Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

16

Smart Sentinel - Activity Detection AI

Divya Mahale

Student

Dept. Of Polytechnic, Dr. Vishwanath Karad's MIT World Peace University

Sulakshana S. Malwade

Assistant Professor

Dept. Of Polytechnic, Dr. Vishwanath Karad's MIT World Peace University

Abstract

A Smart Sentinel platform built using Activity Detection AI can help ensure the safety and well-being of individuals and the community by detecting health incidents, accidents, violence, and intruders. The AI-based solution can monitor people in an area. Once AI detects an incident, it can send alerts to authorized personnel, and health organizations and make announcements to get the attention of people in the area. Further, gesture controls can help to confirm if someone is all right, or the person can make gestures to get help. Traditionally, CCTV or other video monitoring solutions only captured video feed and required someone to manually monitor the feed 24 x 7. It was a resource incentive and error-prone mechanism. New age AI (Computer Vision) can meticulously monitor the video feed 24 x 7 detect incidents, report, and take necessary actions. People can use the platform on mobile or web apps.

Keywords: Artificial Intelligence, Machine Learning, Activity Detection and Gesture Detection

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

Introduction

Activity Detection is made possible by recent significant developments in Artificial Intelligence (Computer Vision) and semiconductors, which have allowed building faster, bigger, and better machines to run AI engines. Cloud technology has also helped reduce end-point processing needs by utilizing cloud resources instead for faster and more aggressive data processing.

AI-based solutions can easily monitor health incidents, hazards, intruders at home, workplaces community centers, etc. IoT devices or traditional CCTV setups can be used to capture video feeds. Computer vision can then process the video streams or recorded video feeds to detect activities. Based on rigorous training the AI model can learn to flag health incidents.

The AI-based solution can be evolved to cater to various use cases viz. security, health tracking, old age homes, community spaces, etc. Furthermore, health issues and illnesses can be recorded so AI can look for anticipated health issues and monitor specific behavior indicators.

AI can be designed to monitor an area, detect health issues like seizures, heart attacks, fits, or other incidents like fighting, detect the presence of unauthorized people, or detect hazards like smoke and/or fire, flood, and electricity failures. Once it detects something, it can be configured to send alerts to authorized personnel and make announcements (using text-to-voice libraries) to guide or warn the subjects.

Gesture detection can be added to avoid false positives. Once an incident is detected, the platform can request the person to confirm if they are okay. They can gesture a “Thumbs Up” to confirm all is okay otherwise show a “Thumbs Down” or do nothing (the most obvious action when they are experiencing a health incident). The platform can decide what to do next. It can either ignore the incident or take necessary actions to address it.

The complete AI solution may comprise a mobile app and a web app. Mobile apps can allow easy access to various feeds and uploading of pictures from mobile cameras. The web app will house the AI model, and database and provide an elaborate front end for further configuration and monitoring. The web app can also be used for live tracking, monitoring, and

announcements by security and other professionals in various community centers, health care centers, etc. This can be a reliable, dependable, and trustworthy system that will work 24/7.

We can team up with prospective users and get real-life data for training and reinforcement training. This will increase the precision of the model and reduce false positives and negatives.

This research is meant to be a high-level survey of various algorithms and computer vision-based solutions for Activity Detection or Face Recognition. This is to have a holistic view of what is possible currently and what it can grow into while looking at what it was years before. This can give us all a timeline view of the evolution of technology as a solution to the problems faced by many.

Literature Review

(Nishat Vasker et al, 2023, 1-2) The YOLOv5-based real-time self-harm detection system is a significant advancement in enhancing safety and well-being. It effectively identifies and addresses self-injurious behaviors, serving as a proactive tool for prevention and intervention. This technology allows individuals to assess their tendencies from the comfort of their homes, fostering prompt support and communication for early interventions.

(Victor E. De S. Silva et al, 2022, 1-3) The system excels in detecting physical violence within videos while ensuring user privacy and adhering to data protection regulations. Its performance is commendable, with improvements noted in F1 score, precision, and recall metrics, underscoring the successful application of federated learning techniques in recognizing instances of violence across both video and audio formats.

(H. Kishara Buddika Jayasanka et al, 2021, 2-5) Utilizing machine learning and image processing, the system locates violent events across videos, audio, text, and thumbnails to mitigate the adverse effects of violent content. Impressively, it achieves an accuracy rate of 83% in detecting hate speech in subtitles and reaches an accuracy of 92% in identifying violence in thumbnails.

(Vu Lam et al, 2013, 3-5) The research highlights the effectiveness of low-level audio and visual features in evaluating new algorithms for violent scene detection, with particular feature combinations significantly enhancing detection capabilities. Findings indicate that local features yield the most impactful results, and integrating various feature types boosts overall performance, culminating in the highest mean Average Precision (mAP) when using a blend of global, local, motion, and audio features.

(Tiago Lacerda, 2022, 1-2) In audio detection of physical violence, the study introduces a method grounded in self-supervised learning and deep learning principles, the HEAR dataset, which balances synthetic audio samples of violence and non-violence. MobileNet stands out, outperforming alternative models with impressive accuracy, F1 scores, and precision, attesting to the efficacy of this approach.

(Vikram Gupta et al, 2022, 2-4) ADIMA, a diverse audio dataset designed for recognizing content, comprises 10 Indic languages with 11,775 samples (equating to 65 hours of audio) and consists of 43.38% abusive and 56.62% non-abusive recordings. Findings reveal that Wav2Vec2 models excel in this domain, surpassing VGG and RNN models across most languages.

(Srijita Ghatak et al, 2023, 1-5) The system demonstrates the capability of monitoring fall incidents among the elderly, contributing to injury prevention and a better quality of life. It achieves strong results in identifying and categorizing fall postures, with over 90% accuracy, an F1 score of 92.5%, precision at 91.2%, and recall of 93.8% on the test set.

(Ibrar Ahmed et al, 2023, 1-4) The proposed architecture is adept at learning appearance and motion features to detect anomalies in various scene settings, employing a joint representations learning technique. The study concludes that object-tracking methods are critical components of video surveillance systems, with kernel and silhouette-tracking methods proving particularly efficient for object-tracking applications.

(Dara Ros et al, 2023, 1-6) The new framework offers a versatile and dependable approach to fall detection in various surveillance settings, effectively navigating different camera angles and configurations. It surpasses traditional methods such as CCTV feeds and web interfaces in numerous test datasets, achieving remarkable accuracy, sensitivity, and F1 scores, underscoring its effectiveness in detecting falls.

(S. Aarthi et al, 2021, 2-4) This framework emphasizes the importance of Human Activity Recognition (HAR) systems in supervising the elderly, highlighting the necessity for trustworthy systems that identify atypical behaviors or falls. It reviews various datasets connected to HAR, focusing on both sensor and video data, and presents publicly accessible datasets along with performance metrics assessed by existing methods, offering valuable insights for future investigations.

(Zhihao Chen et al, 2019, 1-5) The findings indicate that the proposed system successfully detects objects and estimates distances, with YOLO V3 emerging as the optimal choice for real-time applications. Additionally, integrating object detection with distance estimation enhances performance, while tracking objects in videos is achievable by predicting their new locations based on previous frames.

(Hiroaki Kingetsu, 2019, 1-3) In this study, a video-based machine learning system was developed to evaluate fall risk using a standard camera and OpenPose software. Results demonstrated that both SVM and logistic regression effectively predicted fall risks, whereas linear discriminant analysis incorrectly classified a "medium" risk as "none."

(Alex D. Edgcomb et al, 2012, 1-3) The study states that the fall detection algorithm maintains comparable accuracy when using privacy-enhanced videos instead of raw videos, allowing users to prioritize privacy without sacrificing performance. Privacy enhancements can be applied before streaming to local or remote devices, resulting in minimal impact on detection accuracy except in cases of blurred video.

(Jing Tian et al, 2020, 1-3) The study suggests that automated video analysis can reliably assess human seizure behavior, although additional research is necessary to overcome current challenges. Preliminary findings indicate that the proposed framework succeeded in detecting seizure events, highlighting its potential for monitoring seizure activities.

(Rizzah Grace Llanes et al, 2022, 1-3) Moreover, the study illustrates that stress detection through facial action units in videos is effective, with machine learning algorithms adeptly classifying stress levels. The Random Forest model emerged as the most efficient, performing exceptionally well with general and individual data. Its ability to manage intensity and presence features improves when trained on person-specific data.

(Tashreef Abdullah Araf et al, 2022, 1-4) The proposed system adeptly identifies facial emotions using cascade classifiers, with Grad-CAM enhancing explainability. This technology can be implemented in various fields, particularly emotional artificial intelligence.

(Muhammad Abdullah et al, 2020, 1-3) The proposed method suggests that a well-trained CNN followed by an RNN effectively accomplishes Video Facial Expression Recognition. The Xception Net attained an impressive 80% validation accuracy on the FER2013 dataset, whereas the single-layer LSTM network achieved a 65% accuracy rate.

(Xin Song et al, 2016, 2-4) The method suggested in this study shows great promise in recognizing facial expressions, boasting impressive speed and high recognition rates. It effectively addresses limitations in earlier research focused on specific human faces, achieving remarkable results.

(R. Kalaiselvi et al, Mar 2014, 1-4) The findings reveal that the proposed system is adept at identifying emotions through facial expressions in video sequences, showcasing a promising success rate. This technological advancement greatly enhances human-computer interaction.

(Mukesh Choubisa et al, 2023, 1-4) The study highlights the effectiveness of the Mean Shift Algorithm for real-time object tracking in video surveillance, improving accuracy and visibility in public spaces. Its successful implementation allows the tracking of objects in video frames, as evidenced by its ability to place bounding boxes around identified targets, reflecting its

reliability in monitoring dynamic environments.

(K. Ullah et al, 2019, 1-4) The research indicates that the CSRT algorithm excels in head tracking, reaching an accuracy rate of 85%. However, it notes that all algorithms struggle with full-body tracking, with the highest accuracy only hitting about 40%.

(Zijian Zhang et al, 2024, 2-5) The study also presents a novel video analysis method for detecting and classifying hand tremors in Parkinson's patients, yielding reliable data to enhance patient care. It demonstrates high accuracy and precision, outpacing current technologies in evaluating the severity of tremors, indicating a significant improvement in tremor assessment.

(Kuhelee Roy et al, 2013, 1-5) The results are supported by a confusion matrix and a precision-recall graph, with ANOVA analysis revealing a notable difference between tremor and non-tremor classes. The method successfully leverages optical flow and supports vector machines to detect hand tremors in videos and classifies tremor videos.

Application of Activity Detection AI

Activity Detection AI has been used in various social or professional settings. In care centers, staff can use it to monitor the health of patients/residents. Administrators of colleges and community centers can use it to monitor violence, arguments, and group gatherings. Authorized people in offices and homes can use it for intruder alerts and monitor restricted areas.

The applications for Activity Detection can be endless. Listing the following possibilities.

1. Old age homes, Care centers
2. Colleges and community centers
3. Offices and homes
4. Banks
5. Shopping malls, supermarkets, airports
6. High security or restricted areas

7. Forests and wildlife
8. Road/traffic in cities
9. Autonomous Automobiles
10. Social, community functions and gatherings

Activity Detection In Old Age Homes And Care Centers

This is a classic use case. Traditionally people in old age homes and care centres are looked after by attendants in person and/or using CCTV or similar mechanisms. But with the dawn of new age technologies, Activity detection can easily help identify various incidents related to health viz. somebody is having seizures(Jing Tian et al, 2020, 1-3), tremors(Zijian Zhang et al, 2024, 2-5), (Kuhelee Roy et al, 2013, 1-5), accidents, etc.

Basic activities like walking, sitting in a chair, getting up, and doing chores, the many things we take for granted are often points of struggle for the physically challenged. With proper training, AI can detect if a person is falling while trying to sit in a chair or wheelchair and the elaborate solution will raise alarms and make announcements to get help.

Activity detection can also be used to detect the absence of activity. This is a crucial use case for people in old age homes or care centers. The Activity Detection AI solution may not be able to detect the cause in all cases but can highlight the lack of movement by a subject thus ensuring someone can check on them and confirm all is okay. Further, these AI models can also be easily trained for heart attack behaviors or certain sign languages, which can get help along the lines of SOS messages.

The best part of it can be the ability of the End-to-end solution using the Activity Detection AI core, to inform near and dear ones or other healthcare officials or centers immediately in the time of need in real-time. This can be a game changer as imagine an ambulance getting called automatically when someone falls (Hiroaki Kingetsu, 2019, 1-3) due to symptoms similar to a heart attack. Every minute can be a reason for saving a valuable life. Technology cannot be put to a better use than this.

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

Activity Detection in Colleges and Community Centers

It is a rising challenge to manage teens, young adults, and anyone regardless of age and gender who is out there looking for trouble.

The environment in college and community centers requires people to follow rules and regulations in letter and spirit and do their best to maintain the decorum of the facility.

Activity detection can be used to detect early signs of fights or violent arguments breaking out (Victor E. De S. Silva et al, 2022, 1-3), (H. Kishara Buddika Jayasanka et al, 2021, 1-5) in time not only to just detect but to avoid significant issues by raising alerts and taking predefined actions (notifications). It can detect any attempt to damage the property or presence of people in restricted areas.

One thing to note is that health incidents can also be monitored. Activity detection AI can easily identify incidents, and report to officials for help. Looking after people is part of a day's work for an AI model built with passion and hard work.

Facial recognition will flag people and send alarms or warnings to their mobiles. This may be the biggest deterrent to avoid any accidents. Habitual offenders can be proactively monitored and tracked to ensure they are not in a position to negatively affect the decorum of the place.

Activity Detection AI can be trained with relevant datasets (images and videos) to accommodate group behavior. Simple technology can do monumental work if used intelligently.

Activity Detection In Offices And Homes

These environments are meant to be the safest and AI can help keep it that way. The simplest case here for AI is to detect intruders, monitor health, and flag anything that violates the decorum.

It may be unlikely that people will deploy an Activity Monitoring AI solution at home. Authorized people can turn it on when they leave for the office, leaving behind children or senior citizens.

This will put household or delivery personnel in check as their inappropriate behavior (Victor E. De S. Silva et al, 2022, 1-3), (H. Kishara Buddika Jayasanka et al, 2021, 1-5), (Tiago Lacerda, 2022, 1-2), (Vikram Gupta et al, 2022, 2-4) viz. fights, arguments, property damage, stealing, abuse, etc. can easily be detected and reported. Health incident monitoring will remain a key offering in this setup.

Office admin staff can use it to check who is coming and going when they detect the presence of authorized personnel only in restricted areas. They can also check for damage to office property and look for health incidents and abuse (Victor E. De S. Silva et al, 2022, 1-3).

Administrators can use the Activity detection model to monitor meetings in boardrooms, cafeterias, and shared areas to detect the engagement of participants or customers (Tashreef Abdullah Araf et al, 2022, 1-4), (Muhammad Abdullah et al, 2020, 1-3). Traditional solutions lack the intelligence part, they otherwise just record hours and hours of videos. Typically, complacent staff monitor and later look at these feeds post facto i.e., after something troublesome has happened. But AI has the power to deliver intelligence in real time and importantly help avoid negative incidences altogether. Thus, making the world a better place for everyone.

Further, we can train this model for sentiment analysis to check employees' well-being as they check into the office. In recent times, working people are facing stress (Rizzah Grace Llanes et al, 2022, 1-3) and do not know how to manage it, only to lead to fatality (Nishat Vasker et al, 2023, 2-6).

Smart Sentinel using Activity Detection AI

The analysis has confirmed the belief that AI is here to stay and is the undisputed future of humankind. Activity Detection has a far-reaching impact and numerous useful applications in day-to-day life. The key element and enabler that has made this possible is Computer vision. It

has given computers the ability to see, interpret, and understand. The journey has been a long one though.

It is interesting to see how we ended up here. Let's start from the beginning.

Past

The desire and need to detect peoples' activities is not a new phenomenon. As always, we used the means as and when available. Traditionally human resources were abundant, cheap, and reliable for the job.

So we had people watching people checking health incidents and intruders. There was never an issue. But interestingly, the volume and spread of the human population were also very limited.

With the rapid explosion of technology, new tools came to aid. Portable video cameras became cheap and convenient. This enabled CCTV (Closed Circuit Television), which allowed people to monitor and record (evidence) of activities of people. In fact, to date, these solutions are used in every place imaginable and it was a de facto standard for ensuring the security, and safety of people, property, and premises.

As time has passed, the human population has just multiplied, and our way of living has changed significantly. People travel a lot of work, leaving behind kind and elderly people alone at home. There is no community or joint families to look after them. Old age homes and care centers have thus spawned in every city and state.

The need of the hour is to have a solution that can track activities and do face detection, but it may still use man in the middle, meaning a person needs to review to make sense, monitor, and act.

Present

There is no time like the present. Technology is evolving so fast that any implemented solution can be treated as obsolete within months if not days. The big tech firms are pushing the envelope and making breakthroughs in AI viz. ChatGPT, Gemini, and so on.

The open-source community has also matured and sharing production-ready libraries for all needs. Generative AI is taking people on a rollercoaster ride of computer-based creativity.

Here are a few of the open-source frameworks being used for activity detection. They vary in resource usage, ease of integration, and precision.

- **YOLO (You Only Look Once)**

YOLO is an advanced object detection model designed for real-time image processing. It efficiently predicts bounding boxes and class probabilities all in one go, which leads to impressive speed in inference times. This makes YOLO particularly well-suited for time-sensitive tasks, like video analysis. The model has seen multiple iterations. Each latest version has enhanced its accuracy and overall performance.

- **MobileNets**

MobileNets are convolutional neural networks tailored for mobile and edge devices. They utilize depth-wise separable convolutions to minimize the parameters and computations needed, delivering impressive efficiency while maintaining reasonable accuracy. This makes MobileNets perfect for applications that operate in resource-limited environments, such as smartphones and embedded systems.

- **ResNet (Residual Network)**

ResNet represents a breakthrough in deep learning architecture by introducing residual connections. These connections facilitate the smooth flow of gradients through very deep networks, allowing the training of models with hundreds of layers while preventing issues like vanishing gradients. Its outstanding performance on various image classification tasks has established ResNet as a cornerstone model in computer vision projects.

- **Simple CNN**

A Simple CNN is a foundational convolutional neural network commonly used for educational purposes or early-stage projects. This architecture typically comprises a sequence of

convolutional layers, followed by pooling and fully connected layers. While it may not reach the high-level accuracy of more sophisticated models, it is a great starting point for grasping the essential concepts of convolutional neural networks.

- **VGG (Visual Geometry Group)**

Lastly, VGG is a well-known deep-learning architecture known for its straightforwardness and effectiveness in image classification. It features a consistent structure, employing small (3x3) convolutional filters arranged in layers, which allows it to capture intricate patterns within images. Although VGG tends to be more computationally demanding than other models, its high accuracy on benchmark datasets makes it a widely favored option for transfer learning in various applications.

Model	Type	Main Use Case	Architecture	Speed	Accuracy
YOLO	Object Detection	Real-time object detection	Single neural network	Very fast	Moderate to high
MobileNets	Image Classification	Mobile and edge devices	Depthwise separable convolutions	Fast	Moderate to high
ResNet	Image Classification	General-purpose tasks	Residual connections (skip connections)	Moderate	High
Simple CNN	Image Classification	Basic tasks, educational	Basic layers (convolution, pooling)	Fast to moderate	Variable
VGG	Image Classification	General-purpose tasks	Deep architecture with small filters	Moderate to slow	High

Fig. 1. Opensource AI framework comparison

Using these frameworks, anyone can easily build an Activity Detection AI model. The present is leading us to a bright future.

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

Future

The future is undisputedly bright for technology and AI. More and more frameworks are coming into play every day. With time need for human intervention and/or human review may no longer be required. AI solutions will slowly become autonomous and can learn, adapt, and get better. The only thing that can hold us back now is our imagination.

Driverless cars are almost here and AI-based robots are becoming a reality. So when we talk about activity Detection, computer vision will get better and there will be more actions available to detect and we can take corrective measures using robots rather than just be restricted to sending notifications and alerts to people and systems.

There will always be a thin line deciding the morality of the solution. But like any other tool or weapon, if it is used correctly, it can help. We can ensure that new-age AI solutions are ethical and friendly by performing frequent data security audits and implementing other measures as required.

Various bodies and regulations are here to control AI and ensure ethical practices by the firms behind the technology. A good governance framework can help ensure people's interests, and privacy are safeguarded at all times and at all costs.

Comparative Analysis

The comparative analysis lists the various solutions and highlights their offerings. We have analyzed sample research papers. Table 1 shows different solutions, methods, and technology to achieve the desired results, and their strengths and weaknesses.

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

Table 1: Comparative Analysis of Activity Detection Research Papers

References	Proposed Technique	Year	Method/Technology	Strengths	Weaknesses
Nishat Vasker et al [1]	A Real-Time Self-Harm Detection Ensuring Safety in Every Moment	2023	1. It uses YOLOv5, a state-of-the-art object detection technology for identifying and classifying self-harm behaviors. 2. Advanced computer vision technologies enable the system to analyze photographs and videos in real time, ensuring accurate detection with high recall and precision. 3. The YOLOv5 model was trained on a dataset containing images of self-harm and normal harm and fine-tuned using transfer learning.	- Real-time detection for immediate intervention - High accuracy with YOLOv5 - Scalable and remote monitoring - Customizable via transfer learning	- Privacy concerns - Dataset limitations - Requires high computational resources
Victor E. De S. Silva et al [2]	Federated Learning for Physical Violence Detection in Videos	2022	1. It uses a client-server setup where anonymized training data is sent to a central model that combines and redistributes it. 2. It is tested on the AIRTLab Dataset using pre-trained CNN for classifying images.	- Privacy-preserving - High performance (F1 score, precision, recall) - Collaborative learning without data sharing - Effective for	- Communication overhead - Data inconsistency across clients - High computational demands

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

References	Proposed Technique	Year	Method/Technology	Strengths	Weaknesses
			3. The study uses four CNN architectures—Inception v3, MobileNet v2, ResNet-152 v2, and VGG-16—within a Federated Learning framework with three clients over 40 rounds, comparing the results to traditional deep learning methods.	video and audio detection	
H. Kishara Buddika Jayasanka et al [3]	Intelligent Violence Video Detection System	2021	1. It uses public and custom datasets to find violence in videos, audio, text, and thumbnails. 2. It uses machine learning methods GloVe for text classification and Keras and TensorFlow to build and train models. 3. The research aims to improve accuracy by processing different parts of the content, not just the visuals. It combines video, audio, text, and thumbnail detection.	- Multi-modal detection (video, audio, text, thumbnails) - High accuracy (83% for hate speech, 92% for violence in thumbnails) - Versatile machine learning methods (GloVe, Keras, TensorFlow) - Comprehensive content analysis for improved accuracy	- High system complexity - Resource-intensive for processing multiple data types - Potential data quality inconsistencies across modalities

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

References	Proposed Technique	Year	Method/Technology	Strengths	Weaknesses
Vu Lam et al [4]	Evaluation of low-level features for detecting violent scenes in videos	2013	1. It uses the KAORI-SECODE framework to extract different visual features, like global, local, motion, and audio from videos. 2. Keyframes are used to create shot-based features with max and average pooling, and the study looks at 48 different feature combinations. 3. The "Support vector machine (SVM)" classifier with a linear kernel is used for audio and motion features, and the best parameters are found through grid search and 5-fold cross-validation.	- Effective use of low-level audio and visual features for detection - Highest mean Average Precision (mAP) achieved with feature fusion - Comprehensive analysis of 48 feature combinations - Utilizes SVM classifier with optimized parameters for improved performance	- Complexity in feature extraction and combination - Resource-intensive process for extracting multiple feature types - Performance may vary depending on dataset quality and variability
Tiago Lacerda [5]	Detection of physical violence through audio	2022	1. The study will use a systematic review approach with Forward Snowballing to update prior research on audio violence detection. 2. It will focus on mel-spectrograms and deep learning models, adapting Self-Supervised Learning for audio	- High accuracy in detecting physical violence in audio - Utilizes the HEAR dataset with balanced samples - Effective use of MobileNet for improved performance	Dependence on the quality of the HEAR dataset Complexity in implementing Self-Supervised Learning for audio Resource-intensive due to deep learning

References	Proposed Technique	Year	Method/Technology	Strengths	Weaknesses
			applications. 3. Proposed methods will incorporate temporal information into the neural network and apply data augmentation techniques like time-shifting and speed changes.	(F1 score, precision) - Incorporates temporal information and data augmentation techniques	model requirements
Vikram Gupta et al [6]	ADIMA: Abuse Detection in Multilingual Audio	2022	1. Audio recordings were taken from public chatrooms on ShareChat, with three people labeling each as abusive or not. 2. VGG and Wav2Vec2 models extracted features, and a classifier sorted the recordings into abusive or non-abusive categories. 3. The method used audio features like pitch, emotions, and intensity, combining data from all languages for separate training and evaluation.	- Diverse dataset (10 Indic languages, 11,775 samples) - High performance with Wav2Vec2 models - Comprehensive feature extraction (pitch, emotions, intensity) - Community-driven labeling for reliability	- Potential dataset bias from public chatrooms - Resource-intensive model requirements - Language variability affecting performance
Srijita Ghatak et al [7]	A Simple Fall Detection Scheme for Early Detection of	2023	1. The proposed system includes four main components Frame Extraction, Person Detection, Pose Estimation with	- High accuracy (over 90%) - Reliable in identifying and categorizing	- Complex system integration - Performance affected by visual quality

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

References	Proposed Technique	Year	Method/Technology	Strengths	Weaknesses
	Falls in Elderly People		Feature Extraction, and a Decision Tree-based Classifier. 2. It generates Active Energy Images (AEIs) from subject silhouettes to capture dynamic gait variations. 3. Affine Moment Invariants (AMIs) are used as descriptors to extract relevant features for analysis.	fall postures - Comprehensive system architecture - Dynamic gait analysis with Active Energy Images (AEIs)	- Limited to visible falls only
Ibrar Ahmed et al [8]	Object Motion Tracking and Detection in Surveillance Videos Using Resnet Architecture	2023	1. The proposed architecture utilizes ResNet-inspired encoders to extract visual and motion data for effective anomaly detection. 2. The study employs point, kernel, silhouette tracking, object recognition techniques background removal, and optical flow. 3. Key findings highlight the effectiveness of kernel and silhouette tracking in video surveillance and the role of IP camera data for early	- Efficient anomaly detection using joint representation learning - Utilizes ResNet-inspired encoders for visual and motion data extraction - Effective tracking methods (point, kernel, silhouette) enhance object tracking - Emphasizes the importance of IP camera	- Complexity of the proposed architecture may hinder implementation - Reliance on high-quality video data for accurate tracking - Potential challenges with diverse scene settings and lighting conditions

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

References	Proposed Technique	Year	Method/Technology	Strengths	Weaknesses
			warnings and danger prevention.	data for early warnings	
Dara Ros et al [9]	A Flexible Fall Detection Framework Based on Object Detection and Motion Analysis	2023	1. It uses object detection and motion analysis to identify falls in surveillance videos by tracking human objects over time. 2. It allows integration of any pre-trained object detection method and works in single-view and multi-view camera setups. 3. It extracts key features, such as orientation change rate and silhouette area, to train a RUSBoosted tree ensemble classifier for fall detection.	- Flexible and reliable for various surveillance environments - Superior performance compared to traditional methods (CCTV-feed, webUI) - Integrates any pre-trained object detection method - Works with both single-view and multi-view camera setups - Utilizes key features (orientation change rate, silhouette area) for enhanced detection	- Complexity in the integration of multiple detection methods - Dependency on the quality of surveillance footage - May require extensive training data for effective model performance
S. Aarthi et al [10]	A Comprehensive Study on Human Activity Recognition	2021	1. The review analyses state-of-the-art Human Activity Recognition (HAR) methods, focusing on vision-based and sensor-based	- Highlights the importance of HAR systems for elderly monitoring - Summarizes	- May not cover all emerging HAR methods and technologies - Performance metrics may vary across different

References	Proposed Technique	Year	Method/Technology	Strengths	Weaknesses
			approaches and their performance metrics. 2. It highlights the importance of HAR systems for monitoring the elderly, particularly in fall detection and unusual behavior. 3. The study emphasizes the effectiveness of vision-based techniques for activity classification and sensor methods for minimizing false alarms and improving accuracy.	various datasets for HAR, aiding future research - Analyses state-of-the-art vision-based and sensor-based methods - Emphasizes vision-based techniques for classification and sensor methods for accuracy	studies and datasets - Reliance on the quality of datasets for accurate system evaluation
Zhihao Chen et al [11]	Real-Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility	2019	1. The study employs two deep learning approaches, YOLO V3, and SSD, for object detection and tracking. 2. The study combines object detection and distance estimation for tracking objects in video sequences. 3. A modified SSD algorithm is utilized to analyze the behavior of objects such as pedestrians and vehicles.	- Effective real-time object detection and distance estimation - YOLO V3 is optimal for real-time applications - Combining object detection and distance estimation improves the results - Predictive tracking based on previous	- Dependency on high-quality video input for accurate detection - Potential challenges with varying environmental conditions (lighting, occlusions) - Resource-intensive due to the deep learning models employed

References	Proposed Technique	Year	Method/Technology	Strengths	Weaknesses
				frames enhances object tracking accuracy	
Hiroaki Kingetsu [12]	Video-based Fall Risk Detection System for the Elderly	2019	1. It used a video camera to record subjects performing a 5-meter walk test, with OpenPose software detecting 24 key points on the body and hands for feature extraction. 2. The extracted feature data were used for machine learning, and SVM and logistic regression accurately predicted fall risks.	- Uses general camera and OpenPose for easy implementation - Accurate predictions with SVM and logistic regression - Effective feature extraction from 24 key body points - Simple fall risk assessment for the elderly	Misclassification by linear discriminant analysis Performance affected by camera quality and environment Requires feasible video recording conditions
Alex D. Edgcomb et al [13]	Automated fall detection on privacy-enhanced video	2012	1. It uses a foreground-background segmentation algorithm to extract a minimum bounding rectangle (MBR) for fall detection. 2. Time series shapelet analysis is applied to the MBR dimensions, and training is conducted using logical-shapelet	- Comparable accuracy with privacy-enhanced and raw videos - Prioritizes privacy without sacrificing performance - Effective foreground-background segmentation - Robust detection using	- Blurred video may reduce accuracy - Complexity in processing video enhancements - Requires thorough labeling and training for classification

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

References	Proposed Technique	Year	Method/Technology	Strengths	Weaknesses
			analysis software and evaluated via leave-one-out cross-validation. 3. It involves converting raw videos to privacy-enhanced versions, labeling MBR data, training a binary classifier, and testing it on both video types.	time series shapelet analysis	
Jing Tian et al [14]	Automated Analysis of Seizure Behavior in Video: Methods and Challenges	2020	1. It introduces a machine learning-based framework for automated seizure analysis, validated through a literature survey and deployed in a hospital. 2. It is very effective in detecting seizures, the significance of machine learning in seizure motion modeling, and the need for further research. 3. Preliminary trial results indicate the framework's potential to improve diagnostic precision in multi-modal seizure behavior analysis.	- Accurate seizure behavior measurement - Successful seizure detection framework - Highlights machine learning's importance - Potential to enhance diagnostic precision	- Requires further research - Preliminary results may not reflect long-term effectiveness - Detection accuracy depends on video quality

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

References	Proposed Technique	Year	Method/Technology	Strengths	Weaknesses
Rizzah Grace Llanes et al [15]	Stress Detection in Video Feed: Utilizing Facial Action Units as Indicators in Various Machine Learning Algorithms	2022	1. It recorded participants during an arithmetic test in a timed, competitive setting, validating stress levels through their self-assessment. 2. Facial action units were extracted using the OpenFace 2.0 interface to analyze expressions. 3. It used Python Scikit-learn to process data and classify stress levels with multiple linear regression, support vector machines, and random forest models.	- Real-time recognition of facial emotions - Uses Cascade Classifier for effective emotion detection - Grad-CAM enhances model explainability - Diverse dataset with over 1,000 images across gender and age groups	- Performance may vary with different lighting conditions - Cascade Classifiers may struggle with occlusions - Dependency on dataset quality for accurate detection
Tashreef Abdullah Araf et al [16]	Real-Time Face Emotion Recognition and Visualization using Grad-CAM	2022	1. It uses a Cascade Classifier for emotion recognition and Grad-CAM for visualizing model detection. 2. The dataset includes over 1,000 image expressions from diverse gender and age groups, sourced from an online machine learning repository. 3. It is validated with test data before real-time inspection, with Grad-CAM	- Real-time facial emotion recognition - Cascade Classifier enables effective detection - Grad-CAM enhances model explainability and visualization - Diverse dataset with over 1,000 images across	- Performance may vary under different lighting conditions - Cascade Classifiers can struggle with occlusions - Quality of dataset impacts detection accuracy

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

References	Proposed Technique	Year	Method/Technology	Strengths	Weaknesses
			providing insights into the emotion detection process.	gender and age groups	
Muhammad Abdullah et al [17]	Facial Expression Recognition in Videos: An CNN-LSTM based Model for Video Classification	2020	1. Frames are pre-processed to detect faces, which includes cropping and resizing the detected faces. 2. The pre-processed frames are fed into a Convolutional Neural Network (CNN) to extract emotional features. 3. The output from the CNN is then passed to a Recurrent Neural Network (RNN) to learn temporal features.	- Effective CNN-RNN approach for facial expression recognition - Xception Net achieved 80% accuracy on the FER2013 dataset - Includes preprocessing for face detection - RNN captures temporal features for enhanced performance	- Single-layer LSTM only achieved 65% accuracy - Performance affected by video quality and lighting - Computationally intensive due to model complexity
Xin Song et al [18]	Facial expression recognition based on video	2016	1. The method extracts 2D feature points from video frames using Bezier curves and connects them to create temporal characteristic curves. 2. It combines facial region segmentation with Bezier curve-based feature extraction to automatically fit facial contours.	- High speed and recognition rates - Improves upon previous facial recognition methods - Uses Bezier curves for feature extraction and temporal analysis - Combines	- Complex implementation of Bezier curves and fitting - Performance affected by video quality and lighting - Requires significant computational resources for

References	Proposed Technique	Year	Method/Technology	Strengths	Weaknesses
			3. The study also proposes nonlinear function fitting to extract time sequence features from the video stream.	segmentation with contour fitting for accuracy	real-time analysis
R. Kalaiselvi et al [19]	Automatic emotion recognition in video	2014	1. The study employs video segmentation, facial feature tracking, Action Unit recognition, and Dynamic Bayesian Networks for emotion recognition in video sequences. 2. Classification selects Action Units to maximize the likelihood of extracted facial features with Hidden Markov Models (HMMs). 3. The system's accuracy across various scenarios shows favorable comparisons to previous expression recognition methods.	- Effectively recognizes emotions from facial expressions - Promising success rate in tasks Enhances human-computer interaction - Utilizes advanced recognition techniques	- Complex implementation - Varying accuracy across scenarios - High computational resource demands
Mukesh Choubisa et al [20]	Object Tracking in Intelligent Video Surveillance System Based	2023	1. The study uses the Mean Shift Algorithm (MSA) for object tracking, leveraging frame difference principles	- Real-time object tracking with the Mean Shift Algorithm - Improved	- Performance may decline with occlusions or rapid movements - Frame difference

References	Proposed Technique	Year	Method/Technology	Strengths	Weaknesses
	on Artificial System		and digital signal processing to detect moving objects. 2. MSA offers strong computational capabilities for effective object tracking and detection in video surveillance. 3. The algorithm enables body tracking without losing frames, showcasing precision by overlaying functions on identified and user-defined objects.	accuracy and visibility in surveillance - Detects and tracks objects using bounding boxes - Strong computational capabilities for dynamic settings	reliance may limit effectiveness - Requires careful parameter tuning for optimal results
K. Ullah et al [21]	Comparison of Person Tracking Algorithms Using Overhead View Implemented in OpenCV	2019	1. It implements seven different tracking algorithms in OpenCV, including BOOSTING, MIL, KCF, TLD, MEDIANFLOW, MOSSE, and CSRT. 2. These algorithms are compared using the IMS-PT dataset, a newly created overhead view dataset for object tracking. 3. The algorithms are evaluated using the Jaccard similarity coefficient method,	- CSRT achieved 85% accuracy in head tracking - Compares seven algorithms in OpenCV - Uses IMS-PT dataset for evaluation - Performance assessed with Jaccard similarity coefficient	- Full body tracking max accuracy is 40% - Varying effectiveness among algorithms - Complex implementation due to multiple methods

References	Proposed Technique	Year	Method/Technology	Strengths	Weaknesses
			and the implementation is done in Python 3.6 using the OpenCV library.		
Zijian Zhang, Chuangye Wang et al [22]	The automatic detection method for hand tremors in Parkinson's' disease based on video analysis	2024	1. It identifies the position of the hand and cropping localized videos. 2. Uses hand pose estimation and a three-frame difference algorithm to extract key features from the hand region, combining frequency features with image pixel change features. 3. It trains classifiers, such as Logistic Regression and Random Forest, to classify tremors and their severity.	- High accuracy in detecting Parkinson's hand tremors - Outperforms existing assessment technologies - Provides reliable data for patient care	- Complexity in segmenting and processing video frames - Performance may vary based on video quality and lighting - Requires careful tuning of SVM parameters for optimal results
Kuhelee Roy et al [23]	A learning-based approach for tremor detection from videos	2013	1. The method segments video into five-frame sets and extracts features using optical flow and joint entropy for tremor classification. 2. Key steps include skin segmentation, optical flow computation, feature extraction, SVM	- Effectively detects hand tremors in videos - Uses optical flow and SVM for classification - Performance evaluation with confusion matrix and	- Complex video segmentation and processing - Performance may vary with video quality and lighting - Requires careful SVM parameter tuning

References	Proposed Technique	Year	Method/Technology	Strengths	Weaknesses
			training, and leave-one-out cross-validation testing.	precision-recall graph	

Conclusion

Activity Detection is critical in numerous solutions that can greatly enhance our everyday experiences. Advances in technology have not only broadened our horizons but also paved new paths for us to explore. At the heart of it all is a commitment to improving people's well-being through the thoughtful application of technology.

The simplicity of implementation, quick and easy deployment, and ease of use by non-technical people can lead to faster adaptation of this emerging tech trend in various areas of interest viz. old age living, daycares, community centers, sensitive (office) premises, and so on.

The latest trends in robotics and driverless cars are using these core solutions in some way or other. When implemented correctly this solution will have the ability to enrich our lives (period).

Acknowledgment

Our sincere thanks to the Department of Polytechnic, Dr. Vishwanath Karad's MIT World Peace University, for their support and encouragement without which this paper would not have been possible.

Conflict of Interest

The authors declare no competing interests to 'Smart Sentinel - Activity Detection AI."

References

- Nishat Vasker, M. Hasan (2023), Methodology. In Real-Time Self-Harm Detection Ensuring Safety in Every Moment (pp. 2-6), Bangladesh.
- Victor E. De S. Silva, Tiago Lacerda, Péricles B. C. Miranda, André C. A. Nascimento, Ana Paula C. Furtado (2022), Introduction. In Federated Learning for Physical Violence Detection in Videos (pp. 1-3), Brazil.
- H. Kishara Buddika Jayasanka, B. M. Rashitha Dileeshan Batugedara, K. D. C. D. Ruwan Diyunuge, T. H. D. Yehan Malith Jayasekara, Dilani Lunugalage, Udara Srimath S. Samaratunge Arachchilllage (2021), Methodology. In Intelligent Violence Video Detection System (pp. 2-5), Sri Lanka.
- Vu Lam, Duy-Dinh Le, Sang Phan Le, S. Satoh, Due Arm Duong, T. Ngo (2013), Dataset and Evaluation framework. In Evaluation of low-level features for detecting violent video scenes (pp. 3-5), Vietnam.
- Tiago Lacerda (2023), State of Art and Theoretical Background. In Detection of physical violence through audio (pp. 1-2), Brazil.
- Vikram Gupta, Rini A. Sharon, Ramit Sawhney, Debdoot Mukherjee (2022), Related Work and Formulation and Methodology. In ADIMA: Abuse Detection In Multilingual Audio (pp. 2-4), India.
- Srijita Ghatak, Rishita Mitra, Washef Ahmed, Kunal Chanda (2023), Contributions and Results. In A Simple Fall Detection Scheme for Early Detection of Falls in Elderly People (pp. 1-5), Kolkata
- Ibrar Ahmed, Bharat Bhushan Naib (2023), Related Work, Proposed Work and Results & Discussion. In Object Motion Tracking and Detection in Surveillance Videos using Resnet Architecture (pp. 1-4), India.
- Dara Ros, Rui Dai (2023), Introduction and Performance Evaluation. In A Flexible Fall Detection Framework Based on Object Detection and Motion Analysis (pp. 1-6), Ohio.

- S. Aarthi, S. Juliet (2021), Literature Review. In A Comprehensive Study on Human Activity Recognition (pp. 2-4), India.
- Zhihao Chen, R. Khemmar, B. Decoux, A. Atahouet, J. Ertaud (2019), Introduction, Related Work and Object Distance Estimation. In Real-Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility (pp. 1-5), France.
- Hiroaki Kingetsu, Takeshi Konno, Shuji Awai, D. Fukuda, T. Sonoda (2019), Introduction. In Video-based Fall Risk Detection System for the Elderly (pp. 1-3), Japan.
- Alex D. Edgcomb, F. Vahid (2012), Introduction and Fall Detection on Raw Video. In Automated fall detection on privacy-enhanced video (pp. 1-3), California.
- Jing Tian, Weiyu Yu, Jinquan Chen, Junke Lin, Mingfeng Wen, Yingxin Li, Jianxin Zhong, Keqiang Chen, Xuchu Feng (2020), Literature Survey. In Automated Analysis of Seizure Behavior in Video: Methods and Challenges (pp. 1-3), China.
- Rizzah Grace Llanes, Rosula S. J. Reyes (2022), Methodology. In Stress Detection in Video Feed: Utilizing Facial Action Units as Indicators in Various Machine Learning Algorithms (pp. 1-3), Philippines.
- Tashreef Abdullah Araf, A. Siddika, Sadullah Karimi, Md. Golam Rabiul Alam (2022), Literature Review. In Real-Time Face Emotion Recognition and Visualization using Grad-CAM (pp. 1-4), Bangladesh.
- Muhammad Abdullah, Mobeen Ahmad, Dongil Han (2020), Related Wordk and Dataset. In Facial Expression Recognition in Videos: An CNN-LSTM based Model for Video Classification (pp. 1-3), South Korea.
- Xin Song, Hong Bao (2016), Facial expression based on video and Modeling & Classification. In Facial expression recognition based on video (pp 2-4), China.

- R. Kalaiselvi, P. Kavitha, K. L. Shunmuganathan.(2014), Literature survey and architecture. In Automatic emotion recognition in video (pp. 1-4), Chennai.
- Mukesh Choubisa, Vijay Kumar, Mukesh Kumar, Dr. Samrat Khanna (2023), Introduction and experimental methodology. In Object Tracking in Intelligent Video Surveillance System Based on Artificial System (pp. 1-4), Gujrat.
- K. Ullah, Imran Ahmed, Misbah Ahmad, I. Khan (2019), Related work, dataset and results. In Comparison of Person Tracking Algorithms Using Overhead View Implemented in OpenCV (pp. 1-4), Pakistan.
- Zijian Zhang, Chuangye Wang, Ping Liang, Yijing Guo, Hao Gao (2024), Materials & methodology and experiments & results. In The automatic detection method for hand tremors in Parkinson's disease based on video analysis (pp. 2-5), China.
- Kuhelee Roy (2013), Feature detection using SURF, support vector machine and experimental results. In A learning-based approach for tremor detection from videos (pp. 1-5), India.