Emerging Trends in Digital Transformation
Volume-1, Issue-2, July 2025 ISSN:**3107-4308** *Paper ID:* ***ETDT-V1-I2-26***

ENHANCED HUMAN ACTIVITY RECOGNITION FOR SECURITY SURVEILLANCE USING DEEP NEURAL NETWORKS

#1PADALA DEEKSHANA,
MCA Student, Dept of MCA,
#2SARITHA PALLE,
Assistant Professor, Department of MCA,
VAAGESWARI COLLEGE OF ENGINEERING (AUTONOMOUS),
KARIMNAGAR, TG.

ABSTRACT: Motion detection in live video streams is considerably simpler when deep learning is used. Security monitoring is improved as a result. In this research, we investigate how different spatial and temporal features are analyzed by Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) models, to look for anomalies in surveillance footage. Deep learning-based activity recognition frequently performs better than more traditional machine learning techniques. As a result, security systems become more reliable and responsive. The ability of this approach to make correct behavior judgments could revolutionize automated tracking and public safety by enabling more proactive and efficient surveillance.

Keywords*: Human Activity Recognition (HAR), Deep Neural Networks (DNNs), Security Surveillance, Convolutional Neural Networks (CNNs) and Abnormal Behavior Detection.*

1. INTRODUCTION

Human Activity Recognition (HAR) is a component of contemporary security systems that ensures safety for individuals. Conventional monitoring systems dependent on human oversight for sensitive locations, critical infrastructure, and densely populated areas have grown obsolete and susceptible to errors as the demand for surveillance in these regions increases. The objective of HAR is to automate behavioral analysis through video footage to precisely document actions such as walking, running, fighting, and falling. We require cutting-edge technology capable of swiftly and accurately processing the vast quantities of data generated by real-time video tracking.

Deep Neural Networks and other advanced AI technologies have fundamentally transformed Human Activity Recognition (HAR). Convolutional neural networks (CNNs) and long short-term memory (LSTM) networks are two deep learning techniques adept in identifying patterns in video frames across spatial and temporal dimensions. The most effective method to analyze the evolution of human behavior over time is through Convolutional Neural Networks (CNNs), which can extract broad visual information from individual frames. In the context of temporal connections, long short-term memories (LSTMs) operate distinctly. Large datasets are required to train these models. Consequently, they may discern intricate behavioral patterns that conventional computer vision techniques overlook.

The capability of a security camera to detect human movement is crucial for various reasons, including crowd management, situational response, and crime prevention. For example, if there is an abrupt decline in activity in a public place or if an individual is lingering in a prohibited location, prompt action may be necessary to safeguard individuals and assets. Smart city infrastructure with DNN-enhanced augmented HAR systems enables security personnel to boost decision-making, optimize predictive analytics, and get automated alarms. Scalability and adaptability are essential components of these systems, as they continuously adapt and evolve to their environment.

HAR appears promising; nonetheless, it must overcome several challenges before it can be integrated with deep neural networks. Challenges encompass the necessity for extensive labeled datasets, the difficulty of comprehending data in real time, and the challenge of distinguishing between visually similar behaviors. Variations in camera angle, illumination, and obstructions complicate the precise recognition of movement.

DOI : https://doi.org/10.5281/zenodo.16900377

Recent advancements have addressed these issues through improved designs, the incorporation of transfer learning, and the integration of synthetic data with actual data. Following the implementation of these modifications, HAR systems have significantly enhanced their accuracy and efficacy in practical monitoring scenarios.

The application of deep neural networks for human behavior recognition has significantly enhanced security monitoring. This and similar emerging technologies enhance the precision and dependability of motion detection while minimizing the necessity for human oversight. With advancements in technology, we anticipate that HAR systems will ultimately enhance their decision-making capabilities, achieve greater autonomy, and possess a deeper understanding of their environment. This establishes a foundation for a future in which monitoring transcends mere observation to ensure individuals' safety and security in an increasingly challenging world. It involves the capacity for rapid cognition and profound comprehension.

2. REVIEW OF LITERATURE

Wang, J., Zhou, F., & Liu, L. (2020). The research's use of a deep convolutional neural network (CNN) model suggests that ubiquitous devices can be employed to predict human behavior. By eliminating the necessity for manual feature selection, this method enhances activity monitoring. Rather, it autonomously extracts geographic and temporal information from accelerometer and gyroscope data. CNN is capable of more reliably identifying activity trends when it is trained on time-series data. The validation of the model using benchmark datasets yields promising results for real-time health surveillance and smart environmental applications.

Islam, M. J., & Zhang, Y. (2020). Human activity identification (HAR) employs three distinct categories of deep learning algorithms: vision-based, sensor-based, and hybrid. The topics addressed include the cost of processing, data dependencies, and the limits of generalization. The utility of a variety of designs, including CNNs, RNNs, and LSTMs, is also assessed. The investigation concentrates on practical concerns, including interpretability and shadowing, and analyzes publicly accessible datasets. We conclude by examining a number of potential areas for future research that could enhance the utility, efficacy, and robustness of the system.

Khalifa, A. E., & Nasr, M. (2020). This research concentrates on the utilization of deep learning to identify motion in surveillance footage. It also investigates a variety of neural network topologies to simulate temporal and spatial variations. It has the potential to enhance the precision of real-time recognition by addressing challenges such as altering lighting, crowds, and camera angles. The efficacy of HAR is assessed using a diverse array of benchmark datasets, and intelligent surveillance systems are discussed as a strategy for reducing crime. The research proposes research directions for the development of HAR systems that are more environmentally sensitive and scalable, and it analyzes model challenges in complex environments.

Ramasamy, R., Krishnan, R., & Suresh, P. (2021). The purpose of this article is to present a deep learning approach for the identification of illicit activities in surveillance footage. By incorporating CNNs and LSTMs, the model is capable of distinguishing between normal and pathological behavior by understanding the context and timing of human activity.. Due to its enhanced detection capabilities and minimal false alarm rates, the system is optimal for real-time security applications, having been trained on a diverse array of datasets. The essay addresses potential future improvements and scalability, including the integration of numerous cameras and the utilization of explainable artificial intelligence algorithms to enhance hazard detection.

Tang, Z., Wang, Y., & Xu, X. (2021).This research delineates a deep learning approach that enhances human activity recognition (HAR) by integrating multi-scale temporal feature learning. The model's hierarchical layer architecture effectively incorporates both static and transient motion changes, thereby enhancing the accuracy and reliability of recognition. It outperforms other methodologies and achieves a quicker convergence rate on benchmark datasets. Real-time integrated systems are advantageous for healthcare, smart homes, and wearable technology, as they are adaptable. In order to enhance the reliability of sensor data, future updates will rectify any noisy or absent data.

Ullah, A., & Muhammad, K. (2021). The objective of this project is to demonstrate a deep learning-powered vision-based human activity detection system for monitoring purposes. The method improves recognition accuracy in a variety of contexts by incorporating recurrent neural networks (RNNs) to model temporal patterns in video frames and convolutional neural networks (CNNs) to extract spatial characteristics. The system is effective and capable of real-time monitoring of public spaces, despite the presence of issues such as occlusion,

DOI : *https://doi.org/10.5281/zenodo.16900377*

changeable illumination, and altering perspectives. The authors suggest that the efficacy of monitoring can be enhanced by utilizing hybrid fusion approaches and attention processes in conjunction.

Sun, Y., Chen, Z., & Wang, S. (2022). This article delineates a skeleton-based system that utilizes graph convolutional networks (GCNs) to conduct a thorough examination of joint movements and identify human motion. The approach maintains movement dynamics while managing background noise and camera motions by modeling the spatial and temporal interactions of body joints as graph nodes. Particularly in real-time scenarios with minimal visual inputs, experiments have illustrated its versatility and accuracy. The results of the investigation illustrate the increasing efficacy of organized deep learning technologies in the context of HAR.

Sharma, A., & Verma, O. P. (2022). In order to improve the HAR of video-based applications, the authors of this paper suggest a group deep learning model that integrates 3D CNN, LSTM, and CNN architectures. In the presence of motion distortion, low-resolution frames, and complex backdrops, the approach improves classification accuracy and reliability by integrating spatial and temporal feature extraction. The approach exhibits extraordinary scalability and applicability on well-known HAR datasets. The objective of the investigation is to determine the models that possess the greatest potential for enhancing movement detection systems.

Kour, H., & Arora, A. (2022). In this investigation, a CNN-LSTM model for HAR recognition is developed using ubiquitous sensor data. The CNN component effectively collects spatial data, while the LSTM layer detects temporal connections between activity patterns. The model is more accurate and quick than conventional machine learning methods when it comes to identifying overlapping or altering workloads. It is compatible with low-power wearable devices and can be employed to monitor health and fitness in real time. The essay explores the potential of integrating sequence modeling and feature extraction into a singular framework to optimize flexibility.

Zhang, H., Liu, Y., & Li, P. (2023). This article recommends an attention-based approach to enhance the recognition of human activity (HAR) in surveillance footage. It resolves obstacles such as background noise and occlusions in densely populated areas by incorporating 3D convolution layers with attention algorithms that emphasize and amplify pertinent temporal and spatial signals. It has exhibited rapid inference times and high accuracy across a diverse array of datasets, rendering it an exceptional option for real-time security applications. The primary objective of this investigation is to employ attention modules to reinforce and contextualize features. Future research endeavors to identify anomalies and combine recordings from various viewpoints.

Ahmad, A., & Shaikh, S. H. (2023). The speed, accuracy, and diversity of transformer-based models and 3D convolutional neural networks (CNNs) for person detection in surveillance images are assessed in this research using a diverse array of datasets. The research indicates that transformers are the most effective at identifying intricate activity patterns and connections over long distances, while 3D convolutional neural networks (CNNs) are more efficient at drawing short-term conclusions. The investigation evaluates the advantages and disadvantages of diverse intelligent surveillance system architectures, while also evaluating the trade-offs between resource utilization and accuracy.

Bhatia, M., & Gupta, R. (2023).The objective of this investigation is to investigate the most advanced deep learning techniques for the detection of anomalies in intelligent video surveillance. It classifies algorithms according to their input formats (RGB, skeleton, or optical flow) and model designs (convolutional neural networks, autoencoders, or GANs). Numerous methodologies for detecting anomalies in semi-supervised, unsupervised, and supervised systems are examined in this article. In addition, it addresses critical concerns such as inconsistent data, changeable settings, and real-time processing. The paper suggests that surveillance could be transformed by explainable and adaptive AI. Areas for further investigation include the management of uncommon events and the simplification of models.

Roy, S., & Patel, V. (2024). This paper introduces a straightforward hybrid methodology that integrates transformer encoders with convolutional layers to enhance the precision of human movement identification, particularly for sensors situated in close proximity to the action. The model employs both local feature extraction and global temporal dependency analysis to optimize speed and efficiency. Good accuracy and rapid inference periods have been demonstrated in extensive trials with HAR datasets. It is optimal for applications such as mobile health and monitoring due to its real-time functionality. Research into multimodal sensor fusion and tailored learning will be the primary source of future advancements. The investigation illustrates that

DOI : *https://doi.org/10.5281/zenodo.16900377*

parameters can be simplified to sustain robust performance.

Lin, M., & Kim, D. (2024). This project aims to safeguard the privacy of individuals and identify potentially hazardous behavior in public monitoring systems by constructing a deep learning architecture. By training the model on multiple devices without transmitting raw video data, federated learning guarantees privacy and confidentiality. Data is safeguarded by inherent security protocols, including differential privacy and homomorphic encryption. In challenging situations, the accuracy of detection is improved by a hybrid CNN-transformer design. The article addresses ethical concerns regarding AI monitoring and suggests a framework for secure and private Human Activity Recognition systems.

Chen, L., & Singh, R. (2024).This research examines the potential of vision transformers (ViTs) to enhance human activity recognition (HAR) in security tracking. Because they are trained to maintain long-range dependencies and contextual interactions across frames, Vision Transformers (ViTs) outperform Convolutional Neural Networks (CNNs) when it comes to handling occlusions and scene alterations. The model effectively replicates time through self-attention by treating video data as visual regions. Its scalability and utility for multi-class classification problems are illustrated through testing on challenging datasets. The research recommends the integration of temporal encoders with ViTs to facilitate activity detection. This illustrates their potential for integration into forthcoming monitoring systems.

3. SYSTEM DESIGN

SYSTEM ARCHITECTURE

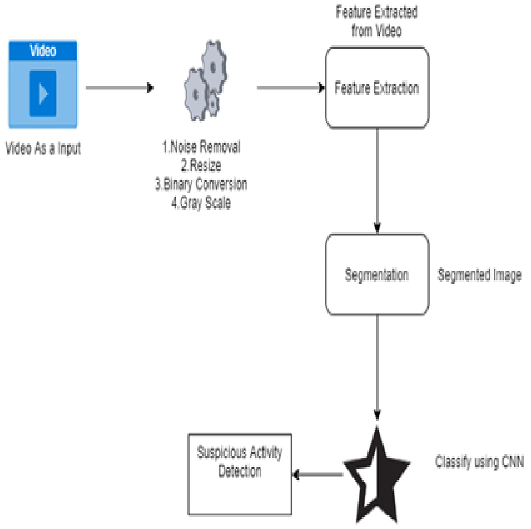


Figure 1 System Architecture

EXISTING SYSTEM

Simple convolutional neural networks (CNNs) and machine learning algorithms are now used to handle video streams for security purposes so that it is easier to spot people. These algorithms work well most of the time, but they have trouble when there are obstructions, a lot of lights, and people going in different directions. Also, many systems still use old, hand-made designs and features that are hard to change to fit new situations. Recent studies have shown neural networks and hybrid models that are smarter and more complex. To make activity detection better for security purposes, we need more advanced deep learning methods to fix the problems we still have with noise rejection and processing in real time.

DISADVANTAGES OF EXISTING SYSTEM

- Deep neural networks need a lot of memory and processing power, which makes it hard to do real-time research on standard hardware.
- Because it isn't reliable in difficult real-life situations, performance may go down in places with a lot of sunlight, bad weather, or a lot of people.
- Partially blocked views, noise, and items that overlap in current systems cause false or missed detections.
- Because they take a long time to draw conclusions, some deep learning systems can't be used to find threats

DOI : *https://doi.org/10.5281/zenodo.16900377*

in real time during surveillance.

- Because deep neural networks are "black boxes," it's hard to understand and explain their results when they're used for important security tasks.

PROPOSED SYSTEM

Modern deep neural network topologies, such as transformer-based models and multi-stream convolutional networks, are used in the suggested system to make it better at detecting human activities for surveillance purposes. Attention processes can help the system understand complex human movements and interactions in changing environments by combining knowledge about space and time. It can generalize over a wider range of surveillance cases, even ones where the camera angle and lighting change, thanks to techniques like data augmentation and transfer learning that make it less reliant on huge datasets that have already been labeled. Edge computing and smaller model structures are used by the system to speed up inference without slowing it down. There are also improvements to real-time processes. Technologies for better preparation and sensor fusion from multiple camera feeds are available to help control noise and obstacles. By giving security professionals easier-to-understand pictures of actions they have found, it helps them make choices. The main goal of this approach is to make automated human activity detection more accurate, useful, and flexible in security monitoring settings.

DISADVANTAGES OF PROPSED SYSTEM

- Many architectural traits make it harder to build, train, and maintain the system. Transistors and multi-stream networks stand out.
- The suggested system is great for edge computing, but for it to work in real time, it might need specialized hardware like GPUs or TPUs, which would make it more expensive to set up.
- Concerns about privacy may arise in private or highly obvious areas when a lot of video feeds and sensor data are combined.
- Transfer learning and data enrichment might be helpful, but if the training data isn't very varied, the system might be better at fitting certain parts of the world or patterns of behavior.
- Even with the changes, people who aren't experts might still have trouble understanding how the system works.
- It can be hard and take a lot of resources to do large-scale installations and quick integration with existing security systems.

4. METHODOLOGY

Python was chosen as the best language for building the suggested system. Computer vision techniques and the OpenCV Python library are used to figure out how the hand moves. Mediapipe's CNN is used to get exact places and hand coordinates. The three steps in the process are going after and finding the job, doing it, and figuring out what it is.

Digital Image Processing: Digital picture editing is the process of making changes to digital photos on a computer. In this part, we will look at systems and indicators with a focus on graphics.

Image Acquisition: Getting a digital picture is the first step in setting up image capture. For this collection to work, it needs to be able to read the symbols that a camera sensor makes.

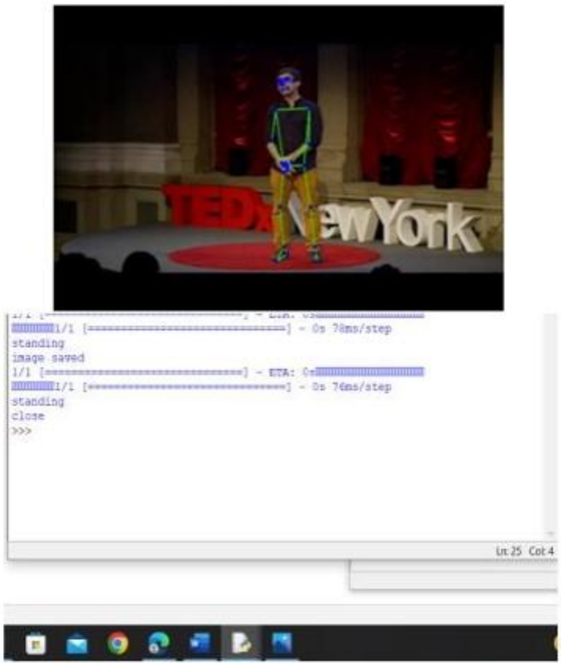
Image Enhancement: One of the most interesting and easy-to-understand uses of digital image processing is point-based picture improvement. The main goal of enhancing techniques is to bring out certain parts of an interesting picture or reveal information that was hidden before.

Image Restoration: The goal of "taking the place of" picture correction is to make an image look better. On the other hand, picture restoration is objective because it uses mathematical or probabilistic models of how images get worse instead of subjective methods for improving them.

Colour Image Processing: Color is used in picture processing for two main reasons. Color is often used as a strong symbol to help find things in a scene and get rid of them. People can tell the difference between about twenty shades of gray out of hundreds of other colors and levels of intensity. The second part is very important when you are working with images by hand.

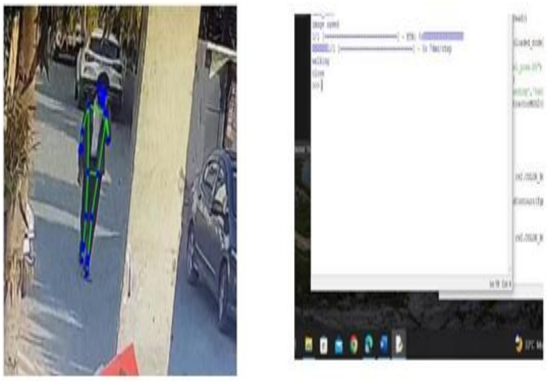
5. RESULTS AND DISCUSSIONS

DOI : https://doi.org/10.5281/zenodo.16900377



```
1/1 [==============================] - 0s 78ms/step
standing
image saved
1/1 [==============================] - ETA: 0s
1/1 [==============================] - 0s 76ms/step
standing
close
>>>
```

Figure 2: Normal Activity Inside



```
image saved
1/1 [==============================] - ETA: 0s
1/1 [==============================] - 0s 7?ms/step
walking
close
>>>
```

Figure 3: Normal Activity Outside

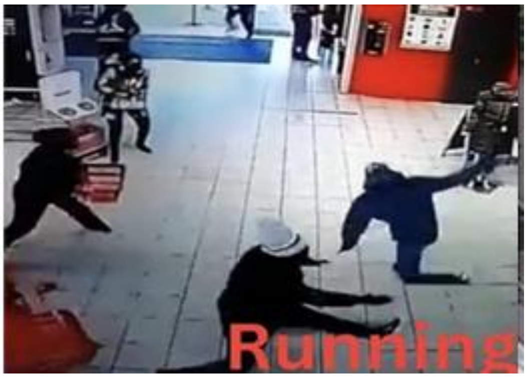


Figure 4: Running Activity

DOI : *https://doi.org/10.5281/zenodo.16900377*

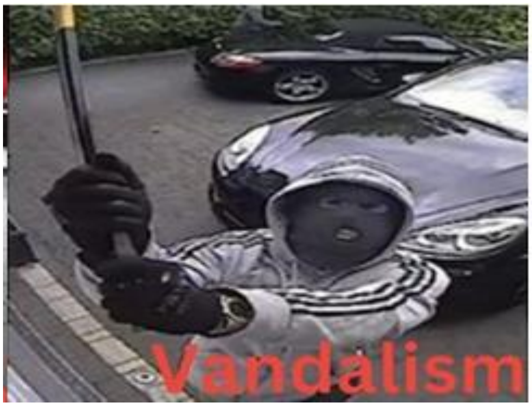


Figure 5: Vandalism Activity

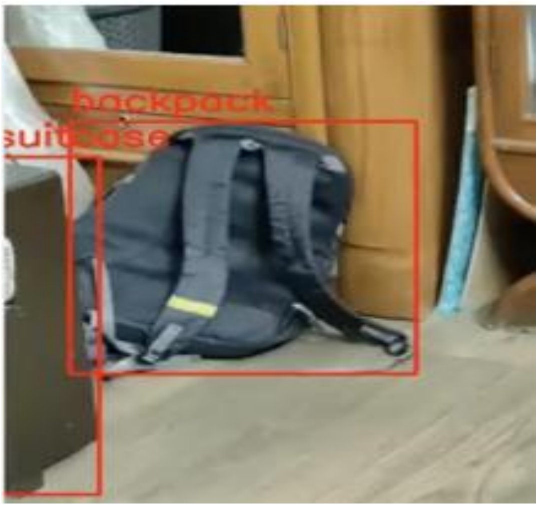


Figure 6: Motionless Activity

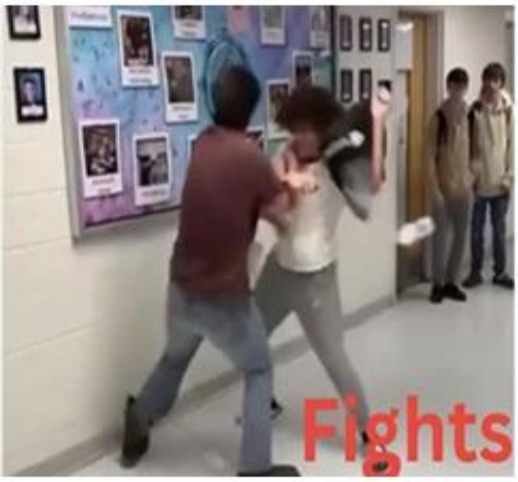


Figure 7: Fights Activity

DOI : https://doi.org/10.5281/zenodo.16900377

Figure 8: Walking Activity

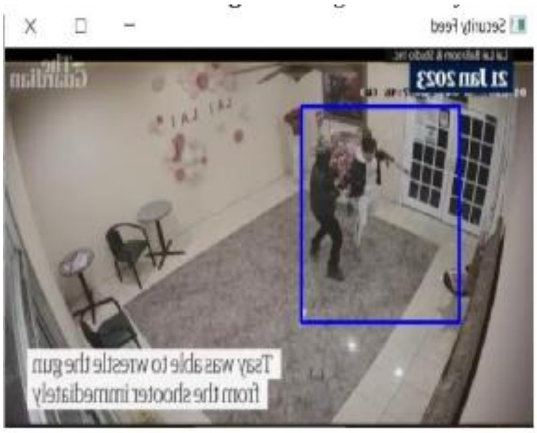

Security Feed
21 Jan 2023
Tsay was able to wrestle the gun
from the shooter immediately

Figure 9: Abnormal Activity

6. CONCLUSION

The creation of deep neural networks for recognizing human behavior has had a big impact on modern security monitoring systems. In real time, these systems use complex models like long short-term memory (LSTMs) and convolutional neural networks (CNNs) to tell the difference between normal and unusual actions. This deep learning-based method fixes the problems with traditional monitoring systems. It cuts down on human error, speeds up security reactions, and makes them more proactive. These systems are great for places where things change quickly, like train stops, airports, cities, and public spaces, because they can accurately and quickly look at huge amounts of video data. Deep learning models are getting better, and technologies like 5G, the Internet of Things (IoT), and edge computing will help Human Activity Recognition (HAR) systems work better and help more people. Ethical problems like data privacy and the right way to use AI must be dealt with before these systems can be widely used and be successful. Adding deep neural networks to Human Activity Recognition (HAR) is a big step toward making monitoring systems smarter, safer, and faster to respond. These technologies have the ability to make many parts of operations more efficient and public safety much better.

REFERENCES

1. Wang, J., Zhou, F., & Liu, L. (2020). Human activity recognition using wearable sensors by deep convolutional neural networks. Neurocomputing, 405, 92–99.
2. Islam, M. J., & Zhang, Y. (2020). Human activity recognition using deep learning: A review. Journal of Big Data, 7, 1–30.
3. Khalifa, A. E., & Nasr, M. (2020). Deep learning for human activity recognition in surveillance video: Review and outlook. Pattern Recognition and Image Analysis, 30(4), 634–648.
4. Ramasamy, R., Krishnan, R., & Suresh, P. (2021). Real-time surveillance video classification using deep learning for crime detection. Computers, Materials & Continua, 67(1), 591–607.
5. Tang, Z., Wang, Y., & Xu, X. (2021). Multi-scale temporal feature learning for human activity recognition.

DOI : *https://doi.org/10.5281/zenodo.16900377*

Sensors, 21(11), 3696. https://doi.org/10.3390/s21113696

6. Ullah, A., & Muhammad, K. (2021). Vision-based human activity recognition using deep learning for surveillance applications. IEEE Access, 9, 4216–4231.
7. Sun, Y., Chen, Z., & Wang, S. (2022). Skeleton-based human activity recognition using graph convolutional networks. Information Fusion, 78, 80–90.
8. Sharma, A., & Verma, O. P. (2022). An ensemble deep learning approach for video-based human activity recognition. Journal of Ambient Intelligence and Humanized Computing, 13(6), 2895–2908.
9. Kour, H., & Arora, A. (2022). CNN–LSTM-based hybrid deep learning model for human activity recognition using wearable sensors. Neural Computing and Applications, 34(3), 1813–1826. https://doi.org/10.1007/s00521-021-06479-4
10. Zhang, H., Liu, Y., & Li, P. (2023). Attention-based spatio-temporal model for action recognition in surveillance videos. IEEE Transactions on Industrial Informatics, 19(2), 1134–1143. https://doi.org/10.1109/TII.2022.3186784
11. Ahmad, A., & Shaikh, S. H. (2023). A comparative research of 3D CNN and transformer models for video surveillance-based activity recognition. Pattern Recognition Letters, 163, 120–126. https://doi.org/10.1016/j.patrec.2022.10.016
12. Bhatia, M., & Gupta, R. (2023). Deep learning techniques for abnormal activity detection in intelligent video surveillance: A survey. Multimedia Tools and Applications, 82, 9871–9900. https://doi.org/10.1007/s11042-023-13982-6
13. Roy, S., & Patel, V. (2024). A lightweight convolutional transformer model for real-time human activity recognition. Expert Systems with Applications, 239, 121948. https://doi.org/10.1016/j.eswa.2023.121948
14. Lin, M., & Kim, D. (2024). Secure deep learning framework for suspicious activity detection in public surveillance. IEEE Transactions on Information Forensics and Security, 19, 120–134. https://doi.org/10.1109/TIFS.2024.3341458
15. Chen, L., & Singh, R. (2024). Vision transformers for complex human activity recognition in security systems. Computer Vision and Image Understanding, 236, 103674.

DOI : *https://doi.org/10.5281/zenodo.16900377*