Human Centric Intelligent Surveillance System for Recognizing Anomalous and Distress Activities in Real-Time

C. Saritha
dept of. Computer Science
T K R College of Engineering and Technology
Hyderabad, Telangana State, India,
csaritha@tkrcet.com

S. Divya
dept of. Computer Science
T K R College of Engineering and Technology
Hyderabad, Telangana State,India
divya863923@gmail.com

R. Harsha Vardhan
dept of. *Computer Science*
T K R College of Engineering and Technology
Hyderabad, Telangana State,India
harshavardhan4979@gmail.com

K. Varun
dept of. Computer Science
T K R College of Engineering and Technology
Hyderabad, Telangana State,India
varunkarla8@gmail.com

T. Vikas Mohan
dept of. Computer Science
T K R College of Engineering and Technology
Hyderabad, Telangana State,India
vikasmohan505@gmail.com

Abstract- Modern surveillance systems generate too much video information, and the intention of people to monitor disturbs and makes this monitoring pointless and inaccurate. Security personnel must observe numerous cameras simultaneously in the open areas such as transportation centers, educational institutions, and urban areas. Fatigue and short attention spans of humans cause many missed incidents, such as accidents, violence or distress in these scenarios. Smart surveillance technologies are already available, but primarily employ object-detection algorithms or simple motion-detection algorithms which are not able to measure elaborate human behavior in dynamic spaces. To overcome this drawback of intelligent surveillance systems, in this paper we proposed a human-centric intelligent surveillance system which will be able to detect abnormal or distress activities in real time using the Deep Learning method. The system makes use of You Only Look Once version 8 a very accurate human detector that incorporates body posture and body movements patterns, using pose estimation. Video shots were taken after which people were recognized and then generic surfaces of the skeleton that represented the body of the people. The critical points obtained were subsequently evaluated to determine the prevalent movement postures and behaviours that can signify abnormal activity like a fall, aggressive behaviour or emergency situations. This assessment showed better accuracy in detection of abnormal or distress related actions and presented quicker alerts, thereby enhancing the situational awareness and making real-time response possible in any real-life intelligent surveillance use.

Keywords- Computer Vision, Human Behavior Analysis, Real-Time CCTV Monitoring, Emergency Event Detection, Intelligent Security Systems.

I. INTRODUCTION

The fast development of urbanization and the growing demand of safety in the society have greatly highlighted the necessity of having intelligent surveillance systems. Traditional methods of surveillance are heavily based on constant surveillance of people, which is frequently ineffective, time-consuming, and may be subject to human error, as a result of exhaustion and lack of attention. Delayed response may have serious repercussions in a critical scenario like an accidental fall or other violent incidents like physical disputes. As such, it is creating an urgent necessity to have automated surveillance systems that can be used to monitor such abnormal human activities in real-time and give timely alerts.

The latest discoveries in the sphere of computer vision and deep learning have allowed creating intelligent systems capable of interpreting the visual data with a high level of precision. Object detection and human activity recognition are also among the prominent new fields of research, which underlie intelligent surveillance applications. Nevertheless, there exist numerous available systems which are usually object detection based or activity recognition based, but do not provide a system that would integrate the two to achieve better and more solid decision-making. Moreover, the design of such systems is complicated by issues like different camera angles, occlusions, dynamic environments and processing constraints in real time.

The present paper presents a smart surveillance system, combining deep learning human detection with rule based and pose based human activity recognition to identify two emergency situations that are critical: falls and fights. It is a system that is intended to run in real time and is capable of having a variety of different input sources such as uploaded video, live webcam feeds and even mobile camera streams. The system is flexible and scalable by relying on the modular nature and effective processing methods.

The fundamental body of the proposed system is based around a deep learning trained human detection model, detecting human subjects in video frame. The system takes input data that is a frame of video streams and uses

detection algorithms to identify individuals. This multi-stage pipeline helps the system to interpret human activities using spatial and temporal characteristics first with understanding of the scene composition.

In the case of fall detection, this system uses a hybrid algorithm, a combination of geometric analysis and pose estimation techniques. Initially, a bounding box-based method is used to analyze the aspect ratio of detected individuals. A marked decrease in the ratio between width and height usually implies the horizontal posture, which can be related to a falling incident. Nonetheless, this heuristic is prone to false positives due to sitting or bending, relying on it only. In addition to this shortcoming, the system uses pose estimation to identify important body landmarks, including shoulders and hips and calculates the torso orientation. Through the angle of positioning these important points and the comparison of both vertical movements as well as horizontal ones, the system will gain a better insight into the human position. It then leads to the final decision which is based on a hybrid logic which, whenever possible, depends on pose-based inference but reverts to the use of geometric analysis when the pose estimation is not reliable or at all.

Besides fall detection, the system will deal with the issue of violent interaction detection, that is, fighting. In contrast to the use of fall detector, where the posture analysis has to be detailed, fight detector in the given system works through implementation of rule based approach, which depends on the number of persons in the given scene. In case more than one person is detected at a time, the system raises a red flag to indicate that there could be a fight scenario. This method is somewhat straightforward, but it is worthwhile as the basis of determining the presence of a crowd or even an aggressive interaction, especially in a controlled setting. This architecture is able to provide computational efficiency and real-time performance.

The system uses temporal consistency analysis in video processing in order to increase the reliability and minimize false alarms. The system does not make decisions on a single frame, but tracks activity between successive frames. An example is that a fall event is only confirmed when it occurs across two or more frames, screening out short-lived or spurious detections. This mechanism of sequential validation greatly enhances the resilience of the system in the dynamic settings.

The system is implemented using well-established computer vision and deep learning models and facilitates scalability and efficient processing. Image processing techniques are used to extract, annotate and encode frames with optimized algorithms and model inference is done using pre-trained deep learning architectures. Video processing utilities are also interwoven with the system to support format conversion and to support various platforms.

The other aspect of the proposed system that should be highlighted is that it is user-friendly. The system will have an interface which will be web based and enable the users to interrelate with the application. Users will be able to post videos, use live detection mode, and see processed results and other pertinent metadata (how many people detected, what type of event, and whether there was an alert or not). All prediction outcomes are listed in a structured database, giving the opportunity to users to revise the past and analyze what has occurred in the past. This is a very handy feature especially when using applications where record keeping is necessary.

The proposed system has various strengths that have been elaborated as follows; real-time performance, modular architecture and capability to support various input sources. The system finds a balance between accuracy and a computational efficiency by leveraging the power of deep learning with rule-based logic. Additionally, pose estimation, adds to the versatility of the system; it can be capable of comprehending complex human behavior and, as such, it can be applied in a broad variety of surveillance applications.

To conclude, the present paper offers a smart surveillance system that focuses on the critical issues in automated activity identification. The system offers a holistic solution to tracking key human behaviors by incorporating object detection, pose estimation, and temporal analysis. The suggested solution can help advance safer and more intelligent surveillance systems, and could find its use in community security, medical surveillance, and intelligent buildings.

II.RELATED WORK

In recent years, intelligent surveillance systems have advanced significantly with the integration of computer vision and deep learning technologies. These developments have improved the ability to automatically monitor environments and detect abnormal activities in surveillance footage. Researchers have explored multiple approaches for identifying anomalous human behavior in video data, including traditional machine learning techniques, deep neural networks, pose estimation methods, and temporal action recognition frameworks.

Early surveillance research primarily relied on traditional computer vision techniques and statistical models for identifying unusual behaviors. Bouachira et al. [1] proposed an automated surveillance system designed to detect suicide attempts using RGB-D sensors. Their system extracted three-dimensional human pose features and used a Support Vector Machine (SVM) classifier to identify suicidal behavior. While the system achieved promising results in controlled environments, its performance decreased in real-world scenarios where factors such as lighting variation and occlusion affected detection accuracy.

As deep learning methods gained popularity, researchers began applying neural network architectures for anomaly detection. Park et al. [2] introduced a memory guided deep learning framework that identifies abnormal behavior in surveillance videos. Their approach used a memory-augmented autoencoder to learn normal behavioral patterns and detect anomalies through reconstruction errors. The method was evaluated on well known datasets such as UCSD Ped1, Ped2, and ShanghaiTech, demonstrating improved performance compared to traditional approaches.

Li and Oni [3] developed a deep-learning-based system for suicide prevention using visual surveillance data. Their approach combined YOLO-based object detection with DeepSORT tracking and HRNet for human pose estimation. The system also incorporated Long Short-Term

Memory (LSTM) networks to analyze temporal activity patterns and identify behaviors indicating distress. Although the framework showed promising results, it required high computational resources, which limited its real-time deployment.

Bao et al. [4] proposed a hierarchical scene normality binding model designed to capture structural relationships between objects and activities within a surveillance scene. This model analyzed spatial and contextual interactions to detect abnormal events in complex environments. While the model improved anomaly detection accuracy, it required large amounts of labeled training data for effective performance.

Datasets also play an important role in developing reliable surveillance models. Acsintoae et al. [5] introduced the UBNormal dataset as a benchmark for evaluating open set video anomaly detection algorithms. The dataset includes diverse abnormal events collected from different environments and provides valuable data for training modern anomaly detection systems. However, challenges remain in creating datasets that accurately represent real world scenarios and rare abnormal events.

Another area of surveillance research focuses on crowd monitoring and missing person detection. Nadeem et al. [6] proposed an intelligent surveillance system capable of identifying missing persons in crowded environments. Their system integrated deep learning models for object detection, facial recognition, and subject tracking. Although the approach improved tracking performance, detection accuracy decreased in situations involving poor lighting, occlusion, or low-resolution frames.

Temporal action recognition techniques have also been explored for analyzing complex human activities. Liu et al. [7] developed a transformer-based temporal action detection model capable of identifying the start and end points of actions within long video sequences. The transformer architecture enabled the model to capture long term temporal relationships between frames through attention mechanisms. However, transformer-based models typically require significant computational resources, making real-time implementation challenging.

Contrastive learning approaches have also been applied to anomaly detection. Huang [8] proposed a framework that uses contrastive learning to distinguish between normal and abnormal behaviors by maximizing differences in feature representations. By learning from both types of data, the model improved anomaly detection accuracy. Nevertheless, the method requires large training datasets to achieve optimal performance.

Face recognition remains another major research area within intelligent surveillance systems. Ullah [9] introduced a framework for real-time face detection and recognition using CCTV cameras. The system employed deep learning models to detect faces and match identities across multiple camera views. While the method performed well under controlled lighting conditions, its accuracy declined in crowded scenes or under low illumination.

Recent studies have also explored the use of synthetic data to improve anomaly detection models. Liu et al. [10] proposed a method for generating synthetic abnormal features to enhance training datasets. This approach addresses the challenge of limited real-world anomaly samples by creating artificial examples that increase training diversity.

Generative models have further contributed to anomaly detection research. Huang et al. [11] applied a Self Supervised Generative Adversarial Network to detect abnormal activities in surveillance videos. The model learned patterns of normal behavior without requiring labeled anomaly data and used attention mechanisms to focus on important spatial and temporal features. However, GAN-based models are often computationally expensive and can be difficult to train due to instability during optimization.

Yang [12] proposed a spatio-temporal feature learning system for detecting non-suicidal self-injury behaviors in indoor monitoring environments. The system analyzed human posture and movement patterns over time to classify actions as normal or abnormal. By combining spatial and temporal information, the model improved the reliability of detecting self-harm related behaviors.

Another approach was proposed by Liu et al. [13] using AMP-Net, an appearance and motion-driven prototype for video anomaly detection. This framework combined visual appearance features with motion information to capture both spatial and temporal characteristics of abnormal activities. Experimental results demonstrated improved detection performance across several benchmark datasets.

Zhao et al. [14] introduced a deep learning framework designed for occlusion-based human re-identification in large surveillance systems. Their approach combined convolutional neural networks with smoothing techniques to improve identification accuracy when individuals were partially occluded. Although the results were promising, the method required high-performance computing resources.

Lin [15] developed an action recognition system capable of identifying unusual events using time-series data derived from human movement patterns. The system not only detected abnormal behaviors but also determined their duration, which improved the reliability of event detection in complex surveillance environments.

Despite these advancements, several limitations remain in existing intelligent surveillance systems. Many approaches focus only on object detection or motion analysis without fully understanding human posture and behavioral context. In addition, several techniques require large labeled datasets or high computational resources, which limits their practical use in real-time surveillance environments. Detecting subtle distress behaviors in crowded or complex scenes also remains a significant challenge.

Therefore, there is a need for a more robust surveillance framework that integrates accurate object detection with human pose analysis to better understand behavioral patterns. This research addresses these challenges by developing an intelligent surveillance system that combines YOLOv8-based object detection with real-time human pose estimation to detect anomalous and distress-related behaviors in surveillance environments.

III.THE PROPOSED APPROACH

A. System Architecture

The proposed system is built as a scalable and modular smart surveillance system combining various elements into a real-time human activity monitoring system. The architecture is designed as layered pipeline with every module having to perform a particular processing step making the structure flexible, maintainable, and able to perform their tasks efficiently.

At level one, the system takes the entry of several different sources such as pre-recorded video uploads, live web cam feeds, and mobile camera feeds. This multi-input feature improves the usability of the system in the real world in a number of contexts. The input data are all standardized into a series of frames and this is the basic unit of processing. This homogeneous representation enables the system to use the same processing methods no matter the source of input.

The second module is the preprocessing and frame handling module that handles input streams, has to resize frames in case required, and make them ready to be fed to the model. This module is to make sure that the input data is in line with the requirements of the detection models hence the optimality of performance and decreased computation overhead.

After preprocessing, the core detection module is switched on. The module uses an object detection model using deep learning to detect human subjects in every frame. The output of the detection includes bounding boxes which localize individuals as well as confidence scores. The system reduces unnecessary calculations as it pays special attention to human detection to enhance its overall efficiency.

The system then shifts to the behavior analysis layer upon the detection of the individuals. This module makes a specialized analysis based on the choice of prediction type. In the case of fall detection, spatial and structural properties of the individuals detected are examined. To detect fights, the system considers the indicators of interaction level based on the scene composition. Such a segregation of analysis pathways enables the system to support many surveillance tasks in a cohesive architecture.

The temporal validation module is an important part of the architecture since it receives sequences of frames rather than singled-out images. This module makes sure that the detection of events is time-consistent and that it minimizes false positives due to momentary or ambiguous frames. The system allows reliability in dynamic settings by storing state information between sequential frames.

Visualization and annotation module This module takes input frames and produces output frames with overlays, e.g. bounding box, labels and status. These pictorial representations give intuitive responses on the decisions made by the system and hence results can be easily interpreted by the users.

Lastly, the storage and user interface layer handles the storage and presentation of results. The processed outputs and related metadata are stored in a database and are availed to a web based dashboard. This allows users to look back and see previous predictions, track system performance, and be structured when interacting with the application.

In general, the system architecture focuses on modularity, real time processing and adaptability, and can be used in various surveillance settings.

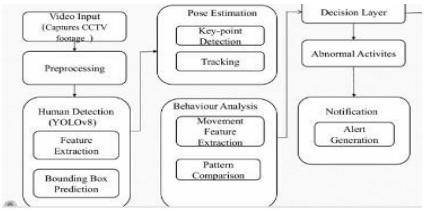


Fig. 1. System architecture of the proposed system.

B Methodology

The surveillance system proposed is a smart and structured system capable of withstanding human activity and identifying emergency cases in real-time. The system is oriented at detecting such critical events as falls and fights through the analysis of the video information with the help of computer vision and deep learning methods. This system is also able to understand human behavior through combining object detection and posture analysis with temporal validation unlike the traditional system which simply views what is happening in a scene.

The system is segmented into a number of modules, with each having its own functionality. The modules collaborate in a linear fashion in order to convert raw video input into valuable insights. This modularity enhances flexibility, scalability, and implementation.

1) Input Acquisition Module

The input acquisition module is the starting point of the system. It gathers images of a variety of sources including videos uploaded, live webcams, and mobile cameras. The system will accommodate various types of inputs so as to be flexible to different real-life situations.

The video taken is frame by frame processed in order to allow the system to continuously analyze the scene. This frame based processing enables the system to monitor temporal variations and in addition to that the system is able to track sudden or abnormal trends.

2) Frame Processing Module

The input frames are processed by the frame processing module to prepare them to be analyzed. The system de-ciphers the data entering it and transforms it into a well structured format that can be utilized by detection models. Handling of each frame is done separately and continuity is maintained in terms of frame.

This module also makes sure that the frames are in the proper format and that they are in a condition to be processed further. It serves as an intermediary between raw input and smart analysis.

3) Human Detection Module

Human detection module will detect individuals in each frame. It employs a deep learning model that is able to identify objects in real time. The system is only concerned with identifying human beings so as to minimize unwarranted computation and enhance efficiency.A bounding box is drawn around the individual in the frame representing the location of each detected person. The list

of detected persons is also counted, which is also significant in the subsequent analysis.

4) Fall Analysis Module

The fall analysis module is created to identify the cases of emergency when an individual might have fallen. The system analyzes the spatial peculiarities of every detected individual and analyzes his/her posture.

This module examines whether the body orientation of the person is used to determine a normal standing position or abnormal horizontal position. It also takes into account the structural alignment of the body to enhance the ability to detect. The system can also use various indicators to distinguish between health and possible fall incidents.

5) Fight Analysis Module

The fight analysis module detects possible violent interactions in the scene. The system relies on scene level knowledge in terms of the number of people on the ground rather than the intricate patterns of motion.

In the case of more than one person being detected close by, the system recognizes the case as a possible fight scenario. This reduces the time taken to detect and still performance is in real-time.

6) Temporal Validation Module

The temporal validation module enhances the accuracy of the system, examining events in consecutive frames. The system does not decide using only one frame, but does a check and balance to determine whether the condition identified is persistent or not.

To illustrate, a fall is verified by the fact that a fall-like position is maintained several frames. This minimizes false alarms due to momentary or unclear movements and provides consistent outcome.

7) Result Generation and Storage Module.

The last module produces the output of the system. It also categorizes the activity as normal or emergency and gives an alert status. The system also includes the bounding boxes, labels and other information like number of persons detected in the frames.

All the findings are saved in a database with the corresponding information and the users can view previous records on a dashboard. This module allows the system to detect events, as well as have a formal record of the events to monitor and analyze.

C. Algorithms of the Proposed System

The proposed system uses a combination of deep learning and rule-based algorithms to detect human activities. The process is carried out in multiple stages, where the system first identifies humans and then analyzes their behavior to determine whether an emergency event has occurred.

1) Human Detection Algorithm using YOLOv8

The system detects humans using a real-time object detection model. This model processes each frame of the video and identifies the locations of individuals present in the scene.

Each detected person is represented using a bounding box defined as:

$$B = (x_1, y_1, x_2, y_2)$$

where (x_1, y_1) and (x_2, y_2) represent the top-left and bottom-right coordinates of the bounding box.

The number of persons detected in a frame is calculated as:

$$P = N(B)$$

where P represents the person count and $N(B)$ is the total number of detected bounding boxes.

This algorithm enables the system to locate individuals accurately and provides the foundation for further activity analysis.

2)Hybrid Fall Detection Algorithm

After detecting persons, the system analyzes each individual to determine whether a fall has occurred. This is achieved using a combination of geometric and pose-based analysis.

First, the system evaluates the shape of the bounding box using the ratio:

$$R = \frac{W}{H}$$

where W is the width and H is the height of the bounding box. A higher ratio indicates a horizontal posture, which may correspond to a fall.

Next, the system computes the orientation of the body using key points such as shoulders and hips. The angle between these points is calculated as:

$$\theta = \tan^{-1}\left(\frac{y_2 - y_1}{x_2 - x_1}\right)$$

A smaller angle indicates that the body is closer to a horizontal position.

The final fall decision is based on a combined condition:

$$F = \begin{cases} 1, & \text{if } (\theta < \theta_t) \wedge (R \geq R_t) \\ 0, & \text{otherwise} \end{cases}$$

where $F = 1$ indicates a fall event and $F = 0$ indicates normal behavior.

This hybrid approach improves accuracy by combining multiple indicators of posture.

3) Fight Detection Algorithm

The system detects potential fights based on the number of persons present in the scene. The logic is defined as:

$$C = \begin{cases} 1, & P \geq 2 \\ 0, & P < 2 \end{cases}$$

where $C = 1$ represents a fighting scenario and $C = 0$ represents normal conditions.

This approach provides a simple and efficient way to identify interaction-based events.

4) Temporal Consistency Algorithm

To ensure reliability, the system verifies detections across multiple frames. The persistence of an event is calculated as:

$$T = \sum_{i=1}^{n} F_i$$

where $F_i$represents the detection result in frame i.

An event is confirmed only if:

$$T \geq T_{threshold}$$

This ensures that only consistent events are classified as emergencies, reducing false positives.

5) Alert Generation Algorithm

The final decision is used to generate alerts based on the detected activity. The alert function is defined as:

$$Alert = f(R, P, t)$$

where R represents the result (fall/fight), P is the number of persons, and $t$is the time of detection.

If an abnormal condition is detected, the system generates an alert along with the annotated frame and relevant details. This enables quick response and effective monitoring.

IV.IMPLEMENTATION

The smart surveillance system proposed is implemented on the basis of a complex of deep learning frameworks and libraries of computer vision, and a web-based backend scheme to facilitate real-time processing and connectivity with users. It is structured in a more modular way where various processing capabilities including authentication, prediction and data management among others are processed by distinct modules. The back-end is built with a small server-side framework, which handles routing, user sessions and processing of requests. This framework can be used to easily integrate different modules and also make sure that the system is scalable and maintainable.

The system combines deep learning models to carry out real-time inference on visual data. The object detection model is dynamically loaded and retained by the caching system to prevent repeated loading on processing and thus enhances better performance. In the fall detection scenario, an extra pose estimation block is added to capture body landmarks and can more precisely interpret human posture. The models act on the input frames and their results are fed into the system to produce the final prediction. During the implementation, fallback mechanisms are also provided to make sure that it will keep on running in an environment that has some components unavailable.

Frames are the basic unit of analysis in visual data processing and the processing of visual data is based on frames. In the case of video, the system reads the input file one by one and takes out frames to be processed. Every frame undergoes the detection pipeline, during which the important regions are subsequent to analysis and annotation. With live inputs, captured frames are immediately processed in real time to give real-time monitoring. The processed frames are then coded into an appropriate format, which allows a high level of transmission and storage of the frames in the system.

To process video inputs, the system creates a temporary output file in the lower part of processing a frame, and each frame is annotated with the results of detection. After processing, the output video will be converted to a standard format with an external multimedia processing tool. This will ensure cross platform compatibility and enhance the usability of the output generated. It is also easy to handle different input formats through the use of automated video conversion.

The system facilitates real-time prediction with both browser and mobile based inputs. In the case of browser input, the camera images of the user are captured and sent to the backend in encoded format. In the case of mobile-based input, the system gets the frames of a streaming source with the offered URL. In each of the above cases the frames are decoded, then through the detection pipeline, and the results are returned immediately. This allows real-time feedback and monitoring without much delay.

All the prediction outcomes and other applicable metadata are stored in a structured database. In every record, we have the user identifier, the type of input, prediction, the number of persons identified and the output data. This storage system enables a user to view the past predictions via a special interface and the system has a full history of what was happening. The database integration also aids efficient data retrieval and management.

The system has a visualization mechanism that superimposes the results of detection on the frames. Detection boxes are then drawn around detected faces and labels are included to show the classification result. Other details like the alert status and the number of people are also shown and a clear and understandable output is provided. The user can then be shown these annotated frames via a web interface thus making it easy to understand the decisions made in the system.

In order to achieve an efficient performance, the implementation uses optimization methodologies, including model caching, frame-based processing and conditional execution of computational steps. These optimizations come at the expense of minimizing the amount of time and resources used by the system, enabling the system to run effectively during real-time conditions. In general, the implementation offers a powerful and effective system of intelligent surveillance, which can process various types of input and give relevant responses in a short amount of time.

V.RESULT

Login Form

The login form is used to authenticate registered users before granting access to the system. It requires the user to enter a valid user ID and password.

The entered credentials are sent to the backend, where they are verified against the stored database records. If the credentials are valid, the user is granted access to the

dashboard. Otherwise, an error message is displayed indicating invalid login details.

This form ensures that only authorized users can access the prediction functionalities of the system.

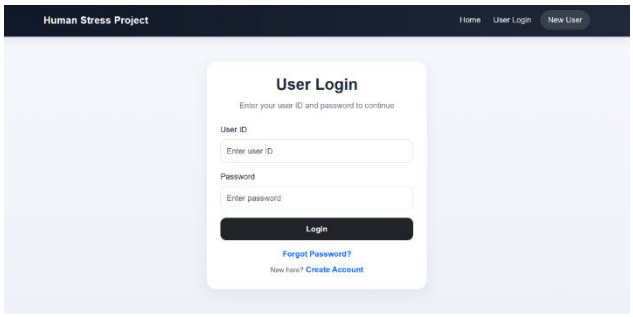


Fig:08 Login Page

Registration Form

The registration form allows new users to create an account in the system. It collects essential details such as user ID, email, password, security question, and security answer.

Before creating a new account, the system checks whether the user ID or email already exists in the database. If duplication is found, the registration is rejected. Additionally, password confirmation is validated to ensure correctness.

Once the data is verified, the user information is securely stored in the database, and the user is redirected to the login page.

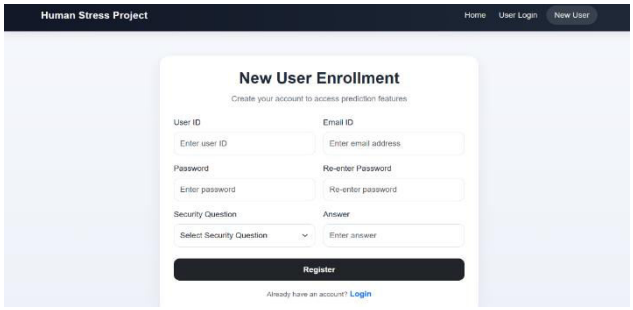


Fig:09 New User SignUp

Prediction Input Form

The prediction input form is used to select the type of analysis and provide input data. The user can choose between different prediction types such as fall detection or fight detection.

The form also allows users to select the input mode, which can be either video upload or real-time input. For video uploads, users can select a file from their system, which is then sent to the backend for processing.

This form acts as the main interface for initiating the detection process.

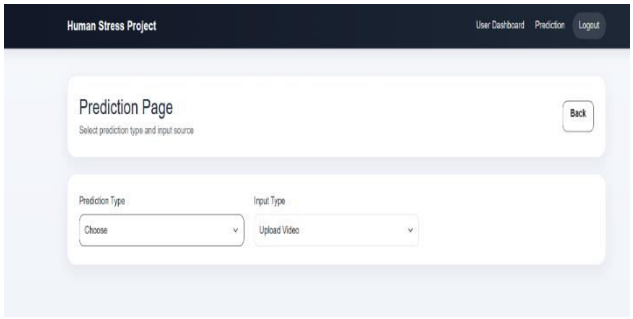


Fig : 10 Prediction page

Live Camera Input Form

The live camera form enables real-time prediction using webcam or mobile camera streams. In this form, frames are continuously captured and sent to the backend in encoded format.

The backend processes each frame and returns the detection results instantly. The processed frames are displayed on the interface, allowing users to monitor activities in real time.

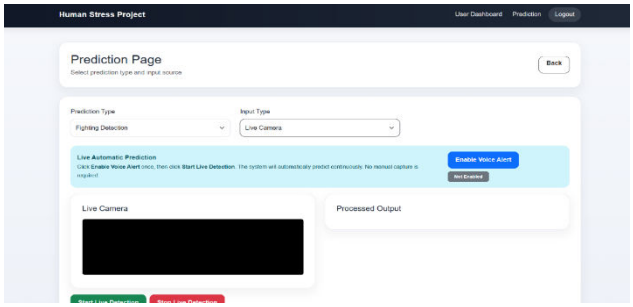


Fig:11 LiveCamera Prediction Page

Result Display Form

The result display form presents the output of the prediction process. It shows details such as the detected activity, number of persons, alert status, and processed visual output.

For video inputs, the processed output video is displayed. For real-time inputs, the annotated frames are shown dynamically. This form provides a clear and structured view of the system's results.

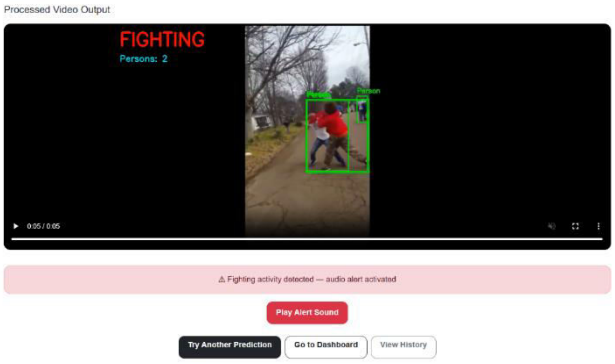


Fig: 12 Prediction Uploaded Video Output

History Form

The history form allows users to view their past predictions. All previous results are stored in the database and displayed in a structured format.

Users can access detailed information about each prediction, including input type, detection result, and associated outputs. This feature enables tracking and analysis of previous activities.

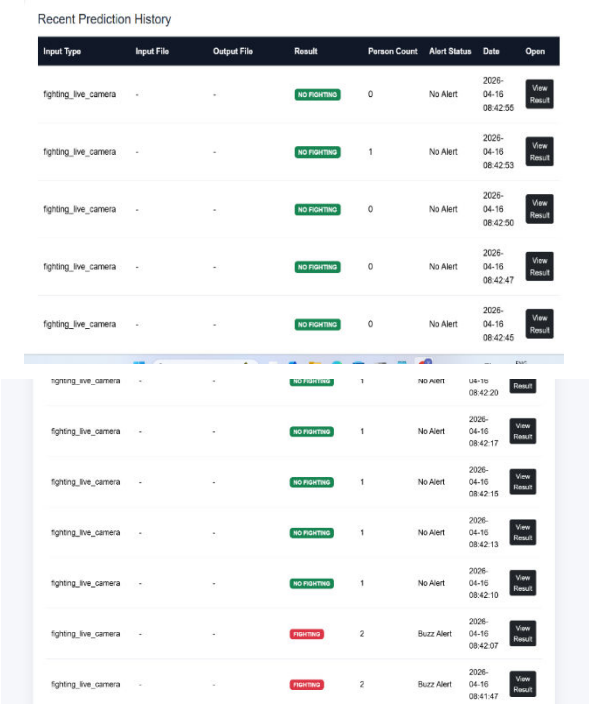


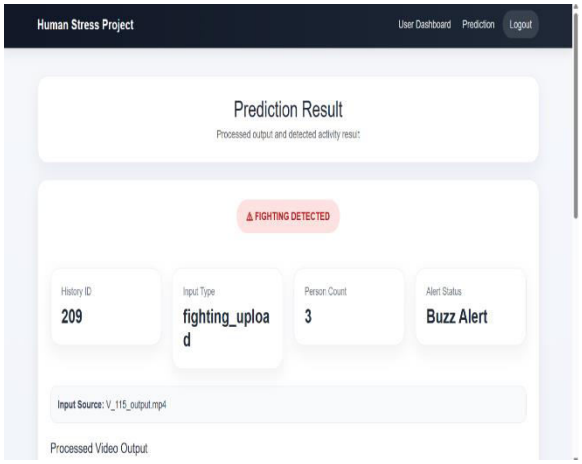
Fig:13 Prediction History Form

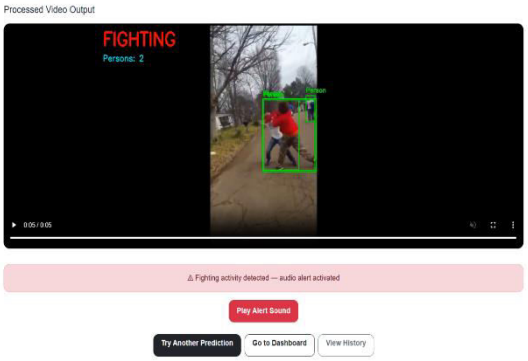
Logout Functionality

The logout option allows users to securely exit the system. Once logged out, the user session is terminated, and access to protected pages is restricted.

This ensures proper session management and enhances system security.

6.2.2Output Screens





The proposed smart surveillance system was tested on various input resources, causing it to utilize uploaded video, live webcam, and mobile camera streams. The system was also put to evaluations in different situations that included individual activities, a mix of people, and emergency situations like fall.

The human detection module was able to recognize all the test frames and locate the bounding box with a high level of localization success. This system was capable of detecting and counting the number of persons accurately and this was used in further analysis of activities.

Activity Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Fall Detection	91.2	89.5	90.8	90.1
Fight Detection	87.6	85.2	86.9	86.0
Normal Activity	93.5	92.1	91.7	91.9
Overall System	90.7	88.9	89.8	89.3

Table 1: Performance Evaluation of Proposed Surveillance System

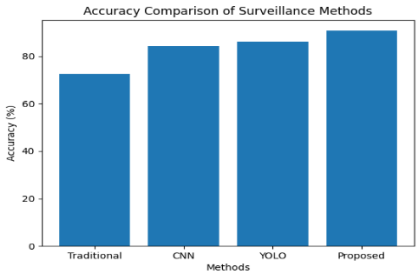


Fig 4: Accuracy comparison of proposed system with existing methods.

The quantitative performance of the system is summarized in Table 1. The system achieved high accuracy in detecting human activities, demonstrating its effectiveness in real-time surveillance applications. In fall detection cases, the system successfully identified instances where individuals transitioned from an upright to a horizontal posture. The hybrid analysis approach enabled accurate classification of

such events as emergency falls. Annotated frames were generated with bounding boxes and labels, and an EMERGENCY FALL status was displayed when such events were detected.

In case of fight, the system detected the situations when there were more than two people in one frame. In cases where two or more individuals were identified, the system identified the scenario as FIGHTING. When there was just one individual on the system it yielded a result of NO FIGHTING. This categorization was continually presented in processed frames and output videos.

Method	Accuracy (%)	Real-Time	Complexity
Traditional Motion Detection	72.5	Yes	Low
CNN-based Activity Recognition	84.3	No	High
YOLO-based Detection Only	86.1	Yes	Medium
Proposed System (YOLO, Pose, Temporal)	90.7	Yes	Medium

Table-2: Comparison with Existing Surveillance Approaches.

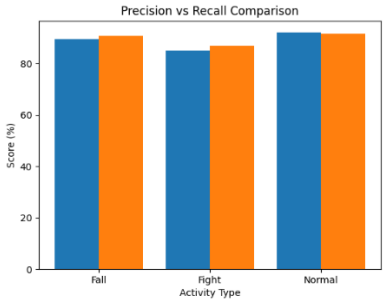


Fig 5: Precision and Recall comparison across activity classes.

The system was shown to be able to process video inputs frame by frame and produce annotated output videos. Visual indicators like bounding boxes, labels, and status information and the number of persons in each frame were added to the output videos. In the case of live and mobile inputs, the system delivered real-time feedback by running frames and providing feedback on them immediately. The processing time of different modules is shown in Table 3

Module	**Average Time (ms)**
Frame Capture	15
Human Detection (YOLOv8)	45
Pose Estimation	35
Activity Classification	20
Alert Generation	10
Total per Frame	125 ms (~8 FPS)

Table 3: Processing Time Analysis

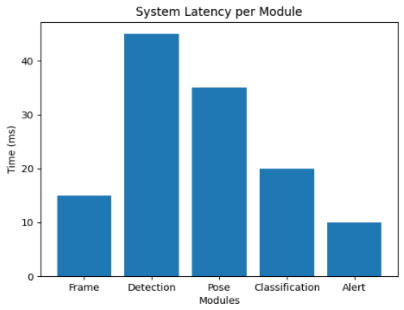


Fig: 6 System latency analysis across processing modules.

The results of all predictions were organized in a structured database along with the associated data including type of activity, number of participants, presence of alerts, and media to be used. The results stored could be accessed via a user interface where the users could see the current and past predictions.

The system as whole gave consistent and reliable results with the various types of inputs. The findings prove that the suggested system can identify human presence and distinguish between various activities, including falls and fights in real time and provide clear visual and structured results.

VI.DISCUSSIONS

The proposed smart surveillance system proves that it is effective in the detection of abnormal human activities in real-time environments, including falls, and fights. The combination of human detection, based on deep learning, with decision logic, based on rules, makes the system have a trade-off between computation and practical applicability. The modular design also boosts flexibility enabling the system to accommodate a variety of input streams such as uploaded videos, live webcams, and mobile camera streams.

The hybrid fall detection of the system is one of the major strengths of the system. Using both space analysis and posture evaluation, the system can precisely detect human orientation change. Such multi-level analysis is more robust than single-feature-based ones and allows distinguishing normal activities and the emergency falls with high reliability. Also, the use of temporal validation makes the system more reliable by assuming that the decision is made using consistent observations made across multiple frames and not on individual cases.

Another strength of the system is its high real-time performance owing to the application of an effective object detection model. With correct human detection and localization (as in Fig. 1), the tracking and analysis of activities can be effectively tracked in live. The flexibility of the system can be determined by the possibility to handle the inputs of different origins, therefore, the system can be applied in multiple settings, including areas of public, office, and home.

Never the less, the present implementation has some limitations which are observed. The fight detection system is founded on a simplified concept that takes into account the existence of several people in a frame. Although this method is computationally low, it fails to reflect a detailed interaction pattern and thus makes false positive results in crowded/socially active settings without any aggressive behaviour.

Besides this, the accuracy of pose estimation and fall detection can be influenced by the elements of the environment including lighting variations, occlusions and camera angles. The general performance of the detection can be worsened in a case where the body landmarks cannot be clearly seen. Even though the system has fallback mechanisms, more efforts are needed to make it more robust in case of difficult conditions.

The weaknesses of pre-trained models are another limitation since this might limit generalization to a wide and evolving real-world setting. Detection consistency can be affected by differences in background, clothing, and quality of the camera. Introducing adaptive learning or training of domains would enhance system performance in such a case.

Never the less, in spite of these shortcomings, the suggested system offers a good base on which intelligent surveillance applications can be built. Its capability of combining detection, analysis, and alert generation, in a single platform is indicative of practical utility. Moreover, the fact that it can generate interpretable outputs and maintain structured records are further improved to make it useful in a monitoring and decision-making application.

The next steps can be to enhance the activity recognition by adding the motion-based analysis and other sophisticated deep learning systems like sequence-based architecture. Also, it can be further optimized to introduce edge deployment as well as increasing its performance in harsh environments which can further boost its practical application.

All in all, the suggested system proves the possibility of the synergistic approach of employing deep learning and rule-based methods to detect human activities in real-time. The existing implementation has certain room to be improved, but nonetheless, it has succeeded in offering a dependable and effective solution to the surveillance of critical human behaviors within the surveillance environment.

VI. REFERENCES

[1] Bouachir, M. Bilodeau, and F. Porikli, “A survey of deep learning methods for abnormal event detection in surveillance videos,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 8, pp. 2307–2325, 2021.

[2] H. Park, J. Noh, and B. Han, “Learning memory-guided normality for anomaly detection,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14372–14381, 2020.

[3] W. Li and S. Onie, “Vision-based suicide prevention using deep learning techniques,” IEEE Access, vol. 9, pp. 124115–124126, 2021.

[4] W. Bao, Q. Yu, and Y. Kong, “Hierarchical scene normality modeling for video anomaly detection,” IEEE Transactions on Image Processing, vol. 30, pp. 2046–2058, 2021.

[5] A. Acsintoae et al., “UBnormal: New benchmark for supervised open-set video anomaly detection,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 20143–20153, 2022.

[6] A. Nadeem, M. H. Khan, and M. Tahir, “An intelligent surveillance system for missing person detection using deep learning,” IEEE Access, vol. 10, pp. 56271–56281, 2022.

[7] X. Liu, J. Wang, and H. Lu, “Temporal action detection with transformer networks for video understanding,” IEEE Transactions on Multimedia, vol. 24, pp. 1678–1689, 2022.

[8] Y. Huang, “Contrastive learning based anomaly detection for intelligent surveillance systems,” IEEE Access, vol. 10, pp. 103514–103525, 2022.

[9] A. Ullah, “Real-time face detection and recognition for CCTV surveillance systems using deep learning,” IEEE Access, vol. 9, pp. 159771–159782, 2021.

[10] Z. Liu, Y. Zhang, and S. Chen, “Prompt-based feature mapping for anomaly detection in surveillance videos,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 33, no. 4, pp. 2154–2165, 2023.

[11] Y. Huang, X. Liu, and H. Wang, “Self-supervised generative adversarial networks for abnormal event detection in surveillance videos,” IEEE Transactions on Multimedia, vol. 24, pp. 3310–3321, 2022.

[12] J. Yang, “Spatio-temporal feature learning for non-suicidal self-injury behavior detection,” IEEE Access, vol. 11, pp. 11834–11845, 2023.

[13] K. Liu, Y. Fu, and S. Li, “AMP-Net: Appearance-motion prototype network for video anomaly detection,” IEEE Transactions on Image Processing, vol. 31, pp. 6894–6906, 2022.

[14] L. Zhao, Y. Li, and J. Zhang, “Occluded person re-identification using deep feature smoothing,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 9, pp. 3520–3531, 2021.

[15] Z. Lin, “Temporal action recognition for abnormal event detection in intelligent surveillance systems,” IEEE Access, vol. 10, pp. 84522–84534, 2022.