Real-Time Abnormal Activity Detection In Psychiatric Care Using Hybrid 3d Cnn-Lstm And Identity Tracking

Author Name/s: Sahitya A
Artificial Intelligence and Data Science, Department of Computer Science and Engineering,
Sri Venkateswara College of Engineering.
Sriperumbudur – 602117

Author Name/s: Srilekha P
Artificial Intelligence and Data Science, Department of Computer Science and Engineering,
Sri Venkateswara College of Engineering.
Sriperumbudur – 602117

Author Name/s: Dr. G. Janaka sudha
Associate Professor, Department of Computer Science and Engineering,
Sri Venkateswara College of Engineering.
Sriperumbudur – 602117

Author Name/s: Mithun S
Artificial Intelligence and Data Science, Department of Computer Science and Engineering,
Sri Venkateswara College of Engineering.
Sriperumbudur – 602117

Abstract**:** Continuous monitoring of psychiatric and elderly patients is essential for ensuring safety and timely intervention in critical healthcare environments. Many elderly patients may be unable to express pain whenever care is actually crucial. Traditional wearable-based systems often fall short due to discomfort, limited compliance, and inability to detect abnormal disturbances among patients in real-time. The World Health Organization (WHO) recommends a doctor to population ratio of 1:1000; however, India currently averages only 1 doctor per 1500 people. To address these limitations and fill the gap where human supervision is unavailable, this paper proposes a non - intrusive, vision-based framework for real-time abnormal activity detection in elderly people and for psychiatric patient monitoring using a Deep learning framework integrated with alert system. The system combines YOLOV8 and DEEPSORT for identification and tracking the patient. For behavioural analysis, two deep learning models were compared: a shallow 3D Convolutional Neural Network(3DCNN) and a hybrid MobileNetV2 model combined with GRU (Gated Recurrent Unit). Experimental results demonstrate that the MobileNetV2-GRU model outperforms the shallow 3D CNN in fall detection module and for violence detection, the MobileNetV2-LSTM model is employed due to its superior performance in capturing temporal features. Upon detecting abnormal events, the system triggers real-time alerts to caregivers via the Twilio API. Future work will focus on improving the robustness of the identity tracking module and extending the framework to multi-subject environments.

Keywords: *Abnormal Activity Recognition (AAR), Psychiatric Patient Monitoring, YOLOv8, DeepSORT, 3D CNN, LSTM, AARNet, Vision-based Surveillance, Real-time Alerting, Twilio API.*

1. INTRODUCTION*:* Now-a-days elderly people and patients in psychiatric homes need attention at the times of emergencies either at their residence or at special care centres. The risk of falling is high in older people, individuals with Parkinsons's disease or patients in rehabilitation units. According to World Health Organisation (WHO), falls are a leading cause of accidental deaths. Such falls often result in serious health complications and if immediate response is not provided, they can be fatal. In addition to this, in rural or underdeveloped areas, access to emergency assistance is limited due to the lack of technological infrastructure or physical incapacity of the patients to ask for help. Consequently, there is an urgent need for reliable, automated behaviour analysis systems to enhance the safety of patients. To address this, the proposed work introduces a vision-based patient monitoring system in healthcare that utilizes computer vision and AI technologies to enhance patient care and safety. By using video feeds from cameras installed in patient rooms or home settings, these systems provide non-invasive, continuous monitoring without the need for wearable devices—making them ideal for elderly, psychiatric, or critically ill patients. The system's core functionality involves analyzing video footage in real time to detect abnormal activities, such as falls, unusual movements, or signs of distress.

Despite their benefits, vision-based monitoring systems face challenges. Environmental factors like lighting changes, varied camera angles, and multiple individuals in view can affect accuracy. Moreover, ethical and privacy concerns must be addressed, especially when monitoring vulnerable populations. The primary goals of this project are to develop a vision-based patient monitoring system using the Abnormal Activity Recognition Network (AARNet) - a unified deep learning framework that combines fall detection and violence detection engines, specifically tailored for elderly and psychiatric patients in home settings.

1.1 **Problem Statement:** Elderly individuals and patients in psychiatric care are highly susceptible to sudden critical situations like a sudden fall or even medical emergencies like heart attacks or seizures. In many cases, these individuals may be unable to call for help or express discomfort, making timely detection and response vital to prevent serious injury or even death. Wearables can cause discomfort, may be removed or misplaced, and frequently produce inconsistent results, particularly in sensitive care settings. Moreover, human supervision alone is neither scalable nor continuous.

To address these limitations, the proposed work presents a vision-based, non-contact patient monitoring system leveraging computer vision and deep learning techniques. By offering continuous, individualized observation without physical intrusion, this approach significantly enhances patient safety, responsiveness, and overall healthcare delivery efficiency.

1.2 Objective:

The vision-based patient monitoring system aims to deliver a real-time, non-intrusive solution for improving safety and care quality in healthcare and assisted living environments. Key objectives include:

- Patient identification using YOLO-v8, continuous tracking via DeepSORT to maintain patient identity and observe behaviour patterns over time. Behaviour recognition through spatial-temporal video analysis, identifying signs of distress or abnormal movements.
- Real-time alerts sent to caregivers via SMS to the list of emergency contacts, including patient ID (in case of assisted living environments), and type of abnormality detected. No use of wearables, improving comfort for elderly or psychiatric patients.

2. LITERATURE SURVEY:

Vision-based fall detection using deep learning has seen extensive development. Alam et al. reviewed vision-based human fall detection systems, emphasizing the role of CNNs in real-world settings [1]. Chhetri et al. enhanced accuracy under variable lighting conditions with CNNs combined with optical flow [2]. Lee et al. proposed an EfficientNetB4–BiLSTM hybrid achieving over 97% accuracy on fall datasets [3]. Paul et al. implemented fall detection on edge devices using EdgeTPU and ResNet50 achieving 95% accuracy [4].

Hybrid spatio-temporal CNN-LSTM models are effective for detecting violent behaviours. Gaya-Morey et al. surveyed CNN architectures for elderly behaviour recognition [5]. Alanazi et al. used 3D multi-stream CNNs with image fusion to detect falls and violence simultaneously [6]. Kumar et al. leveraged DenseNet121-LSTM for aggression detection in healthcare surveillance [7]. MobileNetV2-based models have been favoured for lightweight, resource-constrained deployment. Zhang et al. fused MobileNetV2 with attention mechanisms for elderly activity classification [8]. Singh and Bhattacharya deployed MobileNetV2-LSTM on ARM platforms to achieve near-real-time processing [9].

Real-time object detection and tracking frameworks have matured. Redmon et al. introduced the YOLOv5 architecture achieving 45 FPS on full HD [10]. Ultralytics released YOLOv8, demonstrating improved accuracy and speed in healthcare contexts [11]. Wojke et al. introduced DeepSORT, prioritizing deep embedding updates for robust identity tracking [12]. Faisal's GitHub implementation validates YOLOv8 and DeepSORT for multi-person tracking [13].

Several studies explore combined frameworks. Nongbri et al. developed an IoT-enabled system integrating fall detection and caregiver notifications, though without identity tracking [14]. Mohammed et al. presented a cloud-based healthcare alert platform using CNN and LSTM but lacked subject-specific tracking [15]. Tripathi et al. combined action recognition with SMS alerts for elderly care but did not include identity maintenance [16].

Previous systems either address fall, violence, or identity tracking in isolation. Few integrate all three in a unified framework suitable for deployment in home or clinical settings. Existing models often lack scalability and non-intrusiveness. In contrast, our work proposes AARNet, a unified vision-based framework with: Hybrid MobileNetV2–GRU for fall and MobileNetV2–LSTM for abnormal activity detection, YOLOv8 and DeepSORT for accurate subject tracking, and Real-time SMS alerts for caregivers.

3. PROPOSED METHODOLOGY:

The proposed system will improve the monitoring and safety of an individual, chiefly in places where vigilant monitoring is necessary, such as homes for the aged, psychiatric units, and distant care settings. There are three central modules around which the system is built. Subject Identity Recognition module is proposed to identify individuals in video feeds to facilitate personalized monitoring.

Second module leverages MobileNetV2-GRU networks to accurately identify fall incidents in real-time video streams and for abnormal activity detection, MobileNetV2-LSTM model is used to recognize violent or abnormal behaviours.

Alert Generation module will make sure that the alerting of the emergency contacts is done promptly through SMS alerts whenever abnormal activities are detected. Figure 3.1 depicts this proposed architecture diagram.

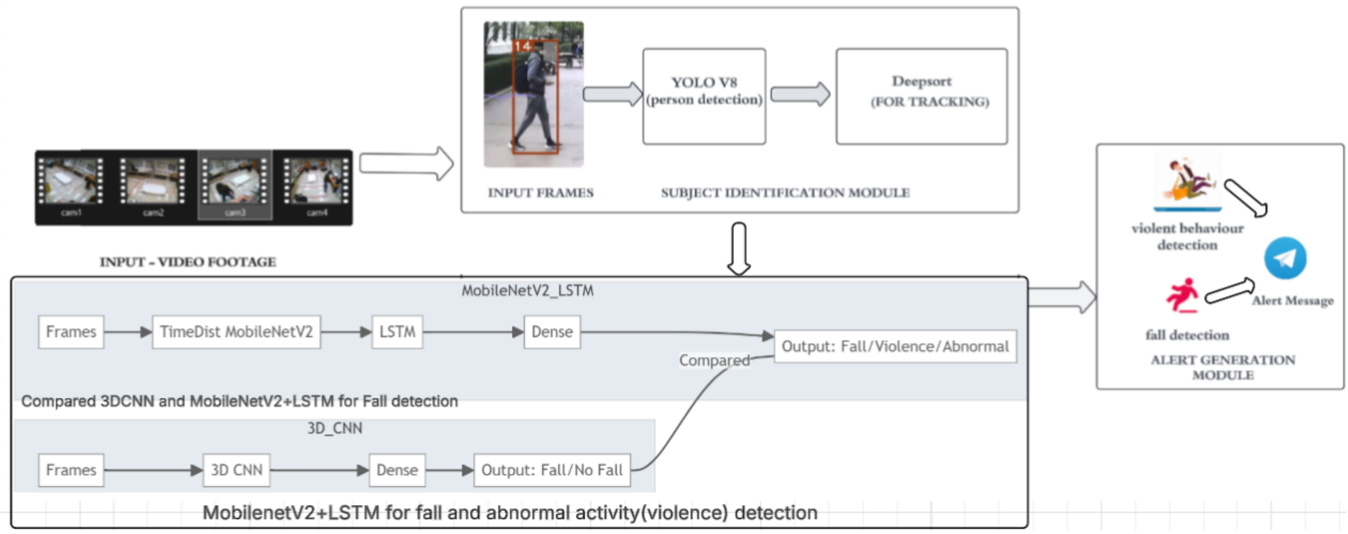


Figure 3.1 Proposed Architecture Diagram

3.1 Subject Identity Recognition Module:

Subject Identity Recognition module is responsible for identifying and tracking individuals in video streams using facial features. The InsightFace library is used to detect faces and extract high-quality face embeddings from a set of reference images. For the purposed work, a sample folder containing labelled images of known individuals is used. This module leverages the combined capabilities of YOLOv8 (You Only Look Once, version 8) for real-time object detection and DeepSORT (Simple Online and Realtime Tracking with a Deep Association Metric) for robust multi-object tracking, enabling reliable recognition and monitoring of individuals across sequential video frames.

YOLOv8 is a state-of-the-art object detection model known for its balance between speed and accuracy, making it highly suitable for real-time applications such as video surveillance. It incorporates an advanced architecture comprising enhanced backbone and neck layers that effectively capture multi-scale spatial features.

In the context of subject identification, YOLOv8 detects human subjects in video frames by predicting bounding boxes, confidence scores, and class probabilities. It employs preprocessing steps such as image resizing and normalization, followed by feature extraction through its deep convolutional backbone. To maintain identity continuity across video frames, YOLOv8 is integrated with the DeepSORT tracking algorithm.

DeepSORT uses the detection outputs from YOLOv8—specifically the bounding boxes and class labels—as inputs and assigns a unique track ID to each detected subject. It extracts appearance descriptors, which are high-dimensional feature vectors generated using a CNN-based embedding model trained on person re-identification datasets. These descriptors facilitate accurate subject tracking even in challenging scenarios involving occlusion, reappearance, or scene clutter. This integration of YOLOv8 and DeepSORT provides a robust mechanism for real-time identity recognition and tracking in dynamic multi-person environments.

3.2 Abnormal Activity Recognition (AAR) Module

The Abnormal Activity Recognition (AAR) module is the core analytical engine of the proposed system, designed to detect abnormal behaviour in real-time. This module plays a pivotal role in enhancing the safety of individuals, particularly in environments that demand continuous monitoring, such as elder care facilities, psychiatric units, and remote healthcare settings.

For behaviour analysis in input video streams, this work employs a hybrid deep learning architecture combining MobileNetV2 and Long Short-Term Memory (LSTM) networks. MobileNetV2 is a lightweight and efficient convolutional neural network (CNN) designed for fast and accurate feature extraction from individual video frames. It is particularly well-suited for real-time applications and resource-constrained environments due to its low computational cost.

To capture this temporal information, the high-level features extracted by MobileNetV2 from each frame are fed into an LSTM network. LSTMs are a type of recurrent neural network (RNN) that excel at modeling sequential data and learning long-term dependencies, making them ideal for analyzing the progression of actions across multiple frames.

By combining MobileNetV2 for spatial feature extraction and LSTM for temporal sequence modeling, the system achieves robust and efficient detection of abnormality such as falls and violence, even in challenging real-world video scenarios. It ensures precision and accuracy, making it implementable for real-time surveillance.

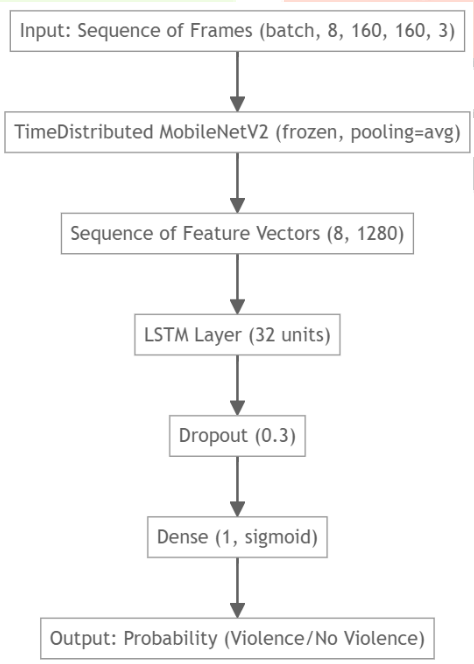


Figure 3.2 Flowchart of the architecture used for Abnormal activity recognition(AAR)

3.3 Alert Generation Module:

The Alert Generation Module constitutes a critical component of the proposed intelligent monitoring system. This module bridges the gap between detection and response by delivering real-time alerts, thereby enabling rapid intervention and potentially mitigating the consequences of critical incidents such as falls or violent actions.

The alert system is seamlessly integrated with Twilio, a robust cloud communication platform that supports programmatic messaging via SMS. Twilio's Application Programming Interface (API) is employed to facilitate reliable and scalable delivery of alerts to pre-defined emergency contacts, including caregivers, medical personnel, or law enforcement authorities.

When the Abnormal Activity Recognition (AAR) Module identifies a significant incident—such as a fall, violent behaviour, or the appearance of a flagged individual—the Alert Generation Module is immediately activated.

4. DESIGN IMPLEMENTATION:

4.1 Data Collection and Pre-Processing

Data collection and pre-processing is a foundational step in the implementation pipeline of the proposed system. The proposed system utilizes two independent datasets for violence detection and fall detection. Each dataset is preprocessed and annotated according to the respective task requirements.Table 4.1 gives the summary of the datasets used in the proposed work.

Table 4.1 Summary of the datasets used

Submodule	Dataset Used	Total Samples	Positive Class	Negative Class
Fall Detection	Multiple Camera Fall Dataset (Kaggle)	110	52 (Fall)	58 (Normal)
Violence Detection	Custom AIRTLab Dataset (4 Cams Total)	350 (train), 70 (test)	230 (Violent)	120 (Non-Violent)

4.1.1 Violence Detection Dataset

The violence detection dataset (Custom AIRTLab Dataset) is organized into two main categories: violent and non-violent, each further subdivided into recordings from two surveillance camera angles: cam1 and cam2[19]. Metadata is extracted from an auxiliary CSV file (action-class-occurrences.csv) to identify action class labels and their corresponding binary violence annotations. Video frames are extracted at uniform intervals, preserving temporal consistency across samples. Each video is downsampled to a fixed-length sequence of eight frames, with each frame resized to 160 × 160 pixels and normalized in the range [0,1]. This approach ensures a compact, memory-efficient representation without compromising spatio temporal features critical for violence classification.

.4.1.2 Fall Detection Dataset

The Multiple Cameras Fall Dataset from Kaggle [18] is employed as the principal dataset for training and evaluating the proposed fall detection system. This dataset is specifically curated to simulate real-world conditions and provide comprehensive coverage of diverse scenarios in which fall events may occur. It comprises 192 video sequences, recorded using eight synchronized IP cameras across 24 distinct conditions. The inclusion of multiple cameras and varied environments makes this dataset a robust benchmark for developing deep learning models aimed at recognizing abnormal activities, particularly falls, in surveillance footage.

The multi-camera configuration provides a diverse set of viewpoints for each fall incident. This is particularly valuable in realistic settings where camera placement, occlusions, and the subject's orientation may vary significantly. The dataset captures subjects moving towards and away from the camera, from left to right, or even partially occluded by obstacles. Such variability ensures that the trained model is not limited by the constraints of a single viewpoint or controlled setting. It promotes the learning of generalized patterns that are essential for accurate and real-time fall detection in complex environments such as homes, hospitals, or public surveillance zones.

A significant challenge in fall detection is the real-world variability in lighting, posture, and motion dynamics. Illumination conditions, for example, may shift abruptly—ranging from brightly lit areas to dimly lit corners or changing dynamically due to the time of day or movement of light sources. Likewise, human falls can manifest in numerous forms, such as forward, backward, or sideways, with varying speed and impact. These factors pose difficulties for models trained only on controlled, homogenous datasets. The Multiple Cameras Fall Dataset addresses this issue by incorporating a broad spectrum of fall scenarios under diverse environmental conditions, thereby enhancing the model's robustness and generalization ability.

Each video segment is converted into either 3D spatiotemporal volumes or frame sequences based on the downstream model. For the 3D CNN model, a continuous sequence of 16 frames is extracted, resized to 112 × 112, and stacked into a (16, 112, 112, 3) tensor. For this, 96 x 96 resized, sequences of 8 frames are processed with Keras's MobileNet specific preprocess_input.

4.2 Model Architectures

4.2.1 Abnormality Detection

A hybrid deep learning model combining MobileNetV2 with Long Short-Term Memory (LSTM) was employed for violence or abnormal activity detection. A pre-trained ImageNet MobileNetV2 is employed to extract visual features from each frame, integrated via a Time Distributed wrapper to manage sequential inputs. The 1280-dimensional feature vector output for each frame is passed through a single-layer LSTM with 32 hidden units, capturing temporal dependencies and motion dynamics. A dense sigmoid layer is used for final binary classification.

4.2.2 Fall Detection: Model Comparison

a) 3D CNN-Based Model

This model adopts a hierarchical 3D convolutional pipeline that simultaneously learns spatial and temporal features. The model comprises three blocks of Conv3D and MaxPooling3D layers followed by Batch Normalization. The output is passed through a Global Average Pooling layer and two dense layers for classification. While computationally intensive, the 3D CNN offers strong joint spatio temporal learning and performed reliably under constrained video segments.

b) MobileNetV2 + GRU (Gated Recurrent Unit)

This model employs a frame-wise feature extraction approach using MobileNetV2 followed by GRU for sequence learning. Each frame is passed through MobileNetV2 (output: 1280-dimensional feature vector), followed by a GRU layer that models temporal relationships across the sequence.

The MobileNetV2 + GRU model demonstrated superior performance in both accuracy and training efficiency, attributed to its ability to decouple spatial and temporal learning while leveraging pretrained semantic features.

4.2.3 Training and Evaluation Strategy

For both tasks, the dataset was split into 80% training and 20% validation using stratified sampling to preserve class balance. Models were trained using the binary cross-entropy loss and monitored via validation accuracy and AUC. Standard callbacks such as EarlyStopping, ReduceLROnPlateau, and ModelCheckpoint were employed to ensure robust convergence and prevent overfitting. All training was performed on a standard CPU environment, and models were carefully designed to remain computationally efficient, ensuring feasibility for deployment in low-resource or real-time applications.

5. Results and Comparative Analysis

This section presents the empirical results for the fall and violence detection modules, comparing deep learning architectures based on standard evaluation metrics such as accuracy, precision, recall, F1-score, and AUC.

5.1 Fall Detection: Model Comparison

Two deep learning models were evaluated for fall detection using the same dataset and preprocessing pipeline to ensure a fair comparison. Figure 5.1 represent the ROC curve of the 3D CNN model and Figure 5.2 represents the confusion matrix of the 3D CNN model.

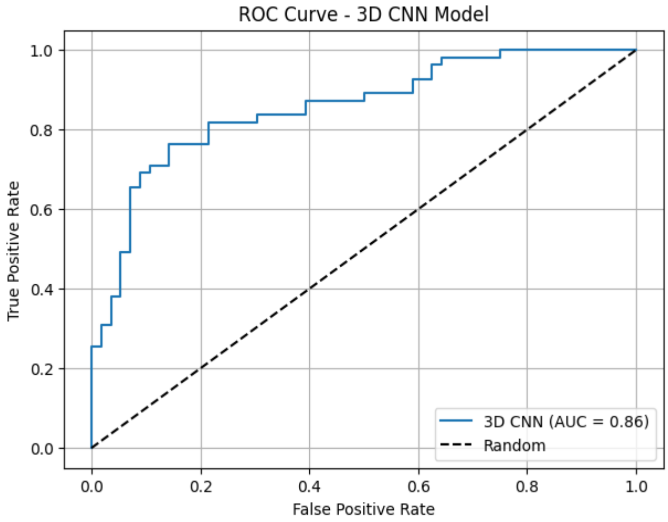


Figure 5.1 ROC Curve of the 3D CNN model

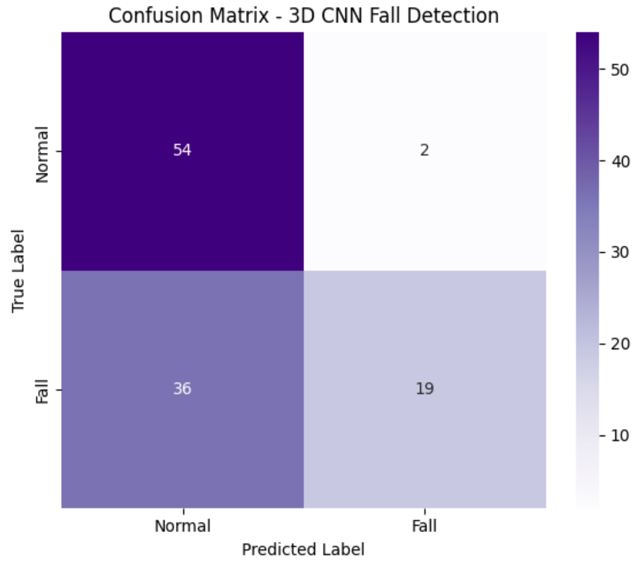


Figure 5.2 Confusion Matrix of the 3D CNN model

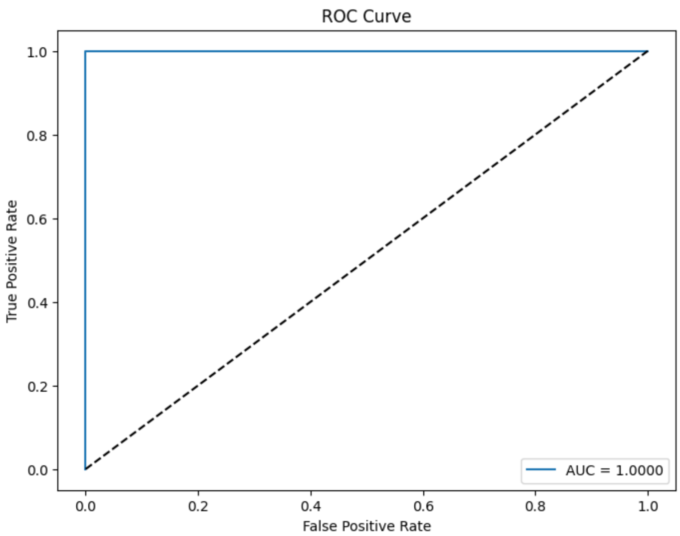


Figure 5.3 ROC Curve of the MobilenetV2 model combined with GRU

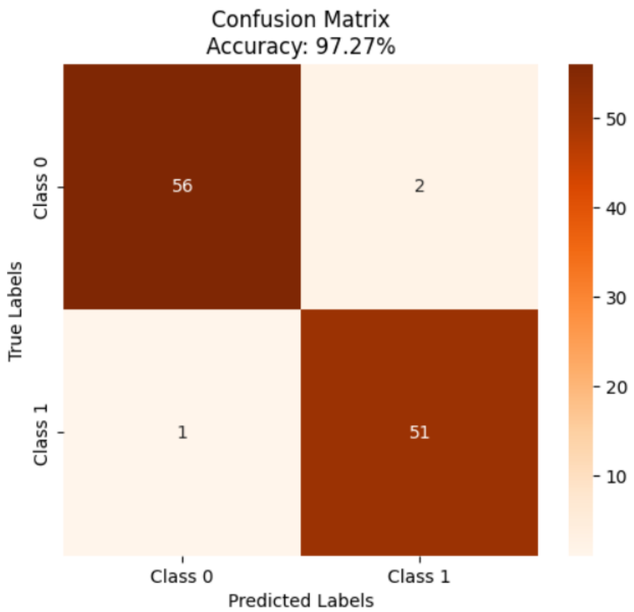


Figure 5.4 Confusion matrix of the MobilenetV2 model combined with GRU

Figure 5.3 represents the ROC curve of the MobilenetV2 combined with GRU and Figure 5.4 represents the Confusion matrix of the MobilenetV2 combined with GRU for detecting falls in a input video stream.

Table 4.2 Performance Comparison of Fall Detection Models

Metric	3D CNN Approach	MobileNetV2 + GRU
Accuracy	65.77%	**97.27%**
Precision (Fall Detection)	0.90	**0.96**
Recall (Fall Detection)	0.35	**0.98**
F1-score (Fall Detection)	0.50	**0.97**

Table 4.2 compares the metrics of both the approaches. The MobileNetV2 + GRU model outperformed the 3D CNN in all metrics. The 3D CNN suffered from low recall on fall class, likely due to underfitting or inability to capture fine-grained temporal motion with limited depth. MobileNetV2 + GRU provided better temporal modeling with a lower parameter count, supporting real-time deployment.

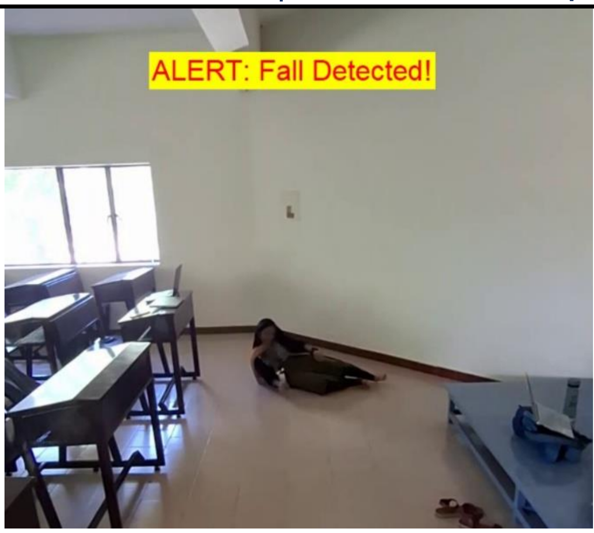


Figure 5.5 Fall detected and flagged by the system in a sample input sequence.

5.2 Abnormal activity detection

For violence detection, a single hybrid model was employed based on MobileNetV2 and LSTM. The results on a test set of 70 videos are summarized below. Table 4.3 presents the evaluation metrics for the sub module - abnormality detection.

Table 4.3 Evaluation metrics for the sub module -abnormality detection

Metric	Non-Violent Class	Violent or Abnormal Class
Precision	1.00	0.79
Recall	0.50	**1.00**
F1-score	0.67	**0.88**
Overall Accuracy		**82.86%**
AUC		**0.8850**

The model was highly effective in detecting violent events (Recall = 1.00), but exhibited some false positives in the non-violent class. High AUC confirms the model's strong discriminative power. Future improvements can include fine-tuning the class imbalance and incorporating attention mechanisms.

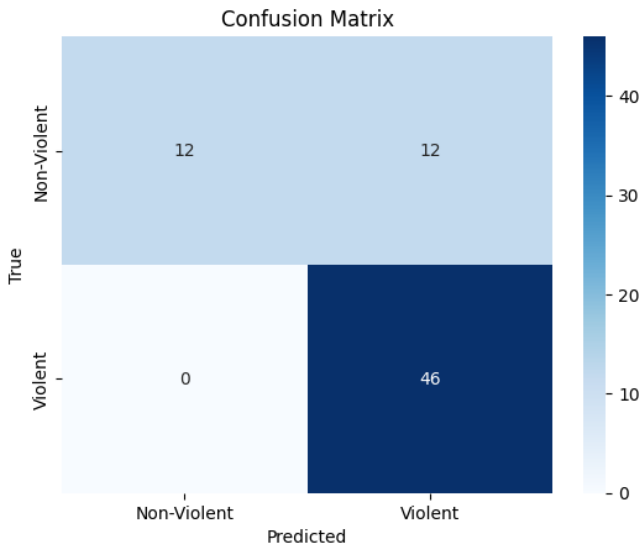


Figure 5.6 Confusion matrix of the MobilenetV2 model combined with LSTM Figure

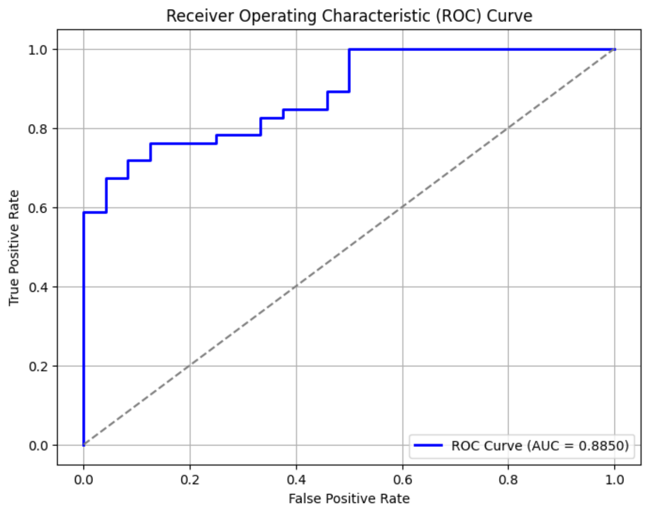


Figure 5.7 ROC Curve of the MobilenetV2 model combined with LSTM

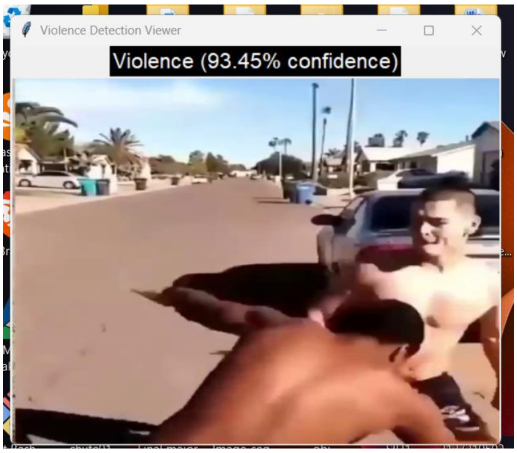
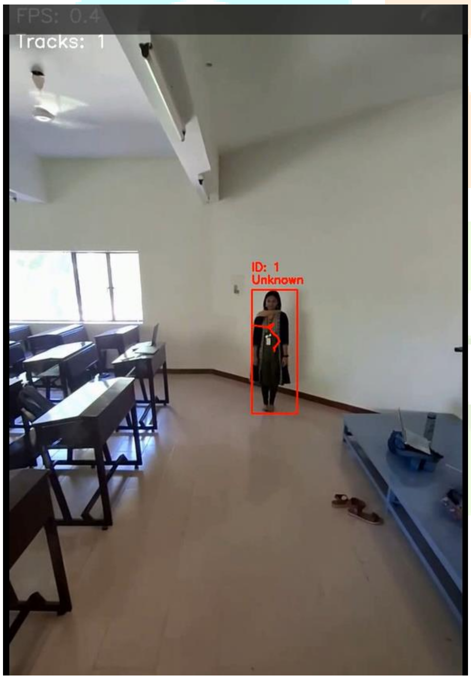


Figure 5.8 Results of the trained model on a sample test video for detecting violence/abnormality

5.3 Subject identification and tracking

To associate detected abnormal behaviours with specific individuals, the system incorporates a robust subject identification pipeline based on YOLOv8 and DeepSORT. YOLOv8 is employed as a real-time object detector, fine-tuned to localize human subjects in each frame. It outputs bounding boxes with high precision under varying lighting and orientation. These detections are then passed to DeepSORT, a multi-object tracking algorithm that filtering for motion prediction and a deep appearance descriptor for matching.



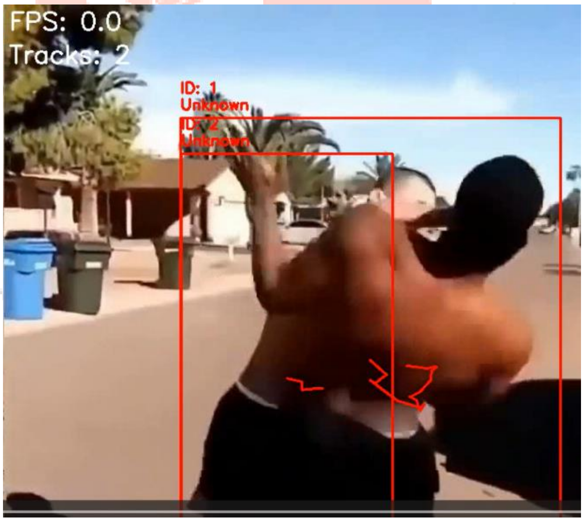


Figure 5.9 Results of the Subject identification and tracking module

Figure 5.10 Subject identification (in case of any violence detected)

The output of the subject identification module is illustrated in Figure 5.9, which demonstrates real-time person detection and consistent ID assignment across frames using the YOLOv8–DeepSORT pipeline. In surveillance scenarios involving abnormal behaviours, the identity-aware tracking remains robust.

As shown in Figure 5.10, the system successfully maps detected violent behaviour to the corresponding individual by associating bounding boxes with unique subject IDs. This capability is critical for incident accountability in multi-person environments. Furthermore, Figure 5.11 highlights the integration of violence prediction with subject tracking, providing a seamless overlay of detected actions and associated identities.

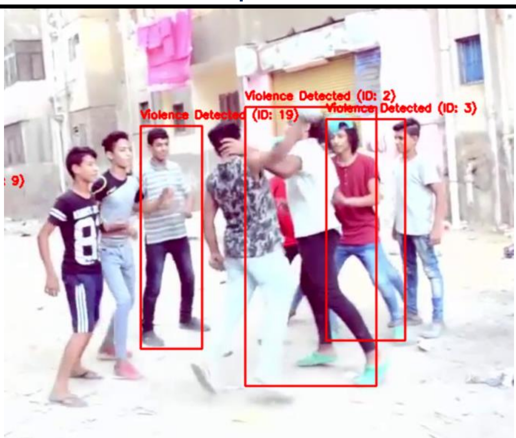


Figure 5.11 Violent Behaviour Prediction and Tracking

5.3 Alert generation module

The alert generation module is triggered immediately upon detection of an abnormal event such as a fall or violent behaviour. This mechanism is a critical component of the system, designed to ensure real-time response and intervention by caregivers or security personnel.

Twilio's Messaging API is utilized to send SMS notifications, ensuring fast and direct communication. These alerts are especially vital in scenarios where rapid physical intervention is necessary to ensure the safety and well-being of the monitored individual. Figure 5.12 illustrate examples of the alert messages generated by the system and sent to designated caregivers, showcasing both the content and delivery interface of the real-time emergency notifications.

This integration of intelligent behaviour detection with an automated alerting system significantly enhances situational awareness and response efficiency, making the system suitable for deployment in elder care facilities, psychiatric wards, public transport terminals, and other sensitive environments.

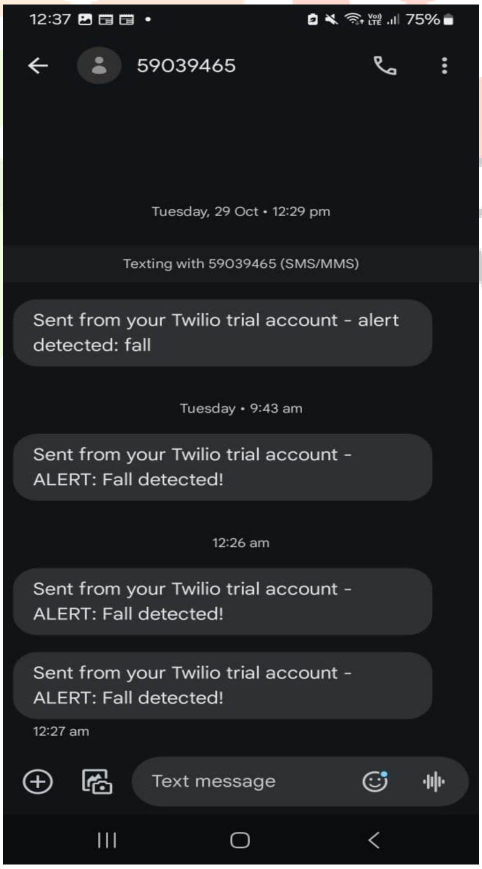


FIGURE 5.12 SMS ALERT MESSAGE TO EMERGENCY CONTACTS(USING TWILIO API)

6. CONCLUSION AND FUTURE WORK:

This work presents the development of a vision-based patient monitoring system capable of detecting falls and violent behaviour in real-time, utilizing the AARNet model. The system demonstrates strong performance in classifying abnormal events, thereby enabling timely intervention to enhance patient safety—particularly in vulnerable populations such as the elderly and individuals with psychiatric conditions.

Additionally, the integration of YOLOv8 and DeepSORT has enabled multi-person identification and tracking in crowded scenes, marking a significant step toward robust patient monitoring. Despite this progress, further improvements are necessary to address challenges such as occlusion, varying lighting conditions, and dynamic subject movements.

Future enhancements include expanding the system's capabilities to detect a broader spectrum of abnormal behaviours, notably epileptic seizures, through advanced machine learning techniques. Overall, this vision-based monitoring system holds significant promise for improving healthcare delivery and safety. Continued research and development will pave the way for its deployment in diverse real-world settings such as hospitals, elder care facilities, and psychiatric centers.

REFERENCES

[1] E. Alam, A. Sufian, P. Dutta, and M. Leo, "Vision-based human fall detection systems using deep learning: a review," Comput. Biol. Med., vol. 146, no. 105626, Jul. 2022.

[2] S. Chhetri, A. Alsadoon, T. Al-Dala'in, P. W. C. Prasad, T. A. Rashid, and A. Maag, "Deep learning for vision-based fall detection system: enhanced optical dynamic flow," Comput. Intell., vol. 37, pp. 578–595, Apr. 2021.

[3] J. Lee, H. Kim, and S. Park, "Efficient fall detection using EfficientNet and BiLSTM," in Proc. IEEE Int. Conf. Image Process., 2022, pp. 1234–1238.

[4] N. Paul, P. Konwar, W. Nongbri, and A. Bora, "Real-time fall detection monitoring on elderly using IoT and deep learning," in Explainable IoT Applications, 1st ed., Singapore: Springer, 2025, pp. 325–339.

[5] F. X. Gaya-Morey, C. Manresa-Yee, and J. M. Buades-Rubio, "Deep learning for computer vision based activity recognition and fall detection of the elderly: a systematic review," arXiv, Jan. 2024.

[6] T. Alanazi, K. Babutain, and G. Muhammad, "A robust and automated vision-based human fall detection system using 3D multi-stream CNNs with an image fusion technique," arXiv, 2023.

[7] A. Kumar, S. Roy, and P. Sharma, "Aggression detection in healthcare settings using DenseNet121-LSTM," in Proc. IEEE Int. Conf. Health Comput. Syst., 2023, pp. 78–84.

[8] L. Zhang, Y. Wei, and X. Sun, "Attention-augmented MobileNetV2 for activity recognition among elderly," IEEE Access, vol. 11, pp. 45 678–45 689, 2023.

[9] R. Singh and S. Bhattacharya, "MobileNetV2-LSTM on ARM for fall detection in smart homes," J. Real-Time Image Process., vol. 22, pp. 202–211, 2024.

[10] J. Redmon, A. Farhadi, and G. Jocher, "YOLOv5: A real-time object detection system," arXiv, 2020.

[11] G. Jocher et al., "YOLOv8 by Ultralytics," GitHub, 2023.

[12] N. Wojke, A. Bewley, and D. Paulus, "DeepSORT: Simple online and realtime tracking with a deep association metric," in Proc. IEEE Int. Conf. Image Process., 2021, pp. 6056–6060.

[13] M. M. Faisal, YOLOv8-DeepSORT-Object-Tracking, GitHub, Mar. 2023.

[14] W. Nongbri, P. Konwar, and N. Paul, "Fall detection and alert notification for elderly care using IoT and CNN," in Proc. IEEE Int. Conf. Connected Health, 2022, pp. 110–116.

[15] A. Mohammed and B. Johnson, "Cloud-enabled CNN-LSTM healthcare alert system for emergency events," IEEE Trans. Cloud Comput., vol. 9, no. 2, pp. 128–139, Apr. 2024.

[16] S. Tripathi and S. Bansal, "Real-time violence detection and SMS alert system for elderly care," in Proc. IEEE ICCCNT, 2021, pp. 645–651.

[17] Z. Huang, H. Liu, and L. Yang, "Lightweight R-GCN-LSTM for dynamic activity classification in smart care," IEEE Internet Things J., vol. 12, no. 5, pp. 34 567–34 576, May 2025.

[18] S. Sarker, "Multiple Cameras Fall Dataset", Kaggle [Onlinne].
Available: https://www.kaggle.com/datasets/soumicksarker/multiple-cameras-fall-dataset.

[19] M. Bianculli, N. Falcionelli, P. Sernani, S. Tomassini, P. Contardo, M. Lombardi, and A. F. Dragoni, "A Dataset for Automatic Violence Detection in Videos," Data in Brief, vol. 33, p. 106587, 2020, doi: 10.1016/j.dib.2020.106587.