Acknowledgements

We would like to thank our project guide Prof. Sayali Badhan and our project co-ordinator Prof. Sayali Badhan for guiding us throughout the project. We would also like to thank our institute i.e. A. P. Shah Institute of Technology, Thane for providing us with a conducive environment to learn and research.

Devesh Sali 21106016

Harshal Deshmukh 22206008

Jeet Manjrekar 21106061

Sakshi Rajeshirke 22206002

Date: