Power System Technology
ISSN:1000-3673

Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

A Detail Study of Support Vector Machine, K-Nearest Neighbours and Convolutional Neural Network for Human Action Recognition in Videos

Mr. Tammmarguddi Mahmad Husen[1] and Dr. Swamy L N[2]

[1,2]Dept.of CSE, VTU-RRC Mysuru, Visvesvaraya Technological University, Belagavi-590018, Karnataka,India.

[1]tammarguddimahmad124@gmail.com, [2]swamyln@gmail.com

Abstract

Over the previous span, Human Activity Recognition (HAR) has evolved into a critical research area within the computer vision, driven by advancements in video-based action recognition techniques. Unlike image-based methods, video-based HAR leverages Spatial and Temporal information, offering a richer understanding of human behaviors. This area has found several applications in diverse domains, including education, intelligent surveillance, healthcare, entertainment, and autonomous systems. The cameras and sensing devices: There is an increasing demand for automated HAR systems utilizing computationally intelligent methods such as Deep learning (DL) and Machine Learning (ML). This paper delivers a detail study of DL and ML techniques applied to HAR between 2014 and 2025. It explores various modalities used for action recognition, including RGB-D cameras, audio, and inertial sensors, and examines their roles in enhancing HAR performance. A detailed analysis of public datasets is presented, highlighting their characteristics, strengths, and limitations. Additionally, this survey explores into how action representation, dimensionality reduction, and actually action analysis methods, identifying their respective advantages and drawbacks. In this paper discusses applications of HAR, including human-computer interaction, remote health monitoring, virtual reality, and abnormal behavior detection, emphasizing its transformative impact on these fields. Key challenges, such as scalability, real-time processing, and environmental variability, are outlined, along with the future research directions aimed at developing robust and efficient HAR systems. This survey serves as a valuable resource for researchers and practitioners, providing insights into the state-of-the-art techniques and to make it easier for further advancements in HAR.

Key Terms**:** *Challenges In HAR, Modes of Activities, Public Datasets For HAR, Video-Based Action Recognition.*

1. Introduction

HAR refers to the process of automatically identifying and classifying physical activities or behaviors of individuals based on the data collected from the diverse sources such as wearable sensors, cameras, or environmental sensors. This goal is to analyze motion, gestures, and interactions to recognize specific activities in real time or from pre-recorded data. As an important part of computer vision, HAR is the key technology for machineries to understand the world, as well as human behaviour. More recently with the continuous development of deep learning and sensor technology [20], the efficiency of HAR has been significantly improved,

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

and it has the most practical applications, including health-care, human-computer interaction, virtual reality, etc [4], [1].

HAR is a significant area in a computer vision research, aims to automatically analyze activities from videos, classify them into appropriate categories, and assist in reducing manual work, such as enhancing security through automated video processing systems that detect malicious activities and alert authorities in real time [1], [24]. It can be installed in public places like metro and railway junctions, bus stands, shopping malls, tourist places, airports, etc., to help prevent explosive attacks. Any person(s) leaving their luggage in a public place can be detected, thus reducing casualties. Identifying abnormal activities in real time might help victims, but this is highly time consuming and requires more resources as lengthy surveillance videos must be processed. As an escape, video files can be compressed using adaptive video compression technology.

When it comes to HAR, an object can be detected either through the human eye or through some form of sensing technology. Human activity can be classified into four categories [1],[24].

- Gesture: Based on the movements of the hands, face or other body parts. Does not require verbal communication.
- Action: It Involves movements performed by a individual person, such as running or walking.
- Interaction: Actions performed by two persons. May include interaction with objects or other individuals.
- Group activities: It Refers to the combination of actions, gestures, or interactions. Involves minimum two individuals, often interacting with objects.

Table 1: The challenges related to the recognition of human activities.

Challenges	Description
Occlusion and Clutter	Difficult to recognize the action when the person are partially represents or substantial background noise.
Viewpoint and Scale Variation	Identifying the actions in the different angles or varying scales remains a significant obstacle.
Scalability	Handling the large datasets and wide range of actions [20], [35].
Real-time Processing	Develop a model to capable of operating in real-time scenario like Surveillance.
Generalization	Ensure the model perform consistently across the datasets and real-world scenarios without overfitting.
Complexity of Human Activities	To Recognize the human activity patterns poses to considerable difficult task [20].

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

Correlation and Interference	In the group of activity detecting the single person action is challenging due to the influence other person [20].
Lighting and Environmental Conditions	Varying the lighting, shadows or reflection of light significantly impact on the model accuracy.
Computational Complexity	High demands for real time applications.
Multi-person Interaction	In multi-person interaction involving the multiple individuals introduces the overlapping challenges.
Dataset Limitations	In limitations of the real-world complexity and adequate annotations.

Table 2: Features and Descriptions of Various Human Action Recognition (HAR) methods.

Ref	Features	Descriptions	Key Example
[2]	Interest Points	It focuses on detecting exclusive spatiotemporal positions that are significant for understanding signal.	Harris3D interest point detector.
[1]	Trajectory-based Features	It can be capturing the movement of key points or areas over time, providing insights into motion patterns. Dense trajectory features and upgraded dense trajectories are frequently used.	Dense Trajectories.
[41]	Depth-based Features	Depth sensors provide 3D information, helping to segregate between overlapping or occluded activities. Depth Motion Maps (DMMs) are popular for briefing the depth data.	Depth Motion Maps.
[1]	Pose-based Features	It extracts the body joint places and angles over time to analyze actions. Open Pose and Media pipe are broadly used tools for pose estimation.	Pose estimation using skeleton sequences.
[24]	Appearance based features	It captures visual patterns in the method of textures, edges, and shapes. Histograms of Oriented Gradients (HOG) and Optical Flow are prominent methods.	HOG3D descriptor for action recognition.
[1]	Spatiotemporal Features	It jointly analyses spatial and temporal aspects of action, such as motion history	Two-stream CNN for action recognition.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

		images (MHI) or convolutional neural networks (CNNs) for video analysis.	
[22]	Histogram Oriented Gradients (HOG)	It captures the structure of human actions based on the distribution of gradients or edge orientations in localized parts of an image. Often utilized in combination with other features.	HOG for pedestrian detection.
[18]	Optical Flow	It is employed to capture the motion of objects among consecutive frames. It is mainly useful in dynamic scene analysis and activity recognition from video.	Dense optical flow for action recognition.
[22]	Frequency Domain Features	Features analyze periodic components in time-series data. Fast Fourier Transform (FFT) and spectral analysis are often employed to capture rhythmic actions or gestures.	Frequency analysis for gesture recognition.
[43]	Deep Learning Features	Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), including variants like LSTM or GRU, are used to automatically derive hierarchical features from raw data.	CNN-LSTM architecture for action recognition in videos.
[17]	Fourier Transform-based Features	Fourier Transform is useful to extract frequency components that represent periodic activity patterns. This is useful for distinguishing repetitive or rhythmic activities.	Fourier features for gesture recognition.
[14]	Skeleton-based Features	In these approaches analyze the body joint positions and their comparative motion over time. This is especially useful for 3D body pose analysis.	Skeleton-based action recognition using joint angles.

1.1 The contributions of this study are summarized as follows:

- A comprehensive review of HAR techniques, with a main focus on advancements in DL and ML methodologies This extensive review helps researchers understand the evolution of methods in HAR.
- This paper presents a detailed analysis of different action recognition techniques, highlighting their pros and cons. This information is the valuable for practitioners in selecting appropriate methods for specific applications.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

- It outlines various applications of HAR, including human-computer interaction, content-based video summarization, abnormal activity detection, and so on. Thereby understanding the real-world applications.
- This paper recognizes key challenges in the HAR and suggests future research directions. This contribution is essential for guiding ongoing and future studies in improving HAR systems

It Compares the accuracy of various widely used ML and DL classification techniques in recognize different activities including standing, running, sitting, walking upstairs, walking downstairs, walking, inactive and lying. Table 2: The above table shows that the Features and Descriptions of Various HAR Methods.

2. Survey Over Various Papers of the HAR System.

2.1 Human Activity

As an important research part, recognition of activities using diverse and advanced algorithms, which can be seen in Figure. 1, play a significant role in improving the quality of life and increasing safety in communities. The complexity and diversity of human behavior necessitate a through the analysis and precise the differentiation of human activity patterns [20]. It examines the various scientific articles that are focus on the identification and classification of human activity patterns under the varying conditions. These studies contribute to a better understanding of the methodologies used for Human activity detection and their effectiveness in real life [20].

Various frameworks and algorithm have been proposed for recognizing the human activities using sensor data. One of these frameworks utilizing the SVM as suggested [29]. In this approaches data is collected through a smartphone, different sensors and then stored in a server. The data is encoded using a feature vector. The conducting the experiments by these researchers demonstrate that this framework accurately detects the static and dynamic activities with an 87.1% accuracy. Numerous features extracted from the inertial sensor in the smartphone, such as mean, median and autoregression coefficients, have been enhanced through the KPCA and LDA.

Other researchers [28] have utilized a DBN to train the numerous features for activity recognition. In this method, features are trained using SVM and ANN. The outcomes shows that DBN outperforms SVM and ANN. Moreover, in [31] PCA is employed to extract the most relevant data from tri-axial accelerometer and gyroscope sensors of mobile phones in the form of signals. Experimental results demonstrate that the proposed algorithm by these researchers achieves the best performance with 96.11% accuracy in rec organizing physical activities, outperforming other ML classifiers on a publicly available dataset.

In [30], a Semi-Supervised Active Learning (SSAL) approach is introduced for self-generating relative marginal annotations for activity recognition based on Self-Training (ST).

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

In this method, SSAL reduces the annotation effort to produce the required volume of marginal annotation data to obtain the best classifier.

In Action recognition in videos are continuous to pose challenges are to variations in the illumination conditions, viewpoints, background clutter and occlusions [1]. Additionally addressing the issues like zooming, camera motion and dynamic backgrounds is critical for the developing a robust action system [24]. Table 1: The above table shows that the challenges in HAR.

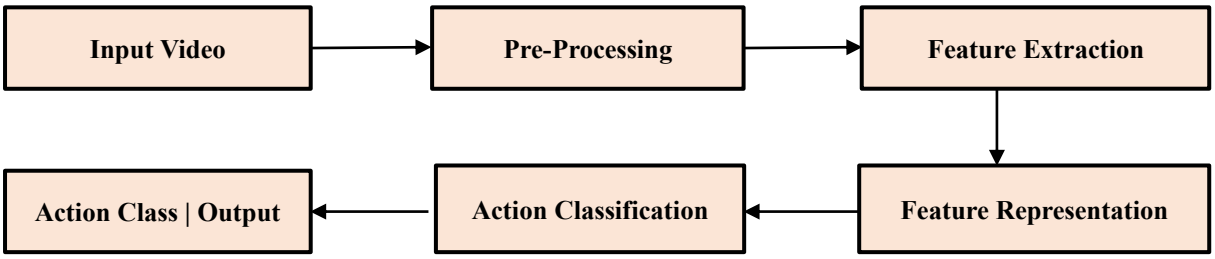

Figure 1: Block Diagram of Human Action Recognition (HAR)

Figure 1. Shows that the Block diagram of HAR. The HAR System designed to analyze the video input on the structured and systematic approach to preprocessing, feature extraction, represent the features, identifying the action classifications give the proper output.

Pre-processing: The accuracy depends on the quality of the of pre-processed data and how it effectively deals with the model. Date cleansing, smoothing, noise removal and grouping the important components of the data preprocessing. After that the dataset undergoes the augmentation, involves the generating perturbed versions of the images to increase the data diversity. These techniques enhance the model ability to generalize and improvement of the performance by creating the variations in the input dataset [6]. where operations such as frame resizing, noise removal, or frame selection are applied to prepare the data for analysis. This step ensures the video data is clean and consistent for subsequent stages.

Features Extraction: Following pre-processing, "Feature Extraction" is performed, involving the identification and capture of distinctive attributes or patterns related to the actions in the video, such as motion dynamics, spatial relationships, or sequential dependencies. Action recognition requires feature extraction, which is an essential component. It entails the removal of body positions and movement patterns from pictures and recordings of human activity [6][1].

- Interest Points: Space-Time Interest Points (STIP), Scale-Invariant feature transformation (SIFT), Color Space- Time Interest Point.
- Shape-Based Feature: Histogram of Oriented Gradients (HOG), Silhouette, Image moments.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

- Motion Based: Motion History Images (MHI), Motion Binary Images (MBI). Histogram of Optical Flow (HOF).
- Pose Based: 2D and 3D spaces.
- Trajectory Based: Harris Corner Detector, Dense Trajectory, Trajectory-pooled Deep Neural Network.

Feature Representation: The extracted features are refined and organized into a Feature Representation format, ensures the data is suitable for subsequent processing and analysis in downstream tasks.

Action Classification: The feature representations serve as input for the Action Classification, where the advanced algorithms categorize actions into the predefined classes. The ML models are usually applied with handcrafted features, while DL models automate feature extraction and classification.

Action Class: The final output is an Action Class, representing the detected or predicted action from the input video. This complete pipeline demonstrates a robust methodology for processing and analyzing complex video data, ensuring accurate and effective action recognition, [6][21].

Table 3: Summary of previous surveys and their key points.

Authors	Year	Main topics / Area of Interest
Guangchun Cheng et al [13]	2015	Summary of recent advances in human action recognition, focusing on automatic recognition of low-level actions and high-level activities, and discussing progress in the field over recent years.
Yu Kong and Yun Fu [3]	2018	Comprehensive review on vision-based action recognition and prediction, covering models, algorithms, technical challenges, popular datasets, evaluation protocols, and future directions in the field.
Valter Estevam et al. [12]	2019	Survey on zero-shot action recognition in videos, discussing methods for classifying instances from unseen classes, techniques for visual and semantic feature extraction, and learning mappings between these features, along with datasets, experiments, and open issues.
Hieu H. Pham et al. [9]	2022	Overview of state-of-the-art deep learning techniques for video-based human action recognition, analyzing deep learning models, their advantages and disadvantages, and identifying current trends and unresolved challenges in the field.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

Zhou Shuchang [4]	2022	A Survey on HAR methods across various input modalities, including RGB-D cameras, audio, and inertial sensors, with a focus on multimodal approaches and an introduction to benchmark datasets and performance comparisons.
Gayathri T and Mamatha HR [9]	2024	A Comprehensive analysis of activity recognition models, comparing their performance and computational requirements, providing an overview of benchmark datasets, and examining applications in sports, surveillance, movies, and robotics.

The above Table 3 shows that the Summary of the previous study and their key points. Action representation involves the low-level processing of human actions. Typically comprising two steps: interest detection and describing the interest boundary. Key features are extracted, and encoding is performed using various techniques [1].

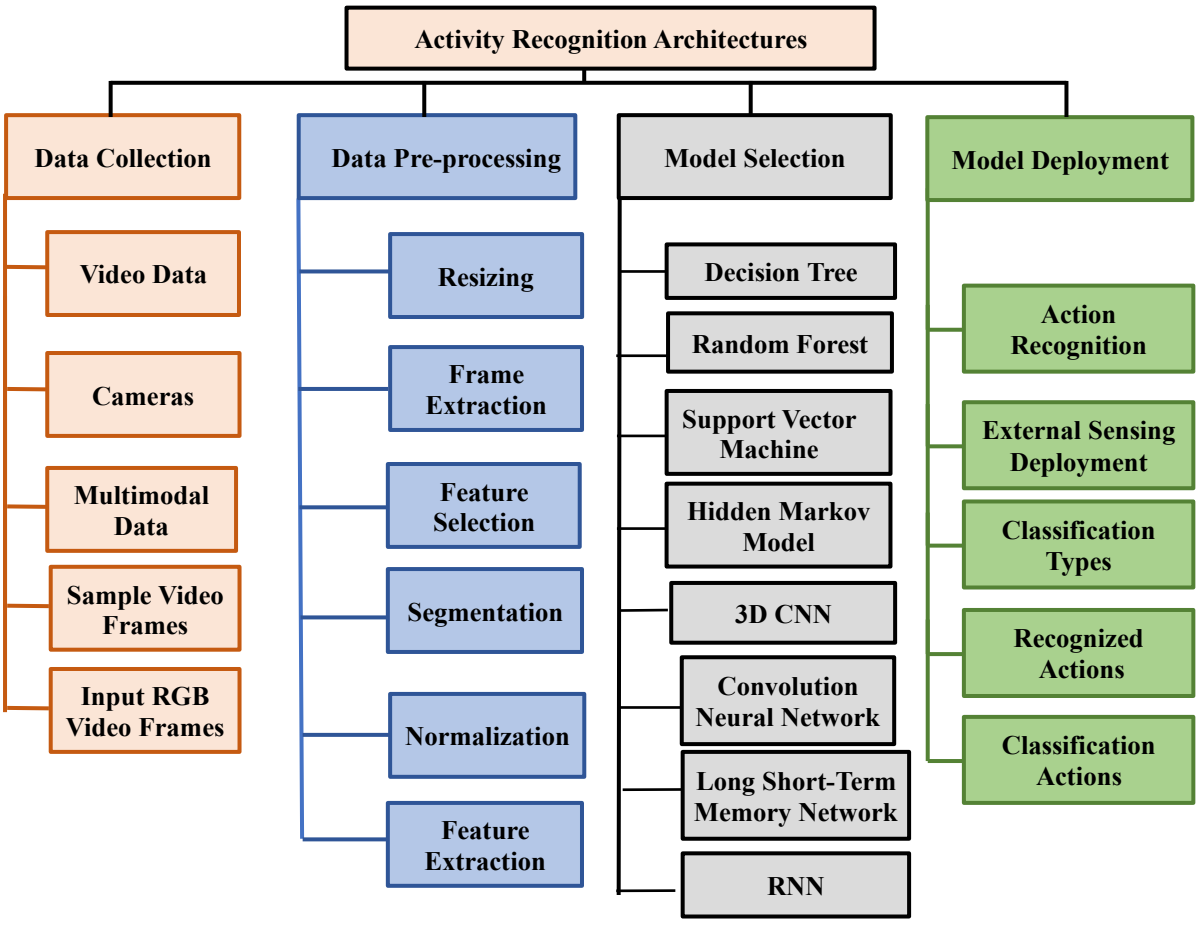

Figure 2: Architecture Steps in HAR [20]

Figure 2 Shows that the Data Architecture Stages in HAR, recognizing actions involves two critical aspects, action representation and action analysis. These actions are required to

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

capture the various types of sensors, like a RGB Cameras, Range sensor, Radar and wearable sensors [1]. Smartphones for a Smartphones for instance, integrate multiple sensors, like accelerometer, magnetometers, gyroscope, GPS receiver, cameras, microphones, light sensors and digital compasses [5]. This study is observed multifactored aspects of HAR including the integration of ML and DL techniques, the critical role of datasets, and their uses [1].

2.2 Various Methods

The domain of HAR has witnessed considerable progress in recent years. with a numerous method being proposed to address this complex challenges. In this section, we study the most recent and relevant literature, with the focus of traditional handcrafted techniques, ML based approaches and DL based models of HAR [14]. Extracting the insights from the major journals and conferences articles [1][3][6].

2.2.1 Traditional Hand-crafted Methods

In the previous HAR in videos methods are heavily on handcrafted feature extraction techniques [4]. In the Traditional handcrafted approaches to HAR often depends on carefully features like motion histograms, optical flow and spatiotemporal interest points to capture the temporal and spatial components of human actions, like Scale-Invariant Feature Transform (SIFT) were used in the combination with the classifiers such as SVM for action recognition [4]. Histograms of Oriented Gradients (HOG) and Global Image Structure (GIST) which extract the similar information from video frames [1]. These methods provided the foundational benchmark for newer advancements and have been extensively studied in the study. This analysis highlights the challenges addressed by these studies and lays the groundwork for our research, which aims to advance human activity recognition by tackling specific, unresolved issues in the field [14].

In these approaches have been widely employed for HAR, they are suffered by the several notable limitations, which have been prompted researchers to shift towards ML Techniques [41]. One key issue is that handcrafted methods depend heavily on domain expertise to manually design feature extraction algorithms, a process that is not only time-consuming but also susceptible to errors. These manually crafted features lack of robustness required to effectively address the variability in the human actions, including the changes in clothing, lighting environments and camera positions [1]. In another way notable challenges are that traditional methods often suffering the overfitting or underfitting, leading to poor generalization when the applied to the hidden data or novel action classes [16]. Finaly, the rigidity of the handcrafted techniques prevents them from adapting the dynamic environments. Has been growing the ML-based approaches, are capable of autonomously extracting the discriminative features directly from the raw data, offering the potential improvements robustness and accuracy of HAR systems. [14] [4].

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

2.2.2 Machine Learning-Based Techniques

In the recent years ML methods have significant prominence in the domain of HAR. In this growth of ML techniques, which are capable to autonomously learning from the data and making accurate predictions [1]. In these capabilities have positioned ML approaches in advancing the state-of-the-art in HAR. In this section offers the complete study of various ML methods applied to the HAR [14].

ML based methods have made extensive contributions to the advancement of HAR. The ML techniques are discussed in this section, diverse strategies, including the feature extraction, fusion approaches and algorithms, all aimed at improve the recognition accuracy and addressing challenges like complex actions, changing the environmental conditions and large-scale datasets. While these methods demonstrated prominent results, there are remains scope for improvement in areas like as robustness, real-time processing and adaptability [1]. In this subsequent, we will delve into DL based techniques, holds the significant potential to further enhance HAR by capabilities of Deep Neural Networks to originally learn and extract high level features directly from the raw data [14]. With this growing availability of large interpreted datasets, ML techniques like Random Forest, K-Nearest Neighbours and SVMs began in the dominate filed. These models learned to recognize the patterns in extracted features without depend on the handcrafting specific characteristics [35]. A notable development was the application of HMM for modeling the temporal sequences in the video data [20]. While these methods improved the performance, they still lack of ability to capture high-level temporal dependencies and complexity of the patterns inherent in actions [5].

2.2.3 Deep Learning-Based Techniques

In these methods have to show significant improvements in HAR, over the performance of traditional handcrafted and ML methods. These advanced techniques capabilities of neural networks to autonomously learn features directly from the video frames, allowing them to capture more and robust representations of the actions [14][1]. The DL methods extend the HAR [20]. The introduction of DL has revolutionized HAR by overcoming the limitations of previous methods, setting a new benchmark for performance and adaptability. In the subsequent sections, we delve deeper into detailed architectures, datasets, and future research directions that can further enhance the capabilities of HAR systems. CNNs serve as the substance for removing spatial features from individual video frames. In addition, LSTM networks and Gated Recurrent Units (GRUs) were widely adopted to model temporal dependencies across frames [15]. Recently, Transformer-based architectures, such as Vision Transformers (ViT) and spatiotemporal transformers, have achieved state-of-the-art results by leveraging consideration mechanisms to focus to the related parts of the video and know long-range dependencies in the data.

2.3 Detection Techniques

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

HAR have to grow significantly with in the integration of DL and ML techniques. In these advancement techniques offers effectively mechanisms for extracting the Deep, meaningful features from the data, leading to the more exact activity recognition. In this study aims to consolidate the recent advancement in detections techniques for HAR, it highlights the Key contributions in the Computer Vision and Explore the applications of ML and DL models in Human activity detection [6].

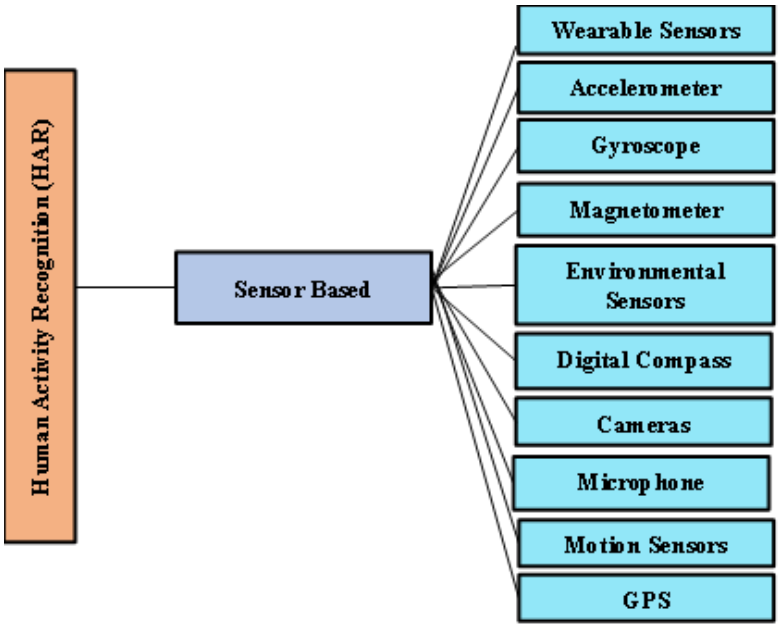

Figure 3: Methods to Detect the HAR in Sensor based System [5],[23], [35].

Figure 3 Shows the various methods of detecting the HAR with sensor-based systems. The summary of HAR based on the sensor data approaches. HAR is a research area dedicated to classifying and analyzing human actions, behavioural with data from the several sensors. These types of sensors are broadly classified into wearable and environmental types, understanding in role of enhancing the accuracy and reliability of human activity recognition systems.

2.3.1 Sensor-Based Approach

In sensor-based approach data are collected from the several sensors to recognize and categorize the human activities. This type of sensor may either wearable attached to the body or embedded in the nearest environment. Smartphones have become a global communication tool and, more recently, a technology for studying humans [26] , [27]. Built-in sensors of smartphones can capture continuous data related human activities [27]. performed the recognition of human activity using smartphone sensors. In this approach, data is retrieved from the smartphone's in-built gyroscope sensor and accelerometer sensors, and then ML

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

techniques were applied to recognize human activity. Making this type of HAR particularly useful for patient monitoring systems, an individual player's activity monitoring during sports, etc., but cannot be applied to the broad application of HAR for security at home/public places, monitoring, etc. [26].

Wearable sensors are fundamental to Human Activity Recognition (HAR) systems, utilizing devices like accelerometers to measure motion or orientation and gyroscopes to track angular velocity for detailed movement analysis. Environmental sensors expand functionality by capturing contextual factors such as temperature, humidity, or light, while cameras provide visual data, enabling image and video-based activity interpretation. Motion sensors detect physical actions like walking, running, or gestures, contributing to accurate activity recognition. HAR systems employ either vision-based methods, leveraging DL and ML models with camera data, or sensor-based approaches using wearable and environmental sensors, with the choice guided by application needs, data availability, and activity complexity [22][23].

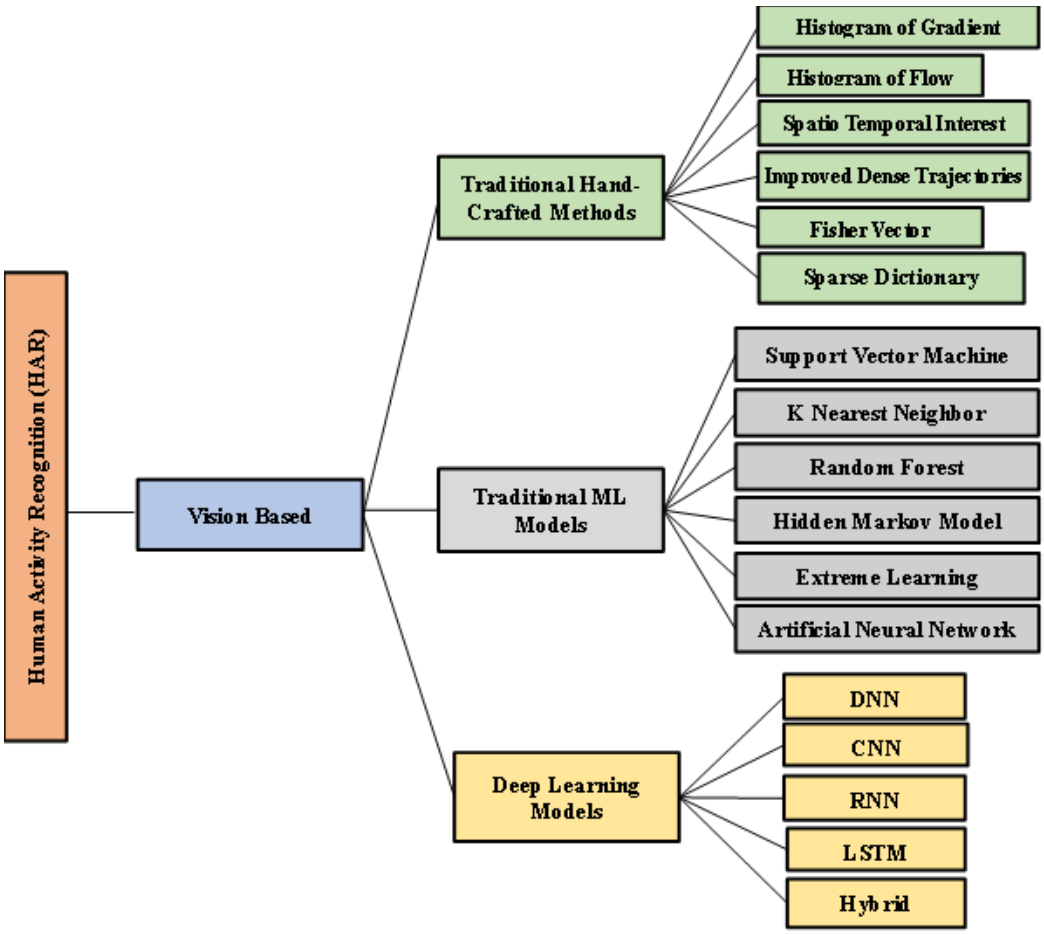

Figure 4: Methods of detecting HAR with vision-based systems [23], [35], [5]

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

Figure 4 Shows that the methods of detecting HAR with vision-based system. The figure outlines a comprehensive framework for HAR, with a particular emphasis on vision-based methodologies. Vision-based HAR techniques are categorized into three major groups: traditional hand-crafted methods, traditional ML and DL models, each offering unique approaches to understanding human actions.

2.3.2 Vision-Based Approach

This approach uses visual data, typically from cameras, to identify activities. The process involves analyzing images or videos to recognize patterns of movement. Videos are captured by static cameras positioned across locations for purposes of surveillance, which are then stored on servers. These captured videos or camera feeds are subsequently used for surveillance, performed human posture recognition for video surveillance applications using one static camera. This type of HAR is used for road safety, public security, traffic management, crowd monitoring, etc [26]. Apply the computer vision techniques like pose estimation and object tracking, it allows to identification of human activities in video or image data [20].

2.3.2.1 Traditional Hand-Crafted Methods

HAR are mainly designed for the manually features extracted from the pictorial data. These are including techniques like Histogram of Gradients (HOG), captures the gradient-based features for object representation and Histogram of Flow (HOF), focuses on analyzing the motion patterns to detect the dynamic activities. Other techniques like, Spatio-Temporal Interest Points (STIP), focuses on the identifying regions of significant motion or changes over time, in the detection of complex activities. Next Improved Dense Trajectories (IDT) provides the motion trajectories for the more accurate activity detection. Fisher Vectors and Sparse Dictionary Learning Techniques improves the features representation and encoding.

2.3.2.2 Traditional ML Models

It utilizes these hand-crafted features to perform the activity recognition. In this commonly used algorithms SVMs, in binary classification, and KNN is straightforward approach based on feature space. Random Forest is an ensemble learning method, and HMM in these model sequential data, also play in vital role. Additionally, techniques like ELM and ANN enhance the modeling the capabilities of these system.

2.3.2.3 Deep Learning Models

It has been enabling to the extraction features directly from the raw data. DNNs extract the hierarchical feature representations, while CNNs specialize the spatial feature learning from images or videos. RNNs modeling the temporal dependencies, making them ideal for sequential data. LSTM networks further enhance temporal modeling by addressing long-range dependencies. Hybrid models that integrate these architectures, offer even more robust solutions for complex activity recognition tasks. This framework highlights the evolution of

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

vision-based HAR from traditional feature engineering to modern data-driven learning approaches, showcasing their respective strengths in accurately recognizing and interpreting human activities.

3. Strategy-based activities

In the real-time systems like surveillance and monitoring, HAR can operates in two primary modes: live stream(online) and pre-recorded (offline). When the working online, HAR processes activities directly from a live video to identify actions in real-time. This model is particularly effective in scenarios that require immediate detection and response, like as in security monitoring or alerting systems. On further, offline HAR works with stored video footage, enabling a detailed and comprehensive analysis without the constraints of real-time processing. It is commonly employed for reviewing past events, conducting forensic investigations, or analyzing behavioural patterns over time. While live HAR ensures timely interventions, offline HAR provides deeper insights and detailed examinations when immediacy isn't a priority [2].

3.1 Offline Strategy-Based Activities

It focuses on analyzing pre-recorded videos, offering the flexibility to apply computationally intensive models and techniques for in-depth analysis. Conversely, online HAR emphasizes real-time processing, decision-making. The modality source in HAR can be categorized into unimodal and multimodal approaches. Unimodal methods rely on a single input modality, such as video or sensor data, for activity recognition. [11]. Multimodal approaches integrate inputs from various modalities capturing nuanced details and improving recognition accuracy. Multimodal systems compensate for the drawback of individual modalities by combining matching information, leading to more robust and reliable HAR systems [20].

3.2 Live Streaming Strategy-Based Activities

The live streaming typically consists of video data captured by sensors or cameras that capture human movements and actions in real-time. The HAR model then processes this data, which analyses and classifies the activities being performed, enabling the VR/AR applications to respond accordingly or self-driving cars to make informed decisions based on the detected activities. This is particularly useful in environments where decisions must be made instantaneously, such as real-time sports events or live strategy games [2].

Jalal et al. [8] used Depth Differential Silhouettes (DDS) and human temporal points to identify online activity, considering skeletal joint characteristics. They used code vectors to reduce computational complexity and used a machine learning algorithm to classify online activities based on extracted features. This approach allowed for real-time monitoring and

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

recognition of specific online behaviors, facilitating a deeper understanding of user engagement and interaction patterns.

Zolfaghari et al. [7] advanced a lightweight algorithm for real-time activity recognition using HMM and depth maps. This approach suits resource-constrained environments and reduces the data required for accurate predictions. Combining the 2D and 3D networks in the ECO architecture allows to the achieve a more comprehensive understanding of human activity by utilizing both temporal and spatial material. Collecting frames from both the current sequence and the following series makes the predictions generated more accurate and efficient, reducing computational complexity and minimizing data overhead. This approach enhances the real-time online activity identification process.

4. Datasets analysis for recognizing human actions in video data

In the realm of computer vision, datasets are essential for together training and testing/validation purposes. With increasing prominence of DL models, a large-scale dataset is necessary to effectively train and evaluate action recognition tasks [35].

Table 4: An overview of datasets for general action recognition tasks.

Ref	Year	Datasets	Key Features	#Classes | Actions	#Instances | Clips	Content
[36] [16]	2012	UCF101	Challenging actions, Diverse actions, real-world settings, RGB-only videos.	101	13K	Human actions, Sports & Daily Activities
[16]	2013	UCF50	Real world challenges, varied viewpoints, camera motion, object scale, cluttered backgrounds.	50	6.7K	Daily Activities, Sports, Human object interactions, microscopic Actions.
[39]	2015	Activity Net	Temporal activity detection, untrimmed videos.	203	28K	Wide range of daily activities
[40]	2016	THUMOS' 14	Temporal Action Detection, Large-scale Dataset, Multiple Action Classes, Diverse Data Sources.	20	2.6K	Video segments, validation Data, Annotations.
[41]	2016	Charades	Large-Scale Dataset, Multi-Label Classification, Segment-Level Features.	157	9.8K	Feature Embeddings, Labels.
[41]	2016	YouTube-8 M	Large-Scale Dataset, Precomputed Features, Diverse Label Set.	3862	8M	Video Features, Labels, Pretrained Models.
[43]	2016	NTU RGB+D	Multimodal Data, Large Scale, Diverse Action Classes, High Quality.	60	56K	Data Modalities, Actions, Subjects, Camera Viewpoints.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

[37]	2017	Multi-THUMOS	Dense temporal annotations for untrimmed videos, multi-label, frame-level action	65	38.7K	multi-label action recognition in realistic scenarios.
[41]	2017	Hollywood 3D	Stereoscopic 3D Data, Real – World Scenarios, Multi-Modal Data.	14	0.7K	Videos, Action Classes, 3D Features
[44]	2017	Kinetics400	Large Scale, Diverse Action Classes, High-Quality, Annotations, Short Clips.	400	306K	Sports and Games, Daily Activities, Social Interactions.
[48]	2017	Something-Something	Action Diversity, Fine-Grained, Large Scale, Crowd-Sourced Data, Short Clips.	174	108.5K	Videos, Action Classes.
[46]	2018	Atomic Visual Actions	Atomic Actions, Spatio-Temporal Localization, High-Quality Annotations, realistic scenarios.	80	385K	Action Annotations, Action Classes.
[49]	2018	Kinetics600	Expanded Action Classes, Large-scale Dataset, high Quality.	600	482K	Video Clips, Action Categories, Annotations.
[47] [43]	2020	NTU RGB+D 120	Large-Scale Action Recognition, Multi-Modality, Realistic and Diverse Data, 3D Skeleton data.	120	114K	Multi-Modal Data, Daily activities, Social Interactions.
[21]	2020	AVA-Kinetics	Hybrid Dataset, Acton Categories, Spatio-Temporal Annotations.	90	624K	Videos, Action Classes, Anatomical Features.
[15]	2022	CZU-MHAD	Multi-Modal data, Multi-Person Action Recognitions, Sensor Fusion, Real World Settings.	22	1.5K	Sensors, Annotations, Action Classes.
[14] [16]	2023	HMDB51	Challenging with noise and low quality, RGB-only videos.	51	6.8K	Movies, web videos.
[44]	2024	CAPTURE-24	Axivity AX3 wrist-worm activity trackers.	20	3K	Accelerometer, Annotations.
[37]	2024	Drone Action	Sensors, Video Frames, motion trajectories, object detection.	13	-	Surveillance, crowded scenes, variations.
[16]	2024	VAID	RGB Cameras, LIDAR, Drones.	4	10K	Vechicle Behaviors, behaviour prediction.

The above Table 4 shows that the overview of the numerous datasets for general action recognition tasks. Some of the key features of datasets, actions containing the dataset, clips of the dataset and content of the datasets.

5. Applications and Future Directions

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

It finds the applications of various areas such as smart homes, elder care, fitness and more. [20]. The below Figure 5. Shows that the classification of HAR with diverse types of applications.

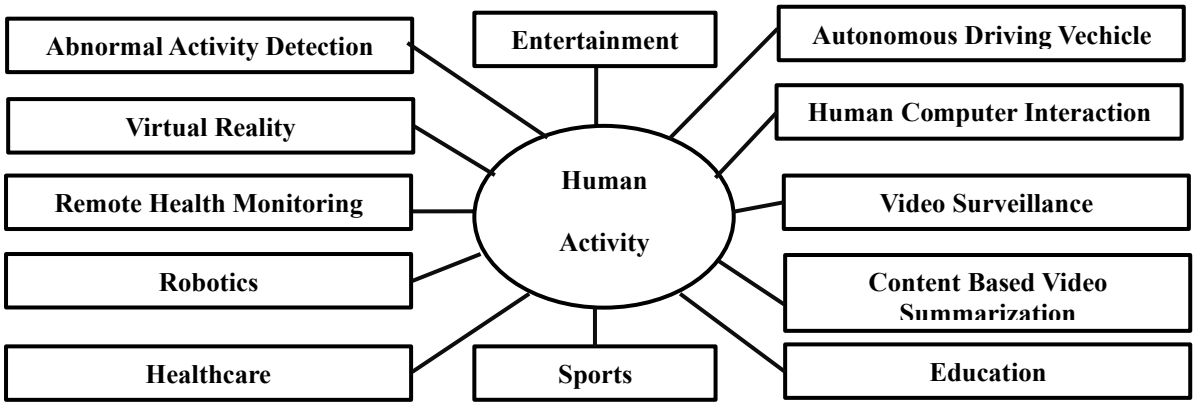

Figure 5: Classification of HAR with diverse types of applications used.

The various applications of HAR, shows the significance of various domains. In Healthcare, HAR system facilitates the remote health monitoring system, assisting the patient care and chronic disease management, it also supporting the robotics in automating healthcare services. In virtual reality and entertainment, enhancing the immersive experiences and interactive environments. The HAR System finds the utility in abnormal activity detection, crucial for security and safety applications. HAR contributes to autonomous driving vehicles by enabling the systems to interpret human actions, improving the decision making and safety management. It also plays a key role in Human-computer interaction. In education, HAR supports innovative learning methods and engagement strategies. Sports benefit from HAR by enabling the performance analysis and injury prevention [24], In content-based video summarization streamlines data processing and retrieval [1]. HAR's critical role in advancing the modern technological solutions [23],[22].

6. Result Analysis

To evaluate the robustness and effectiveness of the model, a set of performance metrics, along with the various validation techniques. To ensures that in-depth understanding of the system's performance.

Performance Metrices

- **Accuracy**

Accuracy measures the proportion of the total predictions that are correctly identified, serving as a fundamental metric in classification tasks. It is calculated using Equation (1):

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Where TP True Positive, TN True Negative, FP False Positive and FN False Negative.

- **Precision**

Precision quantifies the model's ability to accurately identify positive instances. It is the ratio of true positives to the total number of predicted positives, as expressed in Equation (2):

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

- **Recall (Sensitivity)**

Recall, also known as sensitivity, evaluates the model's ability to identify all actual positives instances. It is the ratio of true positives to the sum of true positives and false negatives. As expressed in Equation (3):

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

- **F1-Score**

F1-Score delivers a balanced measures between precision & recall are combined using the F1 score, which is the harmonic mean of precision & recall, calculated as Equation (4):

$$\text{F1-score} = 2 \text{ x } \frac{Precision * Recall}{Precision+Recall} \quad (4)$$

- **Confusion Matrix**

Confusion matrix presents a comprehensive picture of the classification algorithm's performance. Each column represents instances in a predicated class, while each row corresponds to instances in actual class. The diagonal elements show the number of the correct predictions.

Table 5: Comparison of Different Techniques for Human Activity Recognition.

Title	Author Name	Year	Dataset	Techniques	Performance Measure	Limitations	Results
A Hybrid Approach for HAR with SVM and 1D Convolutional Neural Network [32]	M. M. Hossain Shuvo, N. Ahmed, K. Nouduri and K. Palaniappan	2020	UCI-HAR Dataset	Random Forest (RF), SVM, 1D-Convolutional Neural Network (CNN)	The total training time taken for 100 epochs is 170 seconds.	For identification in real time, deploying this to low-power integrated circuits for wearable sensors.	The accuracy percentage was 97.71%.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

Smartphone Based Human Activity Recognition with Feature Selection and Dense Neural Network [33]	Bashar SK, Al Fahim A, Chon KH	2020	UCI-HAR Dataset, Six daily activities.	Neighbourhood Component Analysis, DNN, SVM Four Hidden Layers.	The proposed model outperformed many other deep learning methods.	Exploring other feature selection methods might also be helpful	The accuracy percentage was 95.79%.
Implementation of an Anomalous HAR System [34].	Shreyas, D.G., Raksha, S. & Prasad, B.G	2020	UCF10 1 crime dataset with around 13 Different anomalies	Adaptive Video Compression, CNN, Anomaly Detection.	The proposed system has more accurate favourable rates than the other existing systems	In reality, the chances of occurrences of abnormal activity are meager, so the performance is measured in standard videos without anomalies, as having a low false alarm rate is challenging	This approach outperforms previous approaches regarding accuracy and timeliness
Self-supervised Learning for HAR Using 700,000 Person-days of Wearable Data [16].	Hang Yuan, Shing Chan, Andrew P. Creagh, Catherine Tong, Aidan Acquah, David A	2024	UK-Biobank dataset, containing over 700,000 person-days	Multi-task self-supervised learning (SSL),	The SSL models consistently outperformed baselines, showing a relative improvement of 2.5% to 100% in F1 scores.	The primary limitation was the demographic homogeneity of the UK-Biobank dataset.	Pre-trained models generalized well across diverse datasets, achieving significant performance gains, particularly on smaller datasets.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

Recognizing Human Activities with the Use of Convolutional Block Attention Module [16]	Mohammed Zakariah and Abeer Alnuaim	2024	HMDB51, UCF-101, and UCF-50	Convolutional Block Attention Module (CBAM) integrated with 3D CNN for (HAR)	Accuracy: 94.23% on HMDB51 Accuracy: 83.4% on UCF-101 Accuracy: 88.9% on UCF-50	The adaptability of CBAM to diverse CNN architectures were not extensively studied. Limited validation beyond controlled datasets.	CBAM enhances feature extraction and classification capabilities.
CAPTURE-24: A large dataset of wrist-worn activity tracker data collected in the wild for human activity recognition [31].	Shing Chan, Yuan Hang, Catherine Tong, Aidan Acquah, Abram Schonfeldt, Jonathan Gershuny,	2024	CAPTURE-24, 2562 hours of annotated data collected from 151 participants.	Random Forest (RF), XGBoost, CNN, Recurrent Neural Networks (RNN), and Hidden Markov Models (HMM).	Best performance: CNN + HMM achieved the highest scores across multiple tasks, showing the importance of temporal modeling	Limited demographic diversity as data collection was confined to Oxford participants. Wearable camera data excluded from public release due to privacy concerns.	The dataset is significantly higher and more representative than existing datasets, enabling better generalization for real-world applications
A Graph-based Approach to HAR [24].	Thomas Peroutka, Ilir Murturi, Praveen Kumar Donta, Schahram Dustdar	2024	Biathlon Dataset, Collected using wearable IMU sensors and GNSS receivers.	A graph-based approach for encoding human movements as directed graphs to analyze temporal dependencies and detect complex motion sequences	The approach achieved efficient processing times of 1-2 milliseconds on huge datasets with high granularity	Requires extensive domain-specific knowledge to define and encode movements. Movement variations among individuals may pose challenges for generalization.	Demonstrated accurate and efficient detection of movement sequences in biathlon scenarios. he approaches efficiently handled large, high-resolution datasets.

7. Conclusion

This study comparative analysis of Support Vector Machines (SVM), Convolutional Neural Networks (CNN), k-Nearest Neighbors (k-NN) for Human Action Recognition in

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

videos. HAR has crucial technology in fields of video surveillance, healthcare, human-computer interaction. The challenges are occlusion, varying viewpoints, and complex action dynamics. Our findings emphasize that CNNs outperform SVM and k-NN in accuracy and robustness, especially for the datasets with high variability and complexity. Future research should concentrate on integrating hybrid models that combine the effectiveness of traditional techniques with the deep learning. HAR systems may be further optimized to achieve greater adaptability and precision in real-world applications.

8. Acknowledgements

This research was supported by the Visvesvaraya Technological University, Belagavi under the Jnana Yaana Doctoral Fellowship (VTU-JYDF) program. The authors are grateful for the university's financial assistance and facilities provided during this work.

References:

[1] Pareek, P., Thakkar, A. *A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications*. Artif Intell Rev 54, 2259–2322 (2021). https://doi.org/10.1007/s10462-020-09904-8.

[2] Bukht, T. F. N., Rahman, H., Shaheen, M., Algarni, A., Almujally, N. A., & Jalal, A. (2024). *A review of video-based human activity recognition: theory, methods and applications.* Multimedia Tools and Applications. https://doi.org/10.1007/s11042-024-19711-w.

[3] Kong, Y., Fu, Y. *Human Action Recognition and Prediction: A Survey.* Int J Comput Vis 130, 1366–1401 (2022). https://doi.org/10.1007/s11263-022-01594-9

[4] Shuchang Zhou, *A Survey on Human Action Recognition,* (cs.CV) (2022) https://doi.org/10.48550/arXiv.2301.06082

[5] Thakur, D., Biswas, S. *Smartphone based human activity monitoring and recognition using ML and DL: a comprehensive survey.* J Ambient Intell Human Comput 11, 5433–5444 (2020). https://doi.org/10.1007/s12652-020-01899-y

[6] Rahul Kumar, Shailender Kumar. *A survey on intelligent human action recognition techniques.* Multimedia Tools Appl., 83(17):52653-52709, 2024. DOI:10.1007/s11042-023-17529-6

[7] Zolfaghari, M., Singh, K., Brox, T. (2018). ECO: Efficient Convolutional Network for Online Video Understanding. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science(), vol 11206. Springer, Cham. https://doi.org/10.1007/978-3-030-01216-8_43.

[8] Jalal, A., Kim, Y., Kim, Y., Kamal, S., & Kim, D. (2017). Robust human activity recognition from depth video using spatiotemporal multi-fused features. *Pattern Recognition., 61*, 295-308.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

[9] T, Gayathri & hr, Mamatha. (2024). How to Improve Video Analytics with Action Recognition: A Survey. ACM Computing Surveys. 57. 1 https://doi.org/10.1145/3679011

[10] Pham, Hieu & Khoudour, Louahdi & Crouzil, Alain & Zegers, Pablo & Velastin, Sergio. (2022). *Video-based Human Action Recognition using Deep Learning: A Review*. 10.48550/arXiv.2208.03775.

[11] Y. Wang, Z. Qi, X. Li, J. Liu, X. Meng and L. Meng, "Multi-channel Attentive Weighting of Visual Frames for Multimodal Video Classification," *2023 International Joint Conference on Neural Networks (IJCNN)*, Gold Coast, Australia, 2023, pp. 1-8, doi: 10.1109/IJCNN54540.2023.10192036.

[12] Valter Estevam, Helio Pedrini, David Menotti, Zero-shot action recognition in videos: A survey, Neuroc omputing, Volume 439, 2021, Pages 159-175, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2021.01.036.

[13] Cheng, Guangchun, Yiwen Wan, Abdullah N. Saudagar, Kamesh Namuduri and Bill P. Buckles. "Advances in Human Action Recognition: A Survey." *ArXiv* abs/1501.05964 (2015): n. pag.

[14] El Mehdi Saoudi, Jaafar Jaafari, Said Jai Andaloussi, Advancing human action recognition: A hybrid approach using attention-based LSTM and 3D CNN, Scientific African, Volume 21, 2023, e01796, ISSN 2468-2276, https://doi.org/10.1016/j.sciaf.2023.e01796.

[15] X. Chao, Z. Hou and Y. Mo, "CZU-MHAD: A Multimodal Dataset for Human Action Recognition Utilizing a Depth Camera and 10 Wearable Inertial Sensors," in *IEEE Sensors Journal*, vol. 22, no. 7, pp. 7034-7042, 1 April1, 2022, doi: 10.1109/JSEN.2022.3150225.

[16] Yuan H, Chan S, Creagh AP, Tong C, Acquah A, Clifton DA, Doherty A. Self-supervised learning for human activity recognition using 700,000 person-days of wearable data. NPJ Digit Med. 2024 Apr 12;7(1):91. doi: 10.1038/s41746-024-01062-3. PMID: 38609437; PMCID: PMC11015005.

[17] M. Hanzla, S. Ali and A. Jalal, "Smart Traffic Monitoring through Drone Images via Yolov5 and Kalman Filter," *2024 5th International Conference on Advancements in Computational Sciences (ICACS)*, Lahore, Pakistan, 2024, pp. 1-8, doi: 10.1109/ICACS60934.2024.10473259.

[18] Mohammed Zakariah, Abeer Alnuaim, Recognizing human activities with the use of Convolutional Block Attention Module, Egyptian Informatics Journal, Volume 27, 2024, 100536, ISSN 1110-8665, https://doi.org/10.1016/j.eij.2024.100536.

[19] Thi TH, Zhang J, Cheng L, Wang L, Satoh S (2010) *Human action recognition and localization in video using structured learning of local space–time features.* In: 2010 seventh IEEE international confer ence on advanced video and signal-based surveillance (AVSS). IEEE, pp 204–211.

[20] sedaghati, N., ardebili, S. & Ghaffari, *A. Application of human activity/action recognition: a review.* Multimed Tools Appl (2025). https://doi.org/10.1007/s11042-024-20576-2.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

[21] Simonyan, K., & Zisserman, A. (2014). "Two-stream convolutional networks for action recognition in videos." *Advances in Neural Information Processing Systems*, 568-576.

[22] Dalal, N., & Triggs, B. (2005). "Histograms of oriented gradients for human detection." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 886-893.

[23] Xie, L., & Lu, H. (2015). "Human activity recognition based on frequency analysis." *Journal of Electrical Engineering & Technology*, 10(4), 1329-1337.

[24] Peroutka, Thomas & Murturi, Ilir & Donta, Praveen Kumar. (2024). A Graph-based Approach to Human Activity Recognition. 10.48550/arXiv.2408.10191.

[25] B, Jagadeesh and Chandrashekar M Patil. "Video Based Human Activity Detection, Recognition and Classification of actions using SVM." *Transactions on Machine Learning and Artificial Intelligence* 2019 Jan 5;6(6):22.

[26] Vijeta Sharma, Manjari Gupta, Anil Kumar Pandey, Deepti Mishra & Ajai Kumar (2022) A Review of Deep Learning-based Human Activity Recognition on Benchmark Video Datasets, Applied Artificial Intelligence, 36:1, 2093705, DOI: 10.1080/08839514.2022.2093705.

[27] Wan, S., Qi, L., Xu, X. *et al.* Deep Learning Models for Real-time Human Activity Recognition with Smartphones. *Mobile Netw Appl* 25, 743–755 (2020). https://doi.org/10.1007/s11036-019-01445-x.

[28] Mohammed Mehedi Hassan, Md. Zia Uddin, Amr Mohamed, Ahmad Almogren, A robust human activity recognition system using smartphone sensors and deep learning, Future Generation Computer Systems, Volume 81, 2018, Pages 307-313, ISSN 0167-739X, https://doi.org/10.1016/j.future.2017.11.029.

[29] Lukas Koping, Kimiaki Shirahama, Marcin Grzegorzek, A general framework for sensor-based human activity recognition, Computers in Biology and Medicine, Volume 95, 2018, Pages 248-260, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2017.12.025.

[30] Bota, P.; Silva, J.; Folgado, D.; Gamboa, H. A Semi-Automatic Annotation Approach for Human Activity Recognition. *Sensors* 2019, *19*, 501. https://doi.org/10.3390/s19030501

[31] A. S. A. Sukor, A. Zakaria and N. A. Rahim, "Activity recognition using accelerometer sensor and machine learning classifiers," *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*, Penang, Malaysia, 2018, pp. 233-238, doi: 10.1109/CSPA.2018.8368718.

[32] M. M. Hossain Shuvo, N. Ahmed, K. Nouduri and K. Palaniappan, "A Hybrid Approach for Human Activity Recognition with Support Vector Machine and 1D Convolutional Neural Network," *2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, Washington DC, DC, USA, 2020, pp. 1-5, doi: 10.1109/AIPR50011.2020.9425332.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

[33] Bashar SK, Al Fahim A, Chon KH. Smartphone Based Human Activity Recognition with Feature Selection and Dense Neural Network. Annu Int Conf IEEE Eng Med Biol Soc. 2020 Jul; 2020:5888-5891. doi: 10.1109/EMBC44109.2020.9176239. PMID: 33019314.

[34] Shreyas, D.G., Raksha, S. & Prasad, B.G. Implementation of an Anomalous Human Activity Recognition System. *SN COMPUT. SCI.* 1, 168 (2020). https://doi.org/10.1007/s42979-020-00169-0.

[35] Yin, H., Sinnott, R.O. & Jayaputera, G.T. A survey of video-based human action recognition in team sports. *Artif Intell Rev* 57, 293 (2024). https://doi.org/10.1007/s10462-024-10934-9.

[36] Soomro, K., Zamir, A., & Shah, M. (2012). UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild. *ArXiv, abs/1212.0402*.

[37] Y. Abbas and A. Jalal, "Drone-Based Human Action Recognition for Surveillance: A Multi-Feature Approach," *2024 International Conference on Engineering & Computing Technologies (ICECT)*, Islamabad, Pakistan, 2024, pp. 1-6, doi: 10.1109/ICECT61618.2024.10581378.

[38] Yeung, S., Russakovsky, O., Jin, N. *et al.* Every Moment Counts: Dense Detailed Labeling of Actions in Complex Videos. *Int J Comput Vis* 126, 375–389 (2018). https://doi.org/10.1007/s11263-017-1013-y.

[39] F. C. Heilbron, V. Escorcia, B. Ghanem and J. C. Niebles, "ActivityNet: A large-scale video benchmark for human activity understanding," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 961-970, doi: 10.1109/CVPR.2015.7298698.

[40] Idrees H, Zamir AR, Jiang Y-G, Gorban A, Laptev I, Sukthankar R, Shah M (2016) The THUMOS challenge on action recognition for videos "in the Wild". CoRR 155, 1–23. arXiv: 1604.06182.

[41] Yang, X., Zhang, C., & Tian, Y. (2012). "Recognizing actions using depth motion maps-based histograms of oriented gradients." *Proceedings of the 20th ACM International Conference on Multimedia*, 1057-1060.

[42] Sigurdsson GA, Varol G, Wang X, Farhadi A, Laptev I, Gupta A (2016) Hollywood in homes: crowdsourc ing data collection for activity understanding. CoRR abs/1604.01753, 510–526. arXiv: 1604.01753.

[43] A. Shahroudy, J. Liu, T. -T. Ng and G. Wang, "NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 1010-1019, doi: 10.1109/CVPR.2016.115.

[44] Chan, S., Hang, Y., Tong, C. *et al.* CAPTURE-24: A large dataset of wrist-worn activity tracker data collected in the wild for human activity recognition. *Sci Data* 11, 1135 (2024). https://doi.org/10.1038/s41597-024-03960-3.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 22-05-2025

[45] Kay W, Carreira J, Simonyan K, Zhang B, Hillier C, Vijayanarasimhan S, Viola F, Green T, Back T, Natsev P, Suleyman M, Zisserman A (2017) The kinetics human action video dataset. CoRR abs/1705.06950. arXiv: 1705.06950.

[46] Gu C, Sun C, Vijayanarasimhan S, Pantofaru C, Ross DA, Toderici G, Li Y, Ricco S, Sukthankar R, Schmid C, Malik J (2017) AVA: a video dataset of spatio-temporally localized atomic visual actions. CoRR abs/1705.08421, 6047–6056. arXiv: 1705.08421.

[47] J. Liu, A. Shahroudy, M. Perez, G. Wang, L. -Y. Duan and A. C. Kot, "NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2684-2701, 1 Oct. 2020, doi: 10.1109/TPAMI.2019.2916873.

[48] Goyal R, Kahou SE, Michalski V, Materzynska J, Westphal S, Kim H, Haenel V, Fründ I, Yianilos P, Muel ler-Freitag M, Hoppe F, Thurau C, Bax I, Memisevic R (2017) The “something something” video database for learning and evaluating visual common sense. CoRR abs/1706.04261, 5842–5850. arXiv: 1706.04261.

[49] Carreira J, Noland E, Banki-Horvath A, Hillier C, Zisserman A (2018) A Short Note about Kinetics-600. CoRR abs/1808.01340. arXiv:1808.01340.