# *Real Time Crime Detection Using Improved AI Techniques*

*Sayali Badhan*
*Department of CSE (AIML)*
*A.P. Shah Institute of Technology*
*Thane (M.H), India 400615*
*Email:Sayalibadhan@gmail.com*

*Devesh Sali*
*Department of CSE (AIML)*
*A.P. Shah Institute of Technology*
*Thane (M.H), India 400615*
*Email: dvesh.sali28@gmail.com*

*Harshal Deshmukh*
*Department of CSE (AIML)*
*A.P. Shah Institute of Technology*
*Thane (M.H), India 400615*
*Email: work.harshal22@gmail.com*

*Sakshi Rajeshirke*
*Department of CSE (AIML)*
*A.P. Shah Institute of Technology*
*Thane (M.H), India 400615*
*Email: sakshirajeshirke636@apsit.edu.in*

*Jeet Manjrekar*
*Department of CSE (AIML)*
*A.P. Shah Institute of Technology*
*Thane (M.H), India 400615*
*Email: letsmailjeet@gmail.com*

*Tushar D. Ubale*
*Department of CSE (AIML)*
*A.P. Shah Institute of Technology*
*Thane (M.H), India 400615*
*Email: tdubale@apsit.edu.in.com*

***Abstract*—This paper introduces an intelligent surveillance system with scream and weapon detection to improve public safety. The scream detection module utilizes machine learning on audio signals to recognize distress sounds, while the weapon detection module utilizes deep learning on real-time video streams to recognize threatening objects. With the combination of these technologies, the system enables early crime detection and quick response, lowering the dependence on manual surveillance and improving crime prevention. The fusion of audio and visual analysis improves the accuracy of threat detection, lowering false alarms and making the security system more robust. The system is flexible to operate in various environments like public areas, transportation facilities, and schools, with scalable and efficient surveillance solutions. Real-time alerting mechanisms also enable security personnel to respond quickly, preventing possible incidents from growing out of control. This multi-modal threat detection method opens the door to more sophisticated AI-based security systems, facilitating the development of safer communities.**

***Keywords*—Crime prevention, scream detection, weapon detection, intelligent surveillance, public safety, audio analysis.**

## I. Introduction

With the rise in crime, security has become the most important aspect of global conflict. Traditional surveillance systems are effective in some cases, but suffer from a number of limitations: This requires advanced surveillance techniques that actively tackle modern security-related issues.

This work suggests a unified approach to joint battles against crime through screams and weapon recognition using modern AI and computer vision technologies. Scream detection uses artificial neuron networks to analyze real-time audio and filters most of the background noise and pinpoint high-screams that can be notified of emergencies.

Weapon detection uses OpenCV and Mediapipe to enable the system to accurately identify life-threatening objects, such as firearms and knives in video feeds. The proposed paper provides deeper insight into underlying technologies, methodologies, and practical applicability, and places particular emphasis on the ability to improve public security and crime prevention.

## II. Related Work

A variety of studies have investigated the application of deep learning for weapon detection to reinforce security in open areas. The work by Yadav et al. [1] combines CNNs for extracting spatial features and LSTMs to capture spatial relations, enhancing representation quality with frameworks such as Image Transformer. Likewise, Goenka and Sitara [8] utilize CNN-based classifiers coupled with a region proposal method to minimize false positives for surveillance-based weapon detection. Yet another research by Akshaya et al. [7] applies YOLO for real-time detection of weapons in CCTV surveillance in public places, providing high-speed and efficient threat detection. These methods underscore the efficiency of deep learning in improving surveillance and reducing human involvement.

A comparative analysis by Tamboli et al. [2] evaluates multiple deep learning methods for weapon detection in forensic investigations, emphasizing the role of accurate object detection in reconstructing crime scenes. The study underscores the importance of robust datasets and model performance in improving investigative efficiency. Similarly, Nale et al. [5] propose a real-time weapon detection system designed to minimize security vulnerabilities in large public spaces, ensuring reliable threat detection across various conditions. Both studies stress the need for high accuracy and speed in weapon identification to strengthen crime analysis and security measures.

Advancements in firearm detection have also been explored through different deep learning frameworks. Kiran et al. [9] implement Mask R-CNN with Gaussian deblur techniques to enhance firearm recognition from surveillance footage, particularly in blurred images. Meanwhile, P. T. et al. [11] focus on real-time handgun detection in CCTV videos using deep learning object detection techniques to handle challenges like cluttered backgrounds and varying object sizes. Additionally, Pahuja and Jain [12] introduce a multi-level CNN architecture with three convolutional layers, achieving 97.7% accuracy in classifying different weapon types, demonstrating the efficiency of hierarchical feature extraction.

Beyond visual detection, researchers have explored audio-based security solutions for emergency scenarios. Sharma et al. [3] develop a scream recognition system that detects distress sounds from smartphone recordings, considering environmental factors and microphone variations. Similarly, H. Jantan et al. [4] propose an audio-based safety system that classifies screams using deep neural networks, offering a non-invasive method for detecting danger, particularly for women and children. These studies emphasize the potential of sound-based AI systems in personal safety and emergency response applications.

Automated surveillance systems leveraging computer vision and deep learning are gaining traction for security monitoring. Mathur et al. [10] introduce an AI-powered framework that detects illicit activities from security footage, reducing reliance on human oversight. This aligns with broader efforts to integrate deep learning into surveillance, improving safety through real-time abnormal event detection. The collective advancements in deep learning-based weapon and threat detection demonstrate the growing role of AI in enhancing public security and forensic investigations.

## III. Proposed System Architecture

The system combines Scream detection with OpenCV and Mediapipe for Weapons and Battle Detection Mediapipe to allow real-time threat identification, as shown in Fig 1 Audio and video input analysis ensures rapid and accurate detection that improves public security through automated alerts and rapid emergency response. The system continuously monitors the surroundings and uses deep learning techniques to distinguish between normal and potentially dangerous scenarios. Advanced pre-treatment techniques help reduce noise and extract characteristics, ensuring robust detection even under difficult conditions such as overcrowded areas and large backgrounds. Furthermore, integrating several detection mechanisms will incorrectly reduce the positive aspects and improve accuracy. This means that the system for your real application is very reliable. If a threat is recognized, the system can not only trigger warnings, but also send notifications to law enforcement agencies or specific security personnel. Additionally, user intervention mechanisms allow manual review of warnings, preventing unnecessary escalations and ensuring quick responses when necessary.

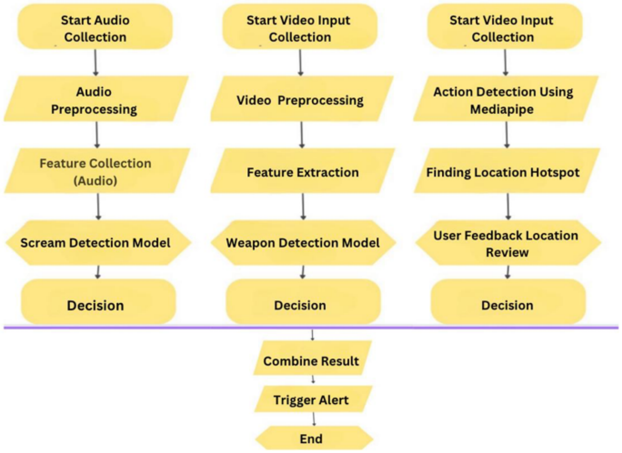


Fig. 1. System Architecture

### A. Scream Detection

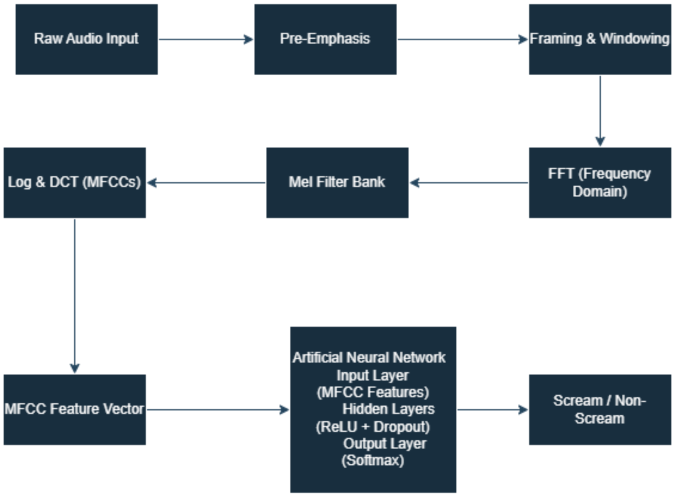


Fig. 2. MFCC-Based Scream Classification Pipeline

Kaggle's SCREAM-ANN dataset is a commercial dataset, designed especially for scream detection for audio security and emergency response applications. The dataset contains a diverse collection of scream recordings, annotated with intensity, emotion, and acoustic attributes. The dataset is essential for training deep learning models to detect screams from ambient sounds, facilitating real-time anomaly detection in surveillance, smart security systems, and emergency alert software. High-quality recordings and well-organized annotations make the dataset suitable for developing robust audio classification models.

The process begins its operation by continuously reading audio input from a microphone, constantly monitoring ambient sounds in real-time. This enables it to sense and screen for any possible distress signals within its audio range. After raw audio is gathered, it is taken through a vital preprocessing phase intended to improve clarity and detection. Throughout this process, several signal improvement techniques are utilized, such as background noise removal, unwanted frequency filtering, and removal of artifacts that could disrupt the recognition process.

Following the preprocessing stage, the system extracts prominent acoustic features utilized in distinguishing distress calls from regular ambient sounds. They include both the spectral and the temporal aspects such as Mel-frequency cepstral coefficients (MFCCs), zero-crossing rates, and spectral centroids, all of which form a general presentation of the features of the sounds. The extraction of the feature is the major input to an already trained artificial neural network (ANN) to classify audio signals as screams or non-screams.

The ANN converts the input features through different layers, with activation functions and transformations to produce the final classification decision. ReLU activation functions and dropout regularization techniques are components of hidden layers, which make the model robust and prevent overfitting. The output layer with a softmax activation function, calculates the probability distribution of classified noise.

In classification, the system makes a well-informed decision on the existence of a scream in the given audio data. In case a scream is identified, the system can trigger pre-set actions, such as alerts to emergency authorities, registration of the event for analysis, or integration into security systems for real-time threat evaluation. The general framework ensures that the system functions effectively and with ease in various environments, positioning it as a useful tool for surveillance, security monitoring, and emergency response systems.

### *B. Weapon Detection*

The provided Python code is an in-real-time weapon detection system built with Flask, utilizing the YOLO deep learning algorithm to detect weapons in a real-time video stream. The system captures video from a webcam, processes each frame with the YOLO (you only look once) technique, and is very accurate in weapon detection. Upon detection of a weapon, the system logs the detection, records the video stream, and sends an alert to the parent application. The YOLO model (50epoch-new-weapon.pt), trained in advance for the detection of various types of weapons, is utilized via the Ultralytics YOLO library.The detection is based on OpenCV (cv2) for continuous video capture and frame-by-frame processing. If a weapon is identified with a confidence value greater than 50%, the detection is logged with a timestamp, video capture starts to capture the video in the ”detected clips” folder, and an alert is triggered in the form of an HTTP POST request to ”http://127.0.0.1:5000/threshold alert”. If the request does not go through, an error is logged. The identified frame is also marked with bounding boxes to identify the detected weapon, and the updated frame is streamed in real time. If no weapon is identified, the system stops recording to save storage space, keeping only relevant video. To support usability, a Flask web application presents an HTML page (weapons.html), which presents controls to initiate and terminate the camera, access the live video feed, and access detection logs. The system offers various REST API endpoints, including video feed to stream the annotated feed, start camera and stop camera to manage the webcam, and get logs to access the last 10 detection logs in JSON format. The Flask application runs on port 5001 in debug mode to support debugging. The system offers an extremely effective security solution by incorporating real-time streaming, automated recording, logging, and alert notification. It is designed to be optimized for public surveillance, law enforcement and crime prevention applications, with rapid response to potential threats through real-time monitoring and automated alerts.

The Weapon Detection dataset, gathered from Roboflow and YOLO (You Only Look Once) fine-tuned, is particularly intended for real-time weapon detection for security and surveillance use. The dataset is made up of labeled images of different weapons such as guns and knives, taken under diverse conditions to enhance the overall generalizability of the models. The dataset is required to train deep learning models to identify and label weapons in real-time video streams accurately to improve automated threat detection systems utilized in law enforcement, public safety, and restricted monitoring areas. Accurate labeling and different image variations provide high detection in real-world environments.

The system starts with video input collection from a dedicated camera that monitors visual activity for potential threats. The captured video is pre-processed to enhance quality in terms of resizing frames, normalization of images, and noise reduction. The system extracts visual features such as shapes of objects, patterns, and motion dynamics from the pre-processed data to aid in detecting weapons. The features are then processed by a machine learning system trained on a wide variety of weapon and non-weapon images, enabling low false alarms during efficient detection. The system then examines the derived data and determines whether a weapon is detected within the video frame.

When a weapon is detected, it triggers an alarm signal to inform security guards or police officers to respond accordingly. The system also captures and processes threats found for forensic investigation later. An automated, real-time detection system like this enhances situational awareness and response time significantly, and hence overall public safety.

**Algorithm 1** Weapon Detection System

```
Input: Video stream (webcam or video file)
Output: Detection of weapons (gun, knife, etc.) with bounding boxes and alerts
BEGIN
Load OpenCV for image processing
Load MediaPipe Holistic model for hand and body keypoints
Load YOLO model with pre-trained weights for object detection
Open video stream (live camera or stored video)
while video stream is active do
    Capture next frame
    Resize frame to $416 \times 416$ for YOLO input
    Normalize pixel values
    Apply MediaPipe Holistic model to detect hand and body keypoints
    Identify hand gestures and store bounding box around detected hands
    Pass frame through YOLO model for object detection
    Extract detected objects and bounding boxes
    for each detected object do
        if detected object class $\in$ {gun, knife, etc.} then
            Store confidence score and bounding box
        end if
    end for
    Apply Non-Maximum Suppression (NMS) to remove duplicate detections
    if weapon detected AND confidence score $>$ threshold (80%) then
        Draw bounding box around detected weapon
        Display label "Weapon Detected"
        Activate security alarm OR send alert notification
        Save detection log and captured frame
    end if
    Show processed video frame with detection overlays
    if user presses 'q' then
        Break loop and stop processing
    end if
end while
Close video stream
Release all loaded models and memory
END
=0
```

The Weapon Detection System Algorithm operates on a video stream (from a webcam or from a file) and detects weapons such as guns and knives. It starts by initializing necessary modules, such as OpenCV for image processing, the MediaPipe Holistic model for hand and body keypoints detection, and a YOLO-based object detection model with pre-trained weights. The system operates continuously, capturing video frames, resizing them to YOLO's input size of 416×416, and normalizing pixel values. The MediaPipe Holistic model detects hand gestures and detects bounding boxes around detected hands, and YOLO operates on the frame to detect objects. If any detected object is from the weapon class (e.g., gun, knife), its confidence score and bounding box are saved. For improving detections, Non-Maximum Suppression (NMS) is performed to remove redundant bounding boxes. If a weapon is detected with a confidence score of over 80, the system marks it with a bounding box, shows a "Weapon Detected" label, triggers an alarm or sends an alert, and logs the detection with a captured frame. The processed video frame, with detection overlays, is shown in real time. The system operates continuously until the user types 'q' to exit, after which all resources and models are released.

### *C. Alert and Location Analysis*

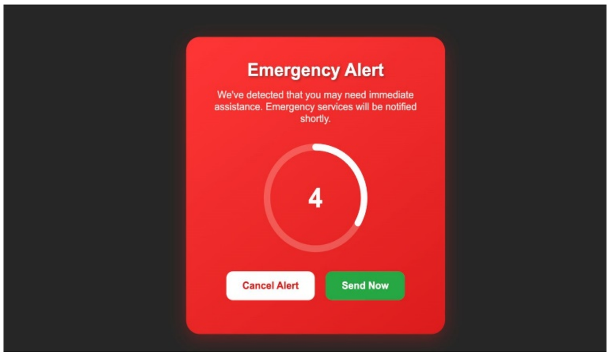
One video input stream is reserved for action recognition and spatial tracking to allow the system to monitor potentially suspicious activity on its own. Utilizing advanced deep learning algorithms and real-time data processing, the system is able to monitor human movement in real-time and detect patterns of aggression, physical confrontations, or other unusual activity. Utilizing MediaPipe-driven action detection, the system monitors body position, limb movement, and velocity to detect threatening behavior, including sudden lunges, raised fists, and frantic arm movements.

Once an abnormal activity is identified, the system not only identifies the case but also delineates high-risk zones, identifying places where such activities are likely to take place. The delineation assists security personnel in focusing their efforts on zones with a past record of suspicious activities, thereby facilitating proactive surveillance instead of merely reactive action. The system can also incorporate contextual elements such as time of day, crowd density, and weather conditions to further improve the precision of its risk assessment.

To further improve detection accuracy, the system incorporates security officer feedback. Officers can review alerted incidents, validate their legitimacy or deny them, and add annotations that allow the model to enhance its decision-making abilities. This human-in-the-loop functionality facilitates continuous learning, and the system improves its ability to learn new suspicious behavior patterns over time. In addition, through the integration of the feedback mechanism, the system minimizes false positives, such that only legitimate concern behaviors trigger security notifications.

The combination of artificial intelligence-based action recognition, geolocation tracking, and real-time threat assessment renders it a valuable tool for applications related to public safety. Implementing it in locations like malls, transportation stations, sports stadiums, or other high-density areas enhances situational awareness, reduces response time, and provides a scalable solution to modern security threats.

## IV. RESULT AND ANALYSIS



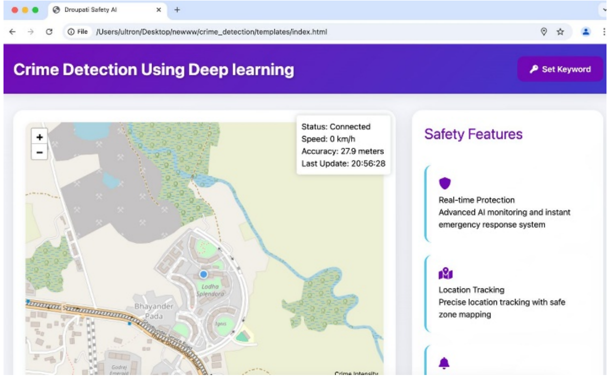
Emergency Alert

We've detected that you may need immediate assistance. Emergency services will be notified shortly.

4

Fig. 3. Emergency Alert

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.



Droupati Safety AI

File /Users/ultron/Desktop/newww/crime_detection/templates/index.html

Crime Detection Using Deep learning

Status: Connected
Speed: 0 km/h
Accuracy: 27.9 meters
Last Update: 20:56:28

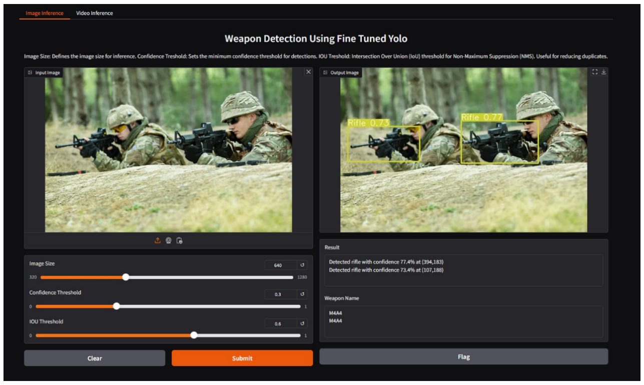
Safety Features

Real-time Protection
Advanced AI monitoring and instant emergency response system

Location Tracking
Precise location tracking with safe zone mapping

Fig. 4. Crime Detection through map

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.



Weapon Detection Using Fine Tuned Yolo

Image Size: Defines the image size for inference. Confidence Treshold: Sets the minimum confidence threshold for detections. IOU Treshold: Intersection Over Union (IoU) threshold for Non-Maximum Suppression (NMS). Useful for reducing duplicates.

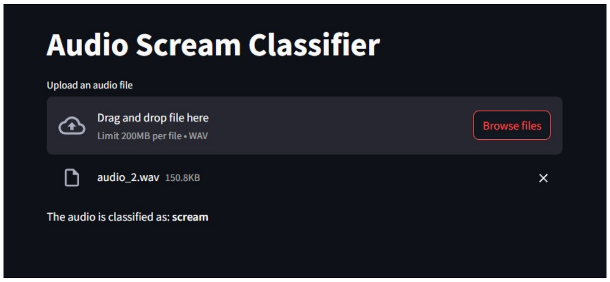
Image Size: 640 (320 – 1280)
Confidence Threshold: 0.3 (0 – 1)
IOU Threshold: 0.6 (0 – 1)

Result
Detected rifle with confidence 77.4% at (394,183)
Detected rifle with confidence 73.4% at (107,188)

Weapon Name
M4A4
M4A4

Fig. 5. Weapon Detection

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.



Audio Scream Classifier

Upload an audio file

Drag and drop file here
Limit 200MB per file • WAV

audio_2.wav 150.8KB

The audio is classified as: **scream**

Fig. 6. Scream Classifier

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.

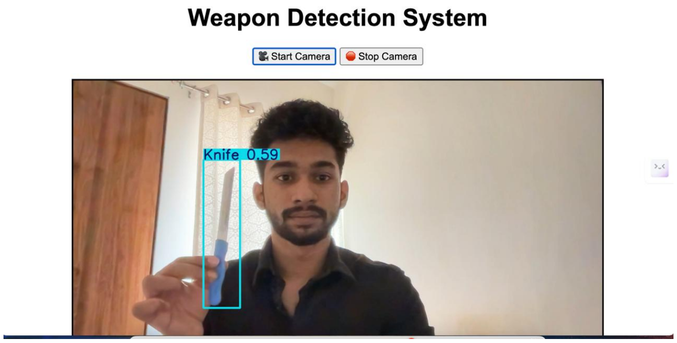


Fig. 7. Weapon Detection

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.

### A. Dataset

#### 1) Scream Detection

For scream identification, we employed an Artificial Neural Network (ANN) model trained on a sound dataset from Kaggle. The dataset includes a vast range of sound recordings, including human screams, normal speech, ambient sounds, and other non-scream sound events. The primary intention of this model is to efficiently distinguish distress sounds, represented by screams, from non-hostile acoustic signals in real-time.

*a) Dataset Details*

- **Source:** Kaggle
- **Audio Types:** Includes human screams, normal conversations, background noises (e.g., traffic, machinery, nature sounds), and other vocalizations.

*b) Preprocessing Techniques*

- **Feature Extraction:** Extracted features include Mel-Frequency Cepstral Coefficients (MFCCs), Spectral Centroids, Zero-Crossing Rates, and Chroma Features to capture frequency domain characteristics.
- **Noise Reduction:** Applied techniques such as spectral subtraction and band-pass filtering to enhance relevant acoustic features.
- **Data Augmentation:** Techniques such as time-stretching, pitch-shifting, and synthetic noise addition were implemented to improve model robustness.

The trained ANN model demonstrated high accuracy in real-time scream classification, making it suitable for emergency detection systems and surveillance applications.

#### 2) Weapon Detection

For weapon detection, we used the YOLO (You Only Look Once) object detection model, which has been trained on the Roboflow annotated dataset. The dataset includes images of handguns and knives under different exposures of light and backgrounds. The goal is to implement real-time weapon detection with high accuracy and low false positives.

*a) Dataset Details*

- **Source:** Roboflow
- **Image Categories:** Includes images of handguns, rifles, knives, and non-weapon objects for contrast.
- **Annotations:** Fully annotated dataset with bounding boxes indicating weapon locations.

*b) Data Preprocessing*

- **Image Resizing:** Resized images to 416×416 pixels to match YOLO's input size.
- **Normalization:** Pixel values were normalized to improve model convergence.
- **Augmentation:** Applied techniques such as flipping, rotation, brightness adjustment, and contrast enhancement to enhance generalization.
- **Balancing the Dataset:** Ensured an equal number of weapon and non-weapon images to mitigate class imbalance.

By fine-tuning YOLO with a diverse dataset, we ensured high accuracy in detecting weapons across various real-world scenarios, making the system highly suitable for security applications such as surveillance, law enforcement, and automated threat detection.

### B. Model Performance Evaluation

To evaluate the effectiveness of our scream detection model, we assessed its performance using key classification metrics: Accuracy, Precision, Recall, and F1-score. The results are visually represented in Figure 8.

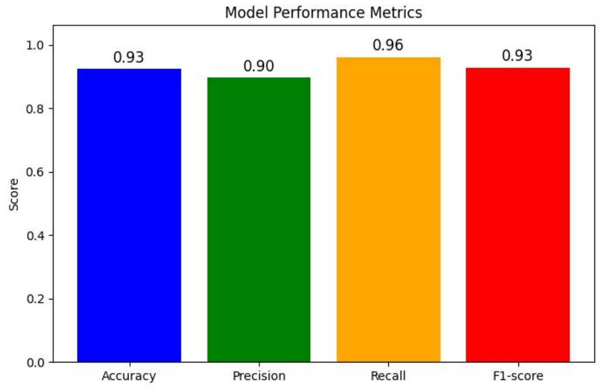


Fig. 8. Overview of Model Analysis

#### 1) Evaluation Metrics

After rigorous training and testing, we achieved the following 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.

### C. Table

TABLE I
MODEL PERFORMANCE METRICS

| Metric | Training | Validation |
|---|---|---|
| **Final Accuracy** | 92.44% | 91% |
| **Final Loss** | 0.2052 | 0.20 |
| **Training Trend** | Increasing | Stable |
| **Loss Trend** | Decreasing | Similar |
| **Overfitting Signs?** | Minimal | None |

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.

In addition, the table evaluates potential Overfitting Signs but can infer little regarding overfitting occurrence since validation accuracy barely varies from the training accuracy. The approximate similarity of values of loss for training and validation also suggests that the model is performing well with unseen data and does not memorize the training set.

### D. Mathematical Formulation

Mel-Frequency Cepstral Coefficients (MFCCs) serve as a powerful feature extraction method in audio processing, capturing characteristics that align closely with human auditory perception. The process involves several key steps:

#### 1) Pre-Emphasis

To enhance high-frequency components and minimize noise interference, a pre-emphasis filter is applied to the raw audio signal:

$$y(n) = x(n) - \alpha x(n-1) \tag{1}$$

where:

- $x(n)$ represents the original audio signal.
- $y(n)$ is the signal after pre-emphasis.
- $\alpha$ is the pre-emphasis coefficient, commonly set to 0.95.

#### 2) Framing and Windowing

The continuous signal is segmented into overlapping frames, typically spanning 25ms with a 10ms overlap. Each frame is then processed using a Hamming window to reduce spectral leakage:[13]

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \tag{2}$$

where $N$ denotes the frame length.

#### 3) Fast Fourier Transform (FFT)

To analyze frequency components, the Discrete Fourier Transform (DFT) is applied to convert the framed signal into the frequency domain:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N} \tag{3}$$

where $X(k)$ represents the resulting magnitude spectrum.

#### 4) Mel Filter Bank Transformation

To align with human hearing perception, the frequency scale is converted from Hertz to the Mel scale using the following equation:

$$m(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \tag{4}$$

A series of triangular filters is then applied to the power spectrum, mimicking the auditory system's response.

#### 5) Logarithm and Discrete Cosine Transform (DCT)

The logarithm of the filtered energy is computed to enhance differentiation between sounds:

$$S_m = \log\left(\sum_{k=f_1}^{f_2} |X(k)|^2 H_m(k)\right) \tag{5}$$

where $H_m(k)$ denotes the triangular filter bank response.

Finally, the Discrete Cosine Transform (DCT) is employed to generate MFCCs by decorrelating the extracted features:

$$C_n = \sum_{m=1}^{M} S_m \cos\left[\frac{\pi n}{M}(m - 0.5)\right] \tag{6}$$

where $C_n$ represents the computed MFCC features.

This transformation plays a crucial role in extracting robust and effective audio features, making it particularly useful for tasks such as scream detection.

## V. Conclusion

This paper introduces an AI-based crime detection system that surmounts the shortcomings of conventional surveillance with the use of real-time audio and video processing. With the application of Artificial Neural Networks (ANN) for scream detection and computer vision technologies like OpenCV, MediaPipe, and YOLO, the system is able to effectively detect distress calls, weapons, and physical fights. Being a multi-modal system, the system improves security by minimizing human error, response time, and enabling proactive threat detection. The system under development is scalable via edge computing and cloud analytics and thus applicable to a wide range of security applications such as smart city surveillance, law enforcement, and emergency response. Future enhancements can include facial recognition, behavior-based anomaly detection, and predictive analytics for further enhanced crime prevention features. Through AI-based automation, the system adds to modern security systems an efficient, scalable, and proactive method of public security. Since AI technology is still evolving, such advanced surveillance systems are expected to have a more central position in safeguarding populations and reducing crime rates.

## VI. Future Scope

AI-powered crime detection holds immense potential for future advancements. Integrating it with security technologies like CCTV, access control, and facial recognition can enhance real-time monitoring and response. Ensuring accuracy in crowded environments is essential, as factors like noise, lighting variations. Advancements in machine learning, hybrid models, and deep learning can help overcome these challenges, improving efficiency and precision. Developing adaptive algorithms for dynamic environments will further strengthen security. A more intelligent and responsive crime detection system can significantly boost public safety and empower law enforcement to prevent and address threats effectively.

## References

[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] A. Sharma, M. Ashikuzzaman, and D. Valles, ”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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] D. Pahuja and S. Jain, “An Automatic Detection of Military Weapons Using Multi-Level CNN Architecture,” in *Proceedings of the 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.

[13] A. Sithara, A. Thomas, and D. Mathew, “Study of MFCC and IHC feature extraction methods with probabilistic acoustic models for speaker biometric applications,” *Procedia Computer Science*, vol. 143, pp. 267–276, 2018, doi: 10.1016/j.procs.2018.10.395.