iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.80624

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Smart CCTV: AI-Based Intelligent Surveillance System Using Deep Learning

SK. Mujafar Ahmed, P. Chandrahas, P. Varun, P. Navadeep, P. Rishi
[1]*Asst. Professor, Department of Artificial Intelligence and Machine Learning (AI&ML), Malla Reddy University, Hyderabad, India*
[2, 3, 4, 5]*UG Scholar, Department of Artificial Intelligence and Machine Learning (AI&ML), MRUH, Hyderabad, India*

Abstract: Smart CCTV: AI-Based Intelligent Surveillance System is an automated security monitoring system that gets around the problems with older surveillance systems. Cameras in the monitored area send real-time video streams to this system. The system uses image processing and machine learning to look at frames and get useful information from them all the time. It can do things like find motion, identify objects, find faces, and find strange behaviour. The system checks the captured frames against trained models and pre-defined patterns to see how dangerous the situation is when it sees movement or suspicious behaviour. The system quickly sends alerts and notifications to the authorized user through a connected app or dashboard if it detects something strange happening. All our detected events and video footages are put into a database which in turn is made available for in depth study at a later time. We also see that this method does away with continuous human supervision which in turn reduces errors. Smart Surveillance System also reports back to us in a timely and precise manner. It is available for use in schools, offices, public places and home settings and we present it as a very reliable, effective, and easy to scale security solution.
Keywords: Artificial Intelligence, Intelligent Surveillance System, Emotion Recognition, YOLOv8 Object Detection, Computer Vision, Motion Detection, Number Plate Recognition, Deep Learning.

I. INTRODUCTION

With the growing need for automated security systems, traditional CCTV surveillance methods fall short because they rely on manual monitoring. Human monitoring is slow, prone to mistakes, and cannot keep up with real-time threat detection.
The Smart CCTV system uses Artificial Intelligence and Deep Learning to automate tasks such as:
• Emotion-based threat detection (Angry/Fear)
• Automatic number plate detection
• Motion-based recording
• Real-time email alert system
• Database logging of security events
The system uses OpenCV, YOLOv8, CNN-based emotion models trained on the FER2013 dataset, and OCR techniques. This creates a complete AI-based surveillance framework.

II. LITERATURE REVIEW

In the field of smart surveillance we have looked at a number of deep learning and computer vision approaches which we have applied to problems. What we did see is that the methods we looked at which which do the processing of different types of input video streams, face images, that which identifies what is moving in the field of view and also what type of vehicles are present are very diverse.

A. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) can be used to analyze the spatial patterns in video and images. CNNs are very efficient at automatically extracting features such as edge, texture and facial characteristics and are therefore highly efficient for tasks such as facial expression recognition and image classification.
Application: CNN-based models are utilized in emotion recognition systems to classify facial expressions into categories of Angry, Happy, Fear, or Neutral. Smart AI CCTV Pro utilizes a pretrained CNN based emotion model to analyze for suspicious emotions (Angry and Fear) found in detected faces to enable identifying threats before they occur.

B. YOLO - Real Time Object Detection

YOLO is a cutting-edge object detection system designed for real-time applications. Unlike previously developed detection models that would require multiple operations to locate an object and determine its type (i.e., to both localize/locate them), YOLO completes this with a single forward pass, allowing it to efficiently perform as an object detector in live surveillance scenarios.

Usage Application:

YOLOv8n (nano) used in "Smart AI CCTV Pro" has been developed using lightweight architecture so that it runs smoothly and quickly on CPU-based systems while delivering very high accuracy. It has been successfully used to detect vehicle license plates as well as for virtually all objects observed in real-time streaming video.

C. Haar Cascade Classifier

The Haar Cascade classifier represents one of the original and widely used algorithms in computer vision that allows for fast detection of people (typically on facial recognition systems), by using an efficient and proven method for detecting objects through the use of a pre-trained cascade classifier - efficiently detecting an object in a given image.

Usage Application:

The Haar Cascade classifier is used in "Smart AI CCTV Pro" to identify (detect) a person's face in order to classify their emotion in real-time, at a very fast pace (suitable for webcam type video surveillance systems), while requiring very little in the way of computing power.

D. Optical Flow/Motion Detection Techniques

Commonly used techniques for detecting movement in video frames today include: frame differencing, background subtraction, and motion thresholding. They have been heavily utilized by video surveillance systems to help detect unusual behaviour without using deep learning algorithms exclusively.

E. Database-Driven Security Systems

Modern intelligent surveillance systems utilize database management to log events, alerts, and user activity. Structured storage enhances traceability and forensic analysis.

Usage Application:

Smart AI CCTV Pro uses an SQLite database to store user credentials, event logs, number plate records, and threat alerts. This provides secure data management and keeps a historical record.

III. LITERATURE SURVEY

M. Salman, Enhancing Surveillance Anomaly Detection Using Keyframes and Explainable Inception Models, Engineering Innovation Journal, 2025.

This paper presents new methods for detecting unusual behaviour in live video feeds from security cameras using deep learning and visual processing techniques such as selecting important frames from the video stream and using the explainable computer vision to make it easier to identify if someone is committing a crime.

Abba, Real-time Object Detection, Tracking, and Monitoring Framework for Surveillance Systems, Heliyon,2024.

A set of algorithms coded in Python that can deteriorate an object (vehicle) and its location in the video using input from multiple screens and/or cameras will be extracted by the methods presented in this paper.

A. Pawar, Intelligence Video Surveillance Using Deep Learning, 2024.

The research also demonstrates an intelligent surveillance system through a combination of background subtraction and various machine learning algorithms, such as convolutional neural networks (CNNs).

R. Zhu, License Plate Detection Based on Improved YOLOv8n Network, Electronics, 2025.

This paper continues to build on existing works by proposing a new improved license plate detection algorithm for the common issue of license plates being mis-read or not being detected due to difficult environmental conditions while still maintaining real-time responsiveness and recognition robustness.

V. Sareen, Video-Based Facial Emotion Recognition using YOLO and Vision Transformer Models, EPJConferences, 2025.

The dataset created as part of this research provides a benchmark for future research on simultaneous face and license plate recognition, and it is designed with real-world challenges in mind and a large amount of data for evaluation purposes.

P. Chaware, Development of Intelligent Video Surveillance System Using Deep Learning, Sensors,2025.
Based on up-to-date information regarding the current challenges and advances in detecting crimes and evaluating the environment using state-of-the-art deep learning methods.
K. Viswanathan et al., FANVID: A Benchmark for Face and License Plate Recognition in Low-Resolution Videos, arXiv, 2025.
The creation of a benchmark data set containing simultaneous detection of both faces and license plates in low resolution video clips from surveillance systems. This includes identification of various obstacles and metrics that can be used to evaluate the performance of camera/video-based systems in real-world environments.
AI Based Smart Surveillance System, ResearchGate
2023
The study of the automatic video and audio analysis of the surveillance system to reduce the amount of human involvement and the assessment of security risks.
AI Driven Smart Surveillance with Motion Detection, IJSRSET, 2025.
A review of the best systems for detecting criminal acts using deep learning and computer vision, including how the systems were developed and trends in the field of intelligent security systems.

IV. PROBLEM STATEMENT

Traditional CCTV systems are simply devices that passively record video and do not have any type of intelligent analysis capabilities. Therefore, these systems have no way to automatically notify security personnel about any suspicious behaviour. As a result, security personnel need to watch over the video footage constantly in order to catch any occurrences of suspicious behaviour. This can delay the response to the occurrence of suspicious behaviour, and also contribute to human error and inefficient detection of threats. In addition, because many of the incidents that occur in the world are at night, and security personnel often work long hours and multitask, they may not always see certain events happen because they are now asleep or distracted by other activities while watching video footage.
In addition, because traditional CCTV systems do not have the ability to automatically analyse behavioural patterns (such as facial emotions), automatically detect abnormal or unusual/erratic movement, automatically identify vehicle license plates, or provide real-time alerts, these systems provide limited support to the detection of threats to the public and private property. While some CCTV systems do integrate multiple sources of data (such as audio and video), they generally do not have the ability to provide a means of storing and logging evidence that has been created by the system. With the continued increase in the demand for proactive security measures in residential, commercial and public spaces, there is an increasing demand for intelligent surveillance systems that are capable of:

- Automatically detecting suspicious emotion/activity
- Identifying motion/abnormal behaviour in real time
- Identifying vehicle license plate for the monitoring purposes
- Recording audio and video at the same time
- Sending automated alerts to a predetermined group of people
- Securely storing the event data so it can be reviewed later

V. METHODOLOGY

The Smart AI CCTV is based on a structured approach incorporating computer vision, deep learning, and sound analysis. In order to provide real-time surveillance at this level, the system has multiple stages: data acquisition; data pre-processing; detection of objects; decision making; and generation of alerts. Each stage will take an input and send the processed input through to the next module in order to effectively monitor threats.

A. Data Acquisition

Video frames from a webcam and sound from a microphone are the two main types of raw data collected by the system through various means in order to allow for real-time monitoring of the environment.

B. Preprocessing

The video frames taken from the camera undergo preprocessing using OpenCV functions; these include resizing, converting to greyscale, applying Gaussian blur, and normalising. By pre-processing video frames, increased accuracy is achieved when

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

detecting objects in a video frame compared to if no pre-processing were done and reduces the amount of noise in a video frame prior to applying any algorithms via machine learning methods.

C. Motion Detection
Motion detection is achieved by using frame differencing and thresholding methods to compare two consecutive frames together to identify significant motion in an area that has been monitored. When a significant amount of motion has been detected, and exceeded the pre-defined threshold, it will flag as an event and begin recording.

D. Face Detection
To detect faces in a video frame the Haar Cascade classifier is applied. Once a face region has been detected, the system captures this region (i.e. face) and sends it to the emotion recognition model for further analysis.

E. Emotion Recognition
Extracted face images are resized to 48 by 48 pixel dimensions in grey scale and given to a Convolutional Neural Network (CNN) model that was trained using several datasets of facial expressions. The model will classify the person's emotional feelings in one of the following seven emotional categories; Angry, Disgusted, Fear, Happy, Sad, Surprised, or Neutral. Any emotion classified as suspicious (i.e. Angry, Fear) is defined as a possible threat.

F. Object Detection
The YOLOv8n object detection model processes the video frames to spot objects such as people and vehicles. YOLO performs both localization and classification in one pass, allowing for quick real-time detection.

G. Number Plate Recognition
Detected vehicle areas are processed to identify number plates. The system extracts the plate region and sends it to Tesseract OCR after preprocessing steps like converting to grayscale and applying thresholding. OCR retrieves alphanumeric characters from the plate image.

H. Audio Monitoring
The system constantly checks audio levels from the microphone. If sound intensity goes beyond a set threshold, the event is flagged as suspicious and may trigger recording

I. Decision and Alert System
The system reviews outputs from all detection modules. If it detects suspicious motion, emotion, or sound, it automatically captures images, records audio and video evidence, and logs the event in the SQLite database.

J. Notification and Storage
If a threat event occurs, an email alert goes to the registered user along with the captured image. All event data, plate numbers, and alerts are stored in the database for future reference.

VI. ARCHITECTURE
The Smart AI CCTV Pro system incorporates a diverse range of advanced technologies including computer vision, deep learning, audio processing and database management in its multi-layered intelligent surveillance system design. Each module operates independently from one another as part of an overall integrated real-time architecture. The input layer's primary function is to continuously receive input data via webcam and microphone so that video and audio can be captured at all times. Motion detection is performed by the processing layer, using Haar Cascade to detect face location, convolutional neural networks (CNN) for emotion recognition and YOLOv8n for object identification. For vehicle monitoring purposes, OCR using Tesseract is also performed on the detected vehicle region in order to determine the license plate number. The various detection modules perform their respective functions continuously together, analyzing both audio and video data streams.

The decision and action layer uses outputs from these various modules to make decisions. For example, if any of the modules detect that someone exhibits suspicious emotion (like anger or fear), or if they detect an appreciable amount of motion (i.e., significant motion), an automated response will be triggered. Automated responses include capturing images, logging events and recording audio and video. Video and Audio will be synchronized using FFmpeg when creating an evidence file within the recording module. All information collected by the system (such as all events that were processed, detail associated with users, vehicle license plate numbers) as well as system alerts will be securely stored in an SQLite Database.

Additionally, an alert system sends real-time email notifications with captured screenshots to authorized users. This layered approach allows for scalability, modularity, and effective real-time threat detection, turning traditional CCTV systems into proactive intelligent surveillance solutions.

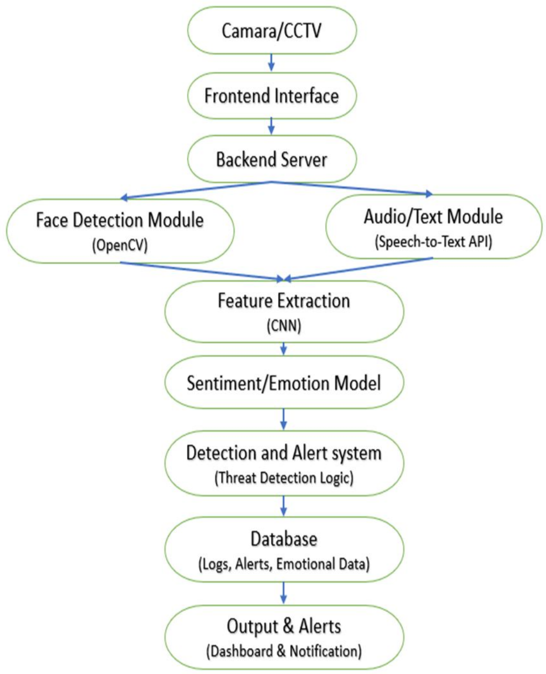

Fig. 6.1. System Architecture of Smart CCTV

VII. MODULES

The Smart AI CCTV Pro System consists of multiple functional modules that separately perform their unique tasks, but also complete all the required work together to create a smart surveillance system using modularity, scalability and efficient real-time processing of video footage.

A. *User Authentication Module*

The user authentication module manages the secure access to the system via login/registration. The user credentials are stored in the SQLite database using the encryption algorithm SHA-256 to hash the password when stored. Only users with verified credentials can log into the surveillance dashboard and/or receive alerts associated with the events.

B. *Video Capture Module*

The video capture module uses OpenCV to connect to the webcam and is used to retrieve continuous frames of live video. The frames will be sent from the video capture module to the various other functional modules for real-time processing (motion detection, face detection, object detection, etc.).

C. *Motion Detection Module*

The motion detection module determines whether the camera feed has detected motion by comparing one frame to the next. Motion detection is accomplished by converting successive camera frames to grayscale, blurring, then threshold through a process of signal processing to identify significant changes before starting the recording of the event and logging it.

D. Module for Face Detection
The process of identifying faces within an image sequence is completed with the Face Detection Module which uses the Haar Cascade Classifier to locate a face in each frame; once detected, this identified face will then be sent to the Emotion Recognition Module to find the type of emotion being represented.

E. Module for Recognising Emotions
Utilising Convolutional Neural Networks (CNNs), the Emotion Recognition Module categorises emotions via Facial Expression Data Sets into 7 categories; Angry, Fear, Happy, Sad, Surprise, Disgust or Neutral. If either Angry or Fear are registered, alerts are activated and recordings made.

F. Module for Object Detection
The Object Detection Module uses YOLOv8n to identify, in real time, the locations and classifications of objects such as People or Vehicles. Consequently, this capability enables YOLO to provide both Object Location and Object Classification simultaneously and, as a result, facilitates an efficient, fast and accurate method for conducting Viable Surveillance.

G. Module for Recognition of Vehicle Registration Plates
The Number Plate Recognition Module can detect vehicles using YOLOv8n and will also identify the area of the vehicle where the registration plate is present. Subsequently, Tesseract OCR will be used on the previously cropped image to obtain details of the registration plate; the identification of a vehicle registration plate will then be logged within a database against the time date stamp of the photograph as well as the photograph itself.

H. Module for Audio Monitoring and Recording
The Audio Monitoring & Recording Module is designed to monitor ambient sound levels via a microphone in real time; should the ambient sound levels exceed a pre-defined threshold then the system is alerted to potential suspicious behaviour. In the event of such an occurrence, audio files shall be recorded and subsequently synchronised to video files by way of FFmpeg.

I. Alert System Module
When suspicious activity is detected, this module sends an automatic email notification to the registered user. The email includes the detected emotion, timestamp, and an attached screenshot as proof.

J. Database Management Module
The database module handles all system records through SQLite. It stores user credentials, event logs, number plate records, and threat alerts. This setup allows structured storage, easy retrieval, and future reference.

VIII. MODEL SELECTION AND ARCHITECTURE
The Smart AI CCTV Pro is designed to use several different ML and CV models to carry out intelligent surveillance such as emotion detection, object detection, and number-plate recognition. The models were chosen based the models' ability to operate in real time, computationally efficient to run, and able to operate effectively in a surveillance environment. Therefore, a CNN model was chosen for the emotion recognition functionality due to CNN's effectiveness in extracting spatial features from images. The model takes a 48x48 pixel grayscale facial image and classifies the image into any one of seven categories: Angry, Disgust, Fear, Happy, Sad, Surprise, or Neutral. As such, the model will assist the system to identify any suspicious emotional state to determine a potential threat.

The YOLOv8n (You Only Look Once), provides very fast object detection and is designed for real-time applications due to its very fast detection speed and lightweight architecture. YOLO, is a one-pass object localization and classification technique, making object detection in continuously captured videos efficient. For detecting faces, Haar Cascade classifiers were used for their fast detection speed and low computational cost. Tesseract Optical Character Recognition (OCR) was used for extracting alphanumeric text from existing number plates. The overall architecture of the system is a layered architecture composed of, an input layer (i.e., camera and microphone), a processing layer (i.e., motion detection, face detection, emotion recognition and object detection) and a decision layer responsible for recording, database logging and alert notification when suspicious activity is detected.

IX. MODEL ACCURACY AND EVALUATION

A data set containing labeled examples was used to test the emotional recognition model. The CNN model's predictions were compared with the actual labels to create a confusion matrix for classification accuracy. The classification of emotions using this system

resulted in an overall accuracy of 75.54%, which is consistent with CNNs trained on FER2013. Emotion classification using CNN is difficult due to the resemblance of certain expressions (e.g., fear vs surprise or sadness vs neutral), one of the reasons for the low accuracy for this application. Thus, this model has the potential for use in real-time surveillance systems but can be improved with larger and more advanced data sets.

A. Using Confusion Matrix

By Using Confusion Matrix, we have predicted the Model accuracy by taking nearly 5000 images from the test datasets of CNN models trained on FER2013 dataset. We calculated each accuracy for each expression separately. Where the Overall Accuracy is 75.54%

Overall Accuracy: 75.54 %
Emotion-wise Accuracy:
Angry: 60.47 %
Fear: 27.68 %
Happy: 93.47 %
Sad: 79.78 %
Neutral: 57.21 %

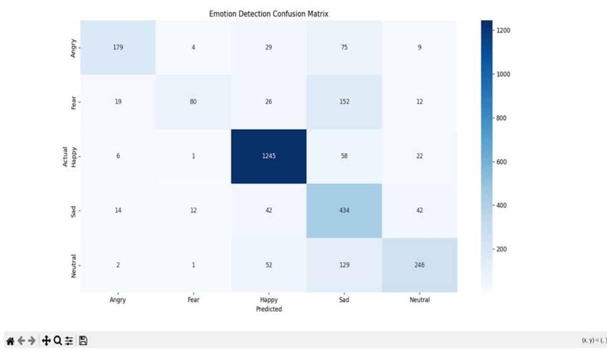

Fig. 9.1 Emotion Detection Confusion Matrix

X. ALGORITHMS

1) *Algorithm 1: Intelligent Artificial Intelligence (AI) Observation System.*
   Input: Live video frames and audio inputs.
   Output: Threat detection, recorded Event log
   (posted to database), email alerts.
   1. Initialize System & Database
   2. Load Models (Haar Cascade, CNN Emotion Model, YOLOv8n).
   3. Activate Webcam + Microphone forMonitoring.
   4. Continuously Capture Video Frames.
   5. Preprocess Frames (Grayscale & Denoising).
   6. Detect Motion by Frame Differencing.
   7. Detect Faces + Classify Emotions Using The CNN.
   8. Detect Vehicles Using YOLOv8n And OCR To Read The Text On The License Plate.

9. Record the event and keep in database, for Any Detected Motion or Emotion which was suspicious.
10. Send Any Evidence Taken Via Email Alert And Continue Monitoring Motion and Emotions.

2) Algorithm 2: Movement Detection.
1. Capture 2 Consecutive Frames
2. Convert the frames to Grayscale
3. Calculate Frame Differences
4. Blur & Threshold Frame
5. Identify Movement By Finding The Contours.

3) Algorithm 3: Emotion Detection
1. Use Haar Cascade To Detect The Face
2. Get The Face Image (48 x 48)
3. Normalize And Input To The CNN
4. Predict Class Of Emotion
5. Indicate Suspicious Emotion

4) Algorithm 4: License Plate Recognition.
1. Use YOLOv8n to Locate the Vehicle
2. Get Area Of Vehicle
3. Preprocess The Image (Grayscale/Filtering)
4. run Tesseract OCR On The Image To Find text on the license plate
5. Store License Plate Number And Time In Database

XI. ADVANTAGES

1) Real-Time Monitoring
2) Automated Threat Detection
3) Efficient Alert System
4) Evidence Recording
5) Database Logging
6) User Authentication
7) Real-Time Performance

XII. APPLICATIONS

1) Home Security Systems
2) Smart City Surveillance
3) Office and Workplace Security
4) Parking Area Monitoring
5) Bank and ATM Security

XIII. FUTURE SCOPE

Improvements can be made to the current system in addition to how the system operates today providing real-time intelligent surveillance and how it can be made more efficient and easier to use. Some examples of new features that can be incorporated into the current system can be, face recognition instead of just face detection, improved emotion detection based upon Vision

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Transformer, and some type of weapon detection to provide better identification of threats. There is the possibility of remote monitoring through cloud storage and/or sending notifications through an application on the phone in real time. Future expansion of the system to support more than one camera, to deploy edge devices (e.g., Raspberry Pi or NVIDIA Jetson), and to integrate with IoT-enabled smart security systems will be part of the overall strategy to develop the system into a fully scalable enterprise intelligent surveillance system by implementing more sophisticated deep learning models and cloud-based infrastructure. In addition, we will implement features like weapon detection, which we believe is invaluable. For example, when a weapon is detected, the system will automatically place a call to emergency services.

XIV. CONCLUSION

The Smart AI CCTV Pro uses a combination of deep learning and computer vision techniques to transform the way traditional video surveillance has been used into a modern, smart, proactive solution for security. By incorporating motion detection, facial recognition, emotion recognition, object recognition with YOLOv8, license plate recognition via OCR and audio surveillance into one system, Smart AI CCTV Pro can provide real-time detection of potential threats and automatically respond to them. If the camera detects a suspicious emotion such as angry or fear, it can automatically record the video, log it in the database, and send an email alert to an authorised person for timely intervention and for the purpose of preserving evidence of the potential crime.

The benefits of combining a number of different AI modules into one architecture are illustrated through the Smart AI CCTV Pro system. By integrating video, audio, database management and alerting systems into a single solution, the Smart AI CCTV Pro system can be scaled and adapted for residential, industrial and commercial uses. This project demonstrates how artificial intelligence can enhance the effectiveness of security systems by reducing the amount of human operations and improving reaction time to completed/attempted criminal acts.

REFERENCES

[1] M.Salman, Enhancing Surveillance Anomaly Detection Using Keyframes and Explainable Inception Models, Engineering Innovation Journal, 2025. https://www.sciencedirect.com/science/article/pii/S1110866525001628

[2] S. Abba, Real-time Object Detection, Tracking, and Monitoring Framework for Surveillance Systems, Heliyon,2024. https://www.sciencedirect.com/science/article/pii/S240584402410953X

[3] A.Pawar, Intelligence Video Surveillance Using Deep Learning,2024. https://journalspub.com/wp-content/uploads/2024/05/9-20-Intelligence-Video-Surveillance-Using-Deep-Learning-1.pdf

[4] R.Zhu, License Plate Detection Based on Improved YOLOv8n Network, Electronics, 2025. https://www.mdpi.com/2079-9292/14/10/2065

[5] V.Sareen, Video-Based Facial Emotion Recognition using YOLO and Vision Transformer Models, EPJConferences,2025. https://www.epj-conferences.org/articles/epjconf/abs/2025/13/epjconf_icetsf2025_01040/epjconf_icetsf2025_01040.html?mb=1&

[6] P.Chaware, Development of Intelligent Video Surveillance System Using Deep Learning, Sensors,2025. https://www.mdpi.com/2673-4591/114/1/11

[7] K.Viswanathan et al., FANVID: A Benchmark for Face and License Plate Recognition in Low-resolution Videos,arXiv,2025. https://arxiv.org/abs/2506.07304

[8] J.Bhavyasri, AI Based Smart Surveillance System, ResearchGate,2023. https://www.researchgate.net/publication/367376895_AI_Based_Smart_Surveillance_System

[9] Ganesh Sonawane, AI Driven Smart Surveillance with Motion Detection, IJSRSET, 2025. https://ijsrset.com/index.php/home/article/view/IJSRSET2512149