**International Journal of Environmental Sciences**
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://www.theaspd.com/ijes.php

# Human Activity Identification And Recognition In Aerial Images Using CNN Deep Learning Models

**Mr. Kazi Azizuddin[1], Dr Premal Patel[2]**
[1]PhD Scholar, Department of Computer Engineering, College of Technology, arkazi.ce@gmail.com
[2]Associate Professor, Department of Computer Engineering, College of Technology
premalpatel.ce@socet.edu.in

***ABSTRACT***: *The Human Action Recognition (HAR) from images, video clips is motivated due to abundant availability of videos and images. There is a need of diverse applications for automatic observation of sick people, security of senior citizens and kids through interfaces between man and machine. Automatic labeling of activities of people in images captured through the cameras of closed circuit television will be the motive of such applications. In recent years there is huge demand of image analysis and in particular -tagging of suspicious actions in surveillance system. The main objective of this work is to develop a model that can identify and classify the predefined activities of an individual. In addition to its technical significance, this research also recognizes the broader relevance of sustainable management practices and their transformative role in environmental sciences. Integrating energy-efficient computational models and responsible data handling mechanisms can contribute to reduced resource consumption and improved ecological outcomes. Moreover, by exploring how management strategies, innovations, and policies influence the deployment of HAR systems, it becomes possible to align technological progress with environmental sustainability goals across sectors and societies. This dual perspective not only highlights the utility of HAR in improving human welfare and security but also emphasizes its potential in supporting global sustainability frameworks.*

***KEYWORDS***: *Deep Convolution Neural Network, Activity Recognition, CNN.*

## INTRODUCTION

The developments in artificial intelligence are more prominent in the reorganization of human action, especially in computer vision and pattern recognition. There are promising results in tagging the images and videos with the relevant actions. The process involves feature extraction from recorded images and videos based on relevance, developing model and identifying acts of human. Research is persuaded in developing programmed observations to detect suspicious activities in both public areas and private properties. The usage of such software can be in key areas for bus terminals, hospitals senior citizens in home, banks, office. It is really difficult to recognize human action from large gathering images and video clips, training the template and next finally recognition. In due process, it has to consider quality of images in terms of resolution, scene appearance, anthropometry which includes altitude & thickness, scale, camera position, motion, illumination, clutter in background and performance issues pose and execution rate. In the present approach the objective is to develop a machine learning prototype to recognize a sub set of human actions which are predefined. The training and testing is to be carried out on datasets which are considered as bench mark datasets. The technique used should also give emphasis on accuracy in classification, reduced computational cost and ability to provide reasonably good performance.

Supervised CNN is utilized to automatically learn and extract optimal feature representations that can effectively differentiate between human and non-human classes. Unlike traditional hand-crafted feature extraction methods, the CNN is able to capture complex spatial patterns and hierarchical information directly from the input data, thereby improving the robustness of the recognition process. Once the discriminative features are obtained, the final classification is carried out using Softmax and SVM classifiers. While the Softmax function provides a probabilistic interpretation of class membership, the use of SVM enhances the decision boundary by maximizing class separability. This combination allows the system to leverage the strengths of both deep learning and classical machine learning approaches, ultimately leading to more accurate and reliable classification results[8].

### 1.1 Applications

1. **Automated surveillance systems:** Daily life tracking systems usually aim to provide a simple guide for

activity reporting, or to assist with exercise and healthy lifestyles. Such devices are fitted with embedded sensors such as accelerometer, gyroscope, GPS; monitoring steps taken by people, climbing stairs, burning calories, sleeping hours, moving distance, sleep quality, suspicious activities for real time reactions like fighting ,stealing etc.

**2. Airports:** Through real-time cameras installed in drones and special security flights sensitive areas such as country borders, sea beaches can be identified for attacking enemies or engaging in terrorist or bomber aircraft entry.

**3. Radar Imaging systems:** Radar and sonar images are used to detect and recognize different target types or in aircraft or missile systems guidance .There is a multi camera system that uses a new calibration-free behavior recognition method for monitoring human activity at subway station.

**4. Intelligence analysis and gathering:** In sensitive areas like the border with the country, sea beaches can be identified and the tracing by means of real time cameras installed on drones and special safety flights by attacking enemies or engaging in terrorist or bomber activities.

Recent studies have shown that deep learning methods are being increasingly adopted across diverse domains, highlighting their adaptability and effectiveness. For instance, deep learning techniques applied to aerial imagery have proven valuable in land search and rescue missions, where accurate visual interpretation supports time-critical operations [10]. Similarly, radar-based approaches have been explored for human activity recognition (HAR), with deep learning methods enabling precise detection and classification of movements [11]. Expanding further, multimodal sensing devices integrated with deep learning frameworks have demonstrated improved recognition accuracy by leveraging information from multiple data sources [12]. Surveys in this field underline the rapid progress of HAR research, particularly emphasizing the advantages of deep models over traditional pattern recognition methods [13]. To enhance recognition performance, strategies such as coarse-to-fine convolutional networks have been employed, allowing systems to capture both global action patterns and fine-grained details [14].

Beyond HAR, attention-based convolutional models have shown their effectiveness in environmental applications like landslide detection, further proving the versatility of deep learning approaches [15]. In parallel, mobile sensor data combined with deep learning has been successfully applied for user identification, expanding the scope of HAR into security and personalization domains [16]. Researchers have also explored data integration methods, where merging heterogeneous datasets enhances the robustness of activity recognition models [17]. Addressing challenges of scalability, domain adaptation techniques have been introduced to allow HAR systems to operate effectively across varied environments and datasets [18]. Several studies reaffirm the strong performance of deep learning methods for HAR, consistently reporting improvements in recognition accuracy and resilience compared to conventional techniques [19]. Hybrid approaches that integrate support vector machines (SVM) with deep networks have also been proposed, combining statistical and deep learning strengths for more reliable classification [20]. Additionally, wearable sensor data analyzed with deep models has opened promising applications in healthcare and personalized monitoring, further extending the real-world relevance of HAR [21].

## RELATED WORK

The problem of categorizing human activity has remained a challenging task in computer vision [1]. An approach to order the movement of human focus in a video clip. The entities which are moving are identified and their boundaries are removed. The inventors recommend utilizing a star skeleton conveyed through the entity's boundary as an aid for human movement examination. It has a star skeleton that gives two movement prompts: pose of the figure, and recurrent movement of skeleton fragments. These two factors plays important role in resolving the movements of exercise postures, for example jogging, running, and hand waving. It doesn't require an earlier human model, computationally affordable Some Issues 1.Creating relative joint images was difficult due to calculation of relative distance between referenced joints and the other joints.2.The joints in the frames where not within the predefined temporal dimension [2]. The point tracking approach is used when detected objects are represented as points. Tracking is performed by evaluating the objects state in terms of position and motion and by associating points across frames. The movement of an entity is to be keenly

tracked to check is it in straight path, which can be trusted or dismissed as surrounding area. But in most applications the entities cannot be restricted to move in straight path. This limits the adoptability of the algorithm. Issues 1. It is expensive in terms of computational cost due its requirement. The computation is to be done for each and every point in stream which makes it less popular for applications in real time.

2.Requires external mechanism to detect in every frame. Problem areas in handling occlusion especially for objects represented by multiple points, misdetection, entries and exits of objects [3]. An approach to order the movement of human focus in a video clip. The entities which are moving are identified and their boundaries are removed. The inventors recommend utilizing a star skeleton conveyed through the entity's boundary as an aid for human movement examination. The system works by extracting key points from the human body, applying quick shift segmentation, and using EM-GMM-based elliptical modeling. These features are then refined through a naïve Bayes optimizer, and finally, the activities are classified using a CNN model [4].

## METHODOLOGY

The first stage of the proposed model is the collection of videos from the bench mark video data sets. i.e KTH dataset. These video clips collected are to be converted to frames. The next step is to determine the difference between the frames. This difference is used for the detection of activity. Once the object of interest has been detected, the background subtraction must be done. The final task is quality enhancement, which plays a vital role where RGB videos are low-quality videos. Human activity recognition involves actions such as running and jogging, if the videos collected are low-quality videos, further activity recognition is a daunting task. Here R-CNN is used to train the data set videos to process the key features [5]. Human Action Recognition (HAR) has become an important area of study in computer vision research. Most of the existing work in HAR focuses on using videos captured in the visible spectrum and considers HAR as a typical pattern recognition problem, generally carried out in two stages: feature extraction and classifier learning. During feature extraction, researchers have explored different approaches for representing actions in both spatial and spatio-temporal domains, with techniques such as Histogram of Oriented Gradients (HOG) being widely used [6]. Among the tested actions, clapping and waving were the most accurately recognized and consistently produced strong results across all classifiers. Both MLP and SVM outperformed deep learning models, mainly because the limited amount of available data restricted the performance of LSTM models, which generally require larger datasets to achieve high accuracy [7]. In the proposed work we tend to initially recognize the activity performed by the person then eliminate the background pictures (by the technique of background subtraction), solely only specific person performing associate activity is known later the specific activity is labeled and recognition of performance takes place. In the validation step the output is validated using trained neural network with original and estimated values. . The important role in the entire process is the key point extraction which helps to categorize the action and is labelled accordingly.

**Step 1:** Load the video and convert in to frames
**Step 2:** Extract features **Step3:** Train CNN
**Step4:**Identify Activity

Pre-processing step in image processing also utilize Gaussian blur method, in order to improve the structures of the images at distinct scale, view scale place description and scale space application. Computationally, using a Gaussian smoothing to a frame is similar to convoluting the frame with a Gaussian-function. It is also called as 2 dimensional weier -strass transformation.

The CNN algorithm propose a bunch of boxes in image.

**Step 1:** Takes video as an input.
**Step 2:** Generates initial segmentation so multiple regions from image is identified.

**Step 3:** Combines the similar group of regions to form larger region ( based on shape similarity, size similarity, color similarity)
**Step 4:** Finally these region proceed to find activity detection

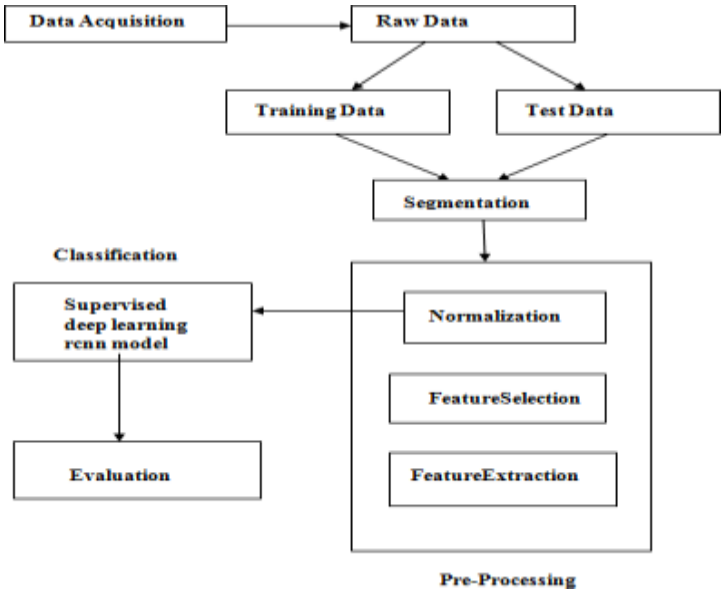


Figure 1-Proposed System for Human Activity Recognition

INPUT DATASET

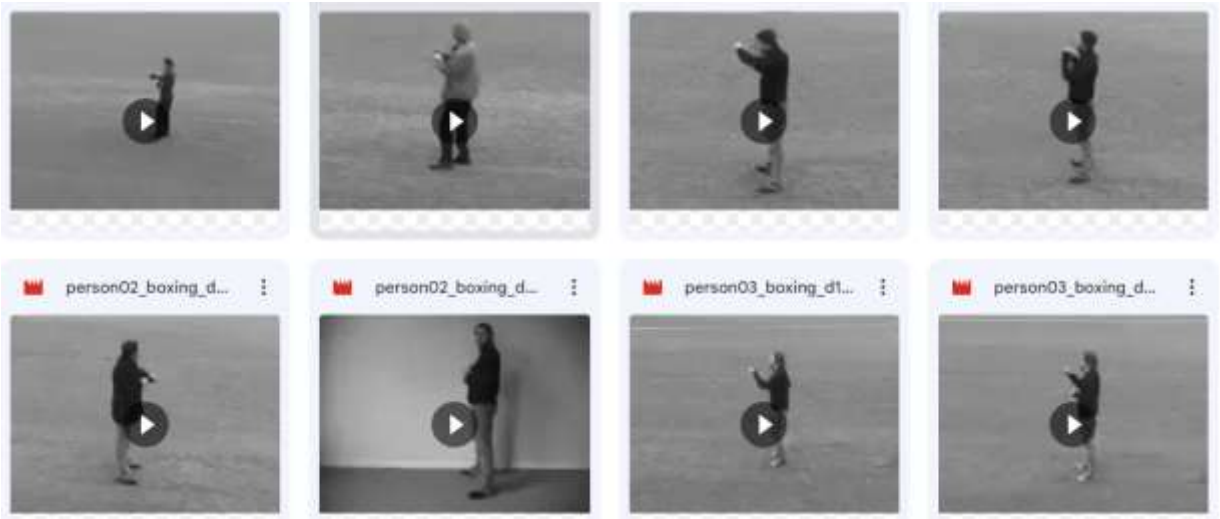


Figure -2: Person Boxing


Figure -3: Person Handclapping

**International Journal of Environmental Sciences**
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://www.theaspd.com/ijes.php


Figure -4: Person Handwaving

**Experimental Results**


Figure -5: Labelled activity recognized Person Handwaving

Figure -6: Labelled activity recognized Person Running

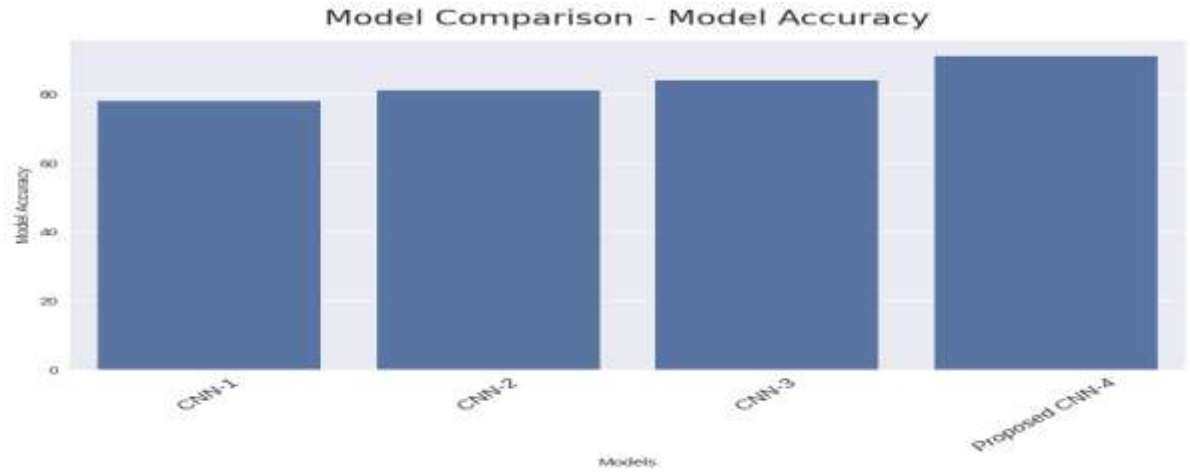


Figure -7: Comparison Chart

| Algorithms | Accuracy % |
|---|---|
| CNN-1 | 78 |
| CNN-2 | 81 |
| CNN-3 | 84 |
| Proposed CNN-4 | 90 |

Figure -8: Comparison Table

The project is a python implementation. The Python program is run from the Idle IDE**.** There are different acts in the videos. First, video is collected from the directory, and then processed where the background containing the unwanted images is removed, and then processing steps are taken to identify activity only by people being moved for activity recognition**.**

The table 1 gives the detailed description about all the possible initial test cases which are required for the project

| **Test case** | **Test case description** | **Input** | **Output** | **Test case result** |
|---|---|---|---|---|
| 1 | Uploading Video | Video dataset | Processing video frames | Pass |
| 2 | Processing of video | Video frames | Recognition activity from | Pass |
| 3 | Motion Detection | frames | Classifying | Pass |

| | | | | |
|---|---|---|---|---|
| 4 | Recognition activity | frames | Activity recognized as walking, running etc | Pass |

## CONCLUSION

In this paper an accurate and efficient activity identification system has been developed which achieves comparable metrics with the existing state-of-the-art system. This paper uses recent techniques in the field of computer vision and deep learning. An important scope would be to train the system on a video sequence for usage in tracking applications. Addition of a temporally consistent network would enable smooth detection and more optimal than per-frame detection. Limitations are when there are limiting conditions, which means the project works only in certain criteria. This project cannot give an ideal, 90% accuracy, which is a limitation that comes naturally. The Model can be trained to make the classification multi-labelled classification. The deep CNN (Convolutional Neural Networks) classifiers are pre-trained. The efficiency of a CNN classifier is more when compared to other neural network classifiers. Beyond these technical contributions, this study also opens discussions on how sustainable management practices can shape the responsible design and deployment of such intelligent systems. Incorporating energy-conscious computational frameworks and environmentally friendly data infrastructures could reduce the ecological footprint of large-scale HAR implementations. Furthermore, aligning this research with broader management strategies, innovative practices, and policy-level decisions can help ensure that the benefits of HAR extend beyond individual applications to societal well-being and long-term environmental sustainability. In this sense, the model is not only a step toward advancing human-centered security and healthcare solutions but also a demonstration of how emerging technologies can be guided by sustainable principles to create lasting impact across multiple sectors.

## REFERENCES

[1] Sehar Un Nisa , Muhammad Imran , Shaheed Zulfikar, “A Critical Review of Object Detection Using Neural Networks”, International Conference on Communication computing and Digital systems, Bhutto Institute of science and technology Islambad, pp 154-159 ,2019.
[2] Dangwei Li, Student Member, IEEE, Zhang Xiaotang Che,and Kaiqi Huang,Senior Member, IEEE “A Richly Annotated Pedestrian Dataset for Person Retrieval in Real Surveillance Scenarios”, University of Chinese Academy of Sciences, Beijing, pp 1-15 VOL. XX, NO. X, OCTOBER 2018
[3] Russell Stewart, Mykhaylo Andriluka, and Andrew Y. Ng,“End-to-end people detection in crowded scenes”,Stanford University, USA Max Planck Institute for Informatics, pp 2325-2333,VOL III 2016.
[4] Azmat, U., Alotaibi, S. S., Abdelhaq, M., Alsufyani, N., Shorfuzzaman, M., Jalal, A., & Park, J. (2023). Aerial insights: deep learning-based human action recognition in drone imagery. IEEE Access, 11, 83946-83961.
[5] Azizuddin, K., Patel, P., & Shah, C. (2022, October). Human Activity Recognition Using Supervised Machine Learning Classifiers. In International Conference on Information and Communication Technology for Competitive Strategies (pp. 87-97). Singapore: Springer Nature Singapore.
[6] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). Ieee.
[7] Kapoor, S., Sharma, A., Verma, A., Dhull, V., & Goyal, C. (2023). A Comparative Study on Deep Learning and Machine Learning Models for Human Action Recognition in Aerial Videos. International Arab Journal of Information Technology (IAJIT), 20(4).
[8] AlDahoul, N., Md Sabri, A. Q., & Mansoor, A. M. (2018). Real-Time Human Detection for Aerial Captured Video Sequences via Deep Models. Computational intelligence and neuroscience, 2018(1), 1639561.
[9] Singh, M., Shah, C., & Patel, P. (2023, August). Lung Cancer Prediction Using Machine Learning Models. In International Conference on ICT for Sustainable Development (pp. 613-618). Singapore: Springer Nature Singapore.
[10] Božić-Štulić, D., Marušić, Ž., & Gotovac, S. (2019). Deep learning approach in aerial imagery for supporting land search and rescue missions. International Journal of Computer Vision, 127(9), 1256-1278.
[11] Li, X., He, Y., & Jing, X. (2019). A survey of deep learning-based human activity recognition in radar. Remote sensing, 11(9), 1068.
[12] Ihianle, I. K., Nwajana, A. O., Ebenuwa, S. H., Otuka, R. I., Owa, K., & Orisatoki, M. O. (2020). A deep learning approach for human activities recognition from multimodal sensing devices. Ieee Access, 8, 179028-179038.
[13] Gu, F., Chung, M. H., Chignell, M., Valaee, S., Zhou, B., & Liu, X. (2021). A survey on deep learning for human activity recognition. ACM Computing Surveys (CSUR), 54(8), 1-34.
[14] Avilés-Cruz, C., Ferreyra-Ramírez, A., Zúñiga-López, A., & Villegas-Cortéz, J. (2019). Coarse-fine convolutional deep-

learning strategy for human activity recognition. Sensors, 19(7), 1556.

[15] Ji, S., Yu, D., Shen, C., Li, W., & Xu, Q. (2020). Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. Landslides, 17(6), 1337-1352.

[16] Alawneh, L., Al-Zinati, M., & Al-Ayyoub, M. (2023). User identification using deep learning and human activity mobile sensor data. International Journal of Information Security, 22(1), 289-301.

[17] Patil, B. U., & Ashoka, D. V. (2023). Data integration based human activity recognition using deep learning models. Karbala International Journal of Modern Science, 9(1), 11.

[18] Khan, M. A. A. H., Roy, N., & Misra, A. (2018, March). Scaling human activity recognition via deep learning-based domain adaptation. In 2018 IEEE international conference on pervasive computing and communications (PerCom) (pp. 1-9). IEEE.

[19] Moola, R., & Hossain, A. (2022, December). Human activity recognition using deep learning. In 2022 URSI regional conference on radio science (USRI-RCRS) (pp. 1-4). IEEE.

[20] Parameswari, V., & Pushpalatha, S. (2020). Human activity recognition using SVM and deep learning. European Journal of Molecular & Clinical Medicine, 7(4), 1984-1990.

[21] Gupta, S. (2021). Deep learning based human activity recognition (HAR) using wearable sensor data. International Journal of Information Management Data Insights, 1(2), 100046.