



Original Article

# Vision-Based Human Action Recognition Using Skeleton Graph Neural Networks

Sajud Hamza Elinjulliparambil  
Pace University, United States.

**Abstract** - HAR is an essential computer vision task, which provides the machine with an automatic representation of human actions and human action classification on the basis of visual representations. Older methods of HAR are based on RGB or RGB-D videos, and, in spite of their success in controlled settings, they are prone to occlusions, variations in lighting, and a high level of computational complexity. HAR based on skeletons has been proposed as an alternative, a representation of human poses in terms of joints or keypoints and tracking both spatial and temporal behavior of movement. This abstraction enhances resiliency to visual noise, lowers computational costs, and supports real time usage. Graph Neural Networks (GNNs), and in particular Graph Convolutional Networks (GCNs), have proven to have a phenomenal success in skeleton-based HAR, by considering the human body as a spatial-temporal graph with joints representing nodes and bones or temporal relationships as edges. The paper provides a systematized review of the vision-based human action recognition based on skeleton GNNs through skeleton representation, graph construction and development, early and advanced GNN architecture, and how they are enhanced by adaptive graphs, attention, and hierarchical modeling. Some of the most critical issues, including skeleton noise, time mismatch, generalization across views, over-smoothing in deep GCNs, and computational complexity, are addressed, as well as solutions to these problems. Also, we emphasize surveillance, healthcare, human-computer interaction, and robotics. Lastly, the future research directions are suggested, which are multi-modality, lightweight models of edge devices, self-monitored learning, and cross-dataset generalization. The current review offers a solid background to researchers who want to come up with strong, effective, and understandable skeleton-based systems of HAR.

**Keywords** - Human Action Recognition (HAR), Skeleton-Based Action Recognition, Graph Neural Networks (Gnns), Self-Supervised Learning.

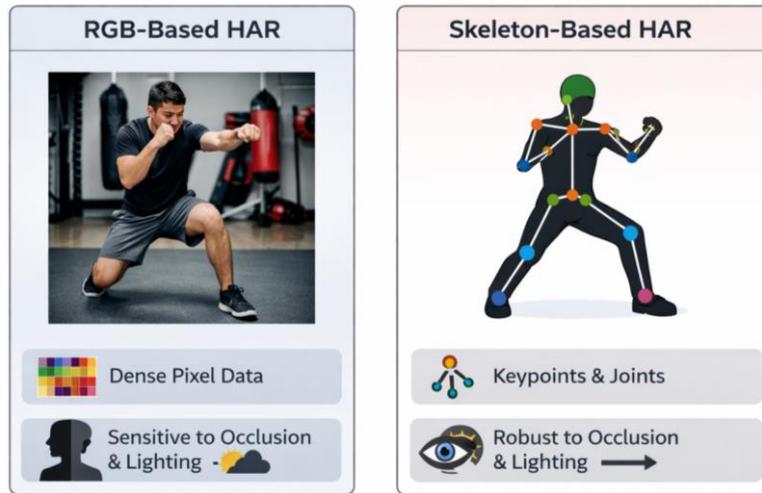
## 1. Introduction

Human Action Recognition (HAR) is a basic issue in computer vision and it seeks to automatically detect and subclassify human actions on the basis of visual data [1]. Within the last few decades, HAR has received considerable interest because of its broad scope of application in surveillance, human-computer interface (HCI), medical surveillance, and robotics. In surveillance, HAR allows identifying suspicious or anomalous behavior automatically, which enhances the safety of the population, and human monitoring is eliminated [2]. In HCI and gaming, user recognition makes it possible to have simply controllable user interfaces and simulation. Action recognition is used to monitor patients, rehabilitation exercises, and falls in elderly care within healthcare. Equally, HAR is imperative in the field of robotics where it is applied to teach robots about human behavior and learn to copy them as a method of working together and providing services [3]. This wide applicability has inspired a lot of research on how to come up with strong, effective and correct recognition systems.

The most common data source is the traditional HAR which uses RGB or RGB-D video sequences [4]. Such techniques retrieve visual representations of motion trajectory, optical flow or deep convolutional representation of raw images. Although RGB-based has been successful in some measure, it has been constrained by environmental factors like occlusion, change in illumination and background clutter. As an illustration, appearance-based features are not reliable when body parts are covered or the difference in light conditions is very high, thus resulting in misclassification. In addition, RGB-based algorithms frequently need significant computing power because image data has a large dimensionality which prevents real-time processing and execution on edge computing. Therefore, RGB-based HAR is still useful in controlled conditions, but other methods are required to be able to handle robust recognition in the real world [5].

Skeleton-based HAR has proved to be one of the promising ones that overcome most drawbacks of RGB-based methods. Skeleton-based methods describe human pose by a collection of keypoints or joints, usually at anatomical points like the head or shoulders, elbows and knees [6]. Skeleton-based representations are by definition resistant to the changes in clothing, lighting and context, owing to the fact that appearance information is abstracted. In addition, skeleton data are small and therefore computationally efficient with important structural and temporal data regarding human motion. Due to the high-quality pose estimation algorithms, research in this field has been faster as it allows skeleton data to be obtained easily and

with high accuracy. The representation of skeletons includes not only spatial configurations of joints but also temporal dynamics, which offers a structured and abundant framework of HAR tasks.



**Fig 1: Comparison of RGB-Based and Skeleton-Based HAR.**

As shown in Figure 1, RGB-based methods rely on dense pixel information, whereas skeleton-based methods represent the human body using keypoints and skeletal structures. This abstraction enables efficient processing and robustness to visual noise.

Recently, the concept of Graph Neural Networks (GNNs) has become a promising solution to skeleton-based HAR because it allows modeling non-Euclidean data structures. Human joints may be regarded as nodes, and bones or temporal relations between them may be regarded as edges, which results in the graph representation of motion. GNNs, and especially Graph Convolutional Networks (GCNs), are a type of information propagation through nodes, preserving the structure of the graph, allowing models to acquire local and global joint dependency. The GNNs capture intricate spatial interactions and compound interactions in a better way compared to other traditional sequence-based models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory networks (LSTMs). With the help of GNNs, scientists have managed to build skeleton-based HAR models that have high accuracy, occlusion resistance, and effective temporal modeling.

The primary aim of the review is to review the literature on the topic of vision-based human action recognition with skeleton graph neural networks, with the emphasis on methods developed. In particular, this paper will set out to:

- Present an overview of skeleton-based HAR approaches and their advantages over RGB-based methods.
- Introduce graph neural networks and their adaptation to skeleton data for HAR.
- Review key GNN architectures for skeleton-based HAR, including spatial, temporal, and adaptive graph models.

## 2. Background

It is human action recognition based on skeleton data (Human Action Recognition, HAR) that has emerged as a key feature because of its ability to help characterize human motion effectively as well as its insensitivity to visual noise, occlusion, and morphological changes. The skeleton-based methods are based on the structural data on the human body and it is posed in the form of a set of keypoints or joints [7]. Such joints record spatial arrangement of the body, which can be linked in time to describe motion dynamics. Skeleton data is also compact in nature, and as such can be used in real-time settings and deep learning models (including Graph Neural Networks (GNNs)). This section is dedicated to the basics of skeleton representation, the introduction to GNNs, and popular skeleton datasets.

### 2.1. Skeleton Representation

Representation of skeletons is the basis of human action recognition as it records the vital structural and movement patterns of the human body. It is concerned with coding of the location of joints and relationships that are applicable in modeling spatial and time dynamics of human activities.

### 2.2. 2D vs 3D Joint Coordinates

Skeletons may be modeled in 2D coordinates or 3D coordinates. Under 2D representation, a joint is characterized by the pixel position in the image plane. As much as this method is easier and usable with RGB videos it is susceptible to camera perspective and depth ambiguity. Conversely, 3D representation describes the spatial position of joints in 3D space to give a

more realistic and view-invariant description of human motion 3D skeletons are commonly computed with depth sensors or by multi-view camera pose estimation algorithms [8].

### 2.3. Temporal vs Spatial Information

Skeleton-based HAR relies on two complementary types of information:

- Spatial information: Captures the structural relationships between joints within a single frame, such as distances or angles between connected joints. This helps recognize postures and configurations.
- Temporal information: Models the motion patterns of joints across consecutive frames, which is critical for distinguishing actions that may share similar static poses but differ in dynamics. Temporal modeling can be performed using recurrent models, convolutional networks, or graph-based approaches.

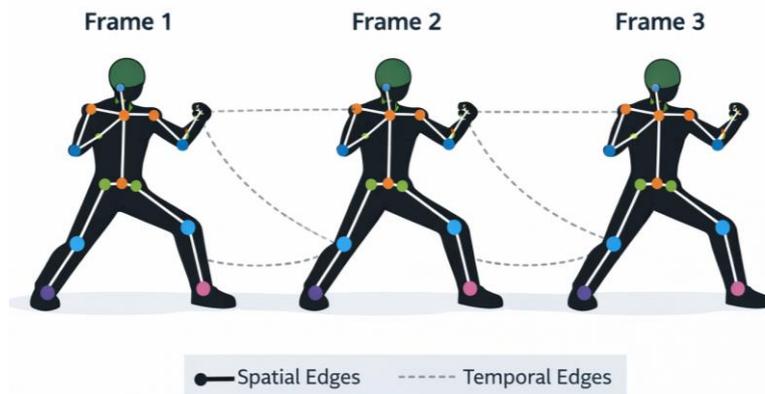
The representation of skeletons allows good representation of human movement and still provides vital structural and time information, so it can be used in graph-based deep learning approaches.

### 2.4. The principles of graph neural networks.

GNNs are especially applicable to skeleton-based HAR as skeletons are non-Euclidean. This is in contrast to grid-like image data wherein skeletons are made up of joints linked haphazardly by using bones, and thus the conventional convolution operations are inappropriate. GNNs can apply deep learning to this type of graph-structured data, enabling joint features to be diffused along edges without disturbing the underlying structure [9].

#### 2.4.1. Graph Construction for Skeletons

A graphical representation of a skeleton has nodes that are joints and edges that are bones that join two joints. It is common to add the temporal relationship between similar joints in different frames, to represent motion dynamics. This space-time graph is able to capture the structural and temporal dependencies of the body and it forms the basis of graph-based HAR [10].



**Fig 2: Skeleton Graph Representation.**

Nodes correspond to joints, edges correspond to bones, and temporal edges connect the same joints across frames. As illustrated in Figure 2, the spatial-temporal graph captures both the connectivity between joints within a frame and their evolution over time.

#### 2.4.2. Graph Convolutional Networks (GCN) Basics

Graph Convolutional Networks (GCNs) extrapolate the conventional convolution functions on graphs. The network learns dependencies together, which is efficient in terms of learning because every node gathers the information of its neighbors. In skeleton-based HAR, GCNs are used to process the joint features and extract the discriminative spatial and temporal patterns which are crucial in action recognition [11].

#### 2.4.3. Message Passing and Adjacency Matrices.

The passing message mechanism enables the nodes to revise their features as a result of the flow of information with the attached nodes via the edges. The adjacency matrix represents the connectivity of the joints and can be fixed (according to the skeleton structure) or learnable (according to optimization of recognition performance). Transmission of messages across the graph allows the joint interactions to be modeled locally and globally to enhance the identification of complex actions [12]

#### 2.4.4. Skeleton Datasets

Labeled skeleton datasets are very important in the development and testing of HAR methods. Popular skeleton databases present the annotations of keypoints/joints in a range of actions, subjects, and camera views.

Table 1: Overview of the most frequently used skeleton datasets to recognize human actions by describing the number of joints, the number of subjects, actions, and modalities.

**Table 1: Comparison of Popular Human Action Recognition Datasets**

Dataset	Joints	Subjects	Actions	Modality
NTU RGB+D	25	40	60	3D skeleton
Kinetics Skeleton	18	Multiple	400+	2D/3D skeleton, RGB
Others	17–25	10–50	20–120	2D/3D skeleton, depth

As shown in Table 1, skeleton datasets vary in scale, joint representation, and modalities. These datasets provide a benchmark for evaluating skeleton-based HAR models.

### 3. Skeleton-Based HAR Using GNNs

The concept of Skeleton-based Human Action Recognition (HAR) has been advanced in multiple, important ways by the innovation of Graph Neural Networks (GNNs). Initial methods used Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks to learn temporal joint sequence dynamics. Although these sequence-based models were moderately successful, they had a hard time making the most of the spatial dependence between joints in a single frame. GNNs overcome this shortcoming by modeling the skeleton as a graph with joints appearing as nodes and bones appearing as edges which enables space and time relationships to be effectively modeled. With the help of a graph structure, GNN-based methods are capable of capturing complex joint interactions, which increase recognition performance and strength [13].

#### 3.1. Graph Construction for Skeletons

Drawing graphs is also an important procedure in skeleton-based HAR as it determines the structure on which features are diffused in GNNs. The skeleton graph is traditionally composed of spatial and temporal elements to obtain the posture and movement.

##### 3.1.1. Spatial Graph

The spatial graph is a representation of the human body structure by shown joints as nodes and bones as the edges [14]. This enables the network to acquire spatial information like joint adjacency and angles among the bones that are connected. Well-constructed space graphs assist the model to identify the postures as well as action specific joint arrangements in a single frame.

##### 3.1.2. Temporal Graph

The addition of temporal edges is done between similar joints between two consecutive frames to create a temporal graph. This allows the network to capture the dynamics of motion and how they change with time, which is crucial in the ability of the network to differentiate between actions that have the same posture but different motions.

##### 3.1.3. Spatial-Temporal Graph Convolution Networks (ST-GCN)

ST-GCN combines the spatial and temporal graph into one structure. Graph convolutions are performed along spatial boundaries in order to reflect joint interaction in frames and along time boundaries in order to reflect motion dynamics. ST-GCN is a skeleton-based HAR framework and a baseline based on GNNs, and has inspired many successors [15].

#### 3.2. Early GNN Architectures for HAR

Some of the initial GNN models have been created to enhance the skeleton-based HAR performance using the framework of ST-GCN.

##### 3.2.1. ST-GCN

The Spatial-Temporal Graph Convolutional Network (ST-GCN) uses graph convolutions on both spatial and temporal edges, therefore, minimizing the joint dependencies and motion patterns. It was able to utilize both spatial structures and temporal sequences, which led to its improved performance over the RNN/LSTM-based approaches [16].

##### 3.2.2. Two-Stream GCNs

Two-stream GCN based on two streams allow the processing of spatial and temporal information independently and then combine them to improve action recognition. This division also enables the model to concentrate its attention on structural and motion cues separately, which enhance the discriminative strength.

### 3.2.3. Attention-Enhanced GCNs

To dynamically give preference to joints that are most informative, attention mechanisms were proposed in GCNs to enable the network to concentrate on the most informative joints to perform certain actions. The method increases performance especially with complex movements that involve the use of some body parts only [17].

**Table 2: Early Skeleton Gnn Comparisons Based On the Accuracy and Model Complexity.**

Architecture	Accuracy	Model Complexity	Key Features
ST-GCN	High	Moderate	Spatial-temporal graph conv
2s-AGCN	Higher	High	Two-stream + attention
AS-GCN	High	Moderate	Adaptive adjacency learning

Table 2 highlights the performance and complexity of early skeleton-based GNN architectures, showing how structural and attention enhancements contribute to recognition accuracy.

### 3.3. Improvements and Variants

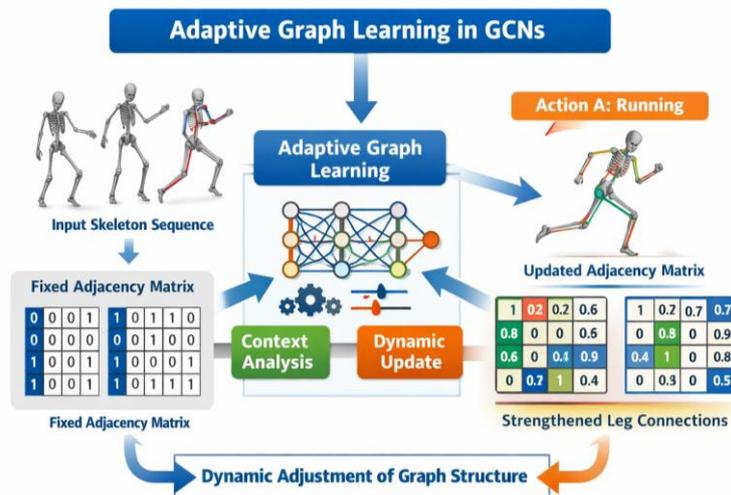
In order to improve skeleton-based HAR, a number of GCNs improvements and variants were offered.

#### 3.3.1. Multi-scale and hierarchical GCNs

Multi-scale and hierarchical GCNs are able to learn dependencies of different levels of joints, including the local bone connections to more global body parts. This enables the network to model both finer grained and global spatial information [18].

#### 3.3.2. Adaptive Graph Learning

Adaptive graph learning Learners use a restricted format of adjacency matrices which can adapt dynamically to the action context to adjust connections between joints. Such a degree of flexibility allows a better representation of joint interactions.



**Fig 3: Diagram Showing Adaptive Graph Learning In Gcns.**

Figure 3 illustrates how adaptive graph learning allows the network to modify adjacency relationships dynamically, improving recognition performance for diverse actions.

#### 3.3.3. Attention Mechanisms for Joint Importance

Attention layers use different weights to the joints, so the most important parts of a body are relevant to a particular action. This enhances concentration of the network and minimizes the noise or irrelevant joints.

**Temporal Modeling Temporal Modeling Improvements** The temporal model is enhanced through the implementation of the new temporal model. The improvements of temporal modeling including dilated convolutions and temporal attention allow the network to detect long-range dependencies between frames effectively to recognize actions with complex natural dynamics.

## 4. Challenges in Skeleton-Based HAR

Although skeleton-based Human Action Recognition (HAR) with the help of Graph Neural Networks (GNNs) has become quite widely successful, several obstacles continue to interrupt the recognition process, its generalization, and real-time

implementation. All these difficulties can be attributed to the specifics of skeleton data and constraints of GNN models. The need to comprehend these issues is paramount in coming up with stronger and effective HAR systems [19].

#### 4.1. Data Limitations

Even though skeleton data is compact and robust, it usually has quality problems and limitations of the dataset that affect the model.

##### 4.1.1. Noise and Missing Joints Skeleton.

Skeleton data provided by pose estimation algorithm or depth sensors can have noise or missing joints because of occlusion, sensor constraints or high motion. These errors may give rise to poor recognition capabilities especially in fine or subtle acts. The models have to be resistant to this noise or preprocessing and data augmentation methods have to be taken to reduce its effects [20].

##### 4.1.2. Miniature Dataset Size and Class Imbalance.

Multiple skeleton datasets are small or with an unbalanced action, and this may diminish the capacity of the model to generalize to previously unknown actions or rare behaviors. Small data sets also lead to the probability of overfitting, particularly to deep GCN models with large model capacity. The usual methods to overcome these problems are data augmentation and transfer learning.

##### 4.1.3. Cross View Recognition Problems.

It is difficult to identify actions based on the other viewpoints of the camera as the positions of the joints can be greatly different in different viewpoints. Cross-view generalization demands models to be trained on view-invariant features or be trained on multi-view generalization to enhance robustness [21].

#### 4.2. Model Limitations

The HAR models based on GNNs have other issues regarding the design of the architecture, computational demands, and learning dynamics.

Deep-GCN layers contain over-smoothing elements, which is addressed by removing them. Over-smoothing occurs as GCN features reduce as they get deeper, and the nodes become less discriminative. This may diminish the capacity of the network to differentiate between subtle joint interactions as well as complex actions especially in the representation of global dependencies.

##### 4.2.1. High Computational Cost of Large Graphs

Multiple joint and long temporal skeleton modeling may result in high computational and memory demands. On-the-fly inference of edge devices or large-scale applications can be a problem particularly in multi-stream inference or attention-enhanced GCNs. Architectures and pruning methods that are efficient in regard to performance and computational cost are frequently needed.

##### 4.2.2. Action Temporal Misalignment.

The ability to recognize can be impaired by the difference in action speed and timeliness across sequences. Quick and slow actions can be represented differently in the temporal graph and models have to represent temporal invariance or adaptively align the sequences.

##### 4.2.3. Requirements of Real-Time Inference.

Surveillance and robotics are just some of the HAR applications which demand real-time recognition. Real-time inference that is not compromised by accuracy is still a challenge, because of the complexity of the model, overhead of graph processing and time model requirements.

##### 4.2.4. Generalization and Bias of Data Sets across Environments.

Trained models can not perform generalization to new environments, patients or sensor setups. Bias in data sets such as changes in pose distribution, mode of action or sensor error requires effective learning methods and domain adaptation algorithms.

**Table 3: Challenges in Skeleton-Based Har and Corresponding Mitigation Strategies.**

Challenge	Mitigation Strategy
Skeleton noise / missing joints	Data augmentation, noise-robust models
Small dataset / class imbalance	Transfer learning, augmentation
Cross-view variations	Multi-view training, view-invariant features
Over-smoothing in GCNs	Residual connections, hierarchical GCNs

High computational cost	Lightweight models, pruning
Temporal misalignment	Temporal attention, sequence normalization
Dataset bias / generalization	Domain adaptation, cross-dataset training
Real-time inference	Efficient architectures, streamlined pipelines

## 5. Applications

Human Action Recognition (HAR) through Skeletal representations with Graph Neural Networks (GNNs) has shown great potential in a diverse spectrum of application. These models offer solutions that are strong, efficient and understandable to be used in different fields by capturing structural and temporal dynamics of human motion.

### 5.1. Surveillance and security Systems.

Skeleton-based HAR applications are used in surveillance to automatically identify anomalous or suspicious behaviors in the workplaces and other areas where people gather in large numbers, as well as infrastructures of significant societal importance. Skeleton-based solutions can also work in different conditions of light and partial cover as well as high density because they emphasize joint movement, instead of appearance. This minimizes the use of manual monitoring and improves the real-time threat detection processes [22].

### 5.2. Healthcare: Elderly Observation and Rehabilitation.

Skeleton-based HAR is important in healthcare related to the detection of falls among elderly persons, rehabilitation exercises, and mobility of patients. The methods based on GNNs can efficiently identify and evaluate the postures and motions, helping to provide personal care, identify mobility problems early, and offer feedback to control exercises. The systems are non-invasive and they do not compromise the privacy of the patient.

### 5.3. Human-Computer Interaction and Gaming

HAR enables the free contact of man and technology. Skeleton-based recognition is used in gaming and interactive applications to enable users to control an avatar, make gestures or interact with a virtual environment by using body movements [23]. Through the use of GNNs, these systems will be capable of strong capture of complex actions such as when there is occlusion or high-speed movement which enhances user experience and interaction.

### 5.4. Robotics: Action Understanding and Imitation Learning.

Skeleton based HAR is useful in robotics to observe, interpret, and repeat human behaviors. This is the capability needed by collaborative robots in an industrial, service, or domestic setting[24]. Through skeleton graph analysis, robots will be able to master complex tasks, adjust to human motions and engage with human beings safely, which will improve human-robot cooperation.

## 6. Future Direction

Although HAR with skeletons through Graph Neural Networks (GNNs) has demonstrated impressive progress, there are a number of promising areas of research that can be taken to improve performance, efficiency, and generalization. The directions are aimed at capitalizing on other modalities, enhancing computational performance, and data limitations.

The performance of HAR can be improved by combining skeleton data with other modalities (RGB frame, optical flow, and depth). Although skeletons are good at offering strong structural and temporal signals, the ability to combine the appearance and motion signals of RGB or optical flow should be used to test actions that differ slightly in movement or context. Multi-modal approaches allow the models to combine complementary information, which increases the accuracy particularly in tricky situations when the skeleton data information is the only one that is unclear.

Lots of applications e.g. surveillance, robotics and human-computer interaction need action recognition on edge devices with limited processing capabilities in real time. One of the most important directions is the development of lightweight GNN architecture that would preserve high accuracy and minimise model size and inference time. The model pruning, knowledge distillation techniques, and efficient graph convolution operations are techniques that can be used to improve real time performance without compromising on recognition quality.

Deep models usually cannot generalize due to the fact that large-scale labeled skeleton datasets are often constrained. Self-supervised learning is a potentially practical direction because it allows models to learn meaningful skeleton sequence representations without any labels. Effective supervision can be given by pretext tasks, including predicting the future joint positions, recreating corrupted skeletons or solving the temporal orderings. The methods minimize the reliance on annotated datasets and enhance robustness and transferability.

HAR models based on skeletons are usually not capable of generalization across datasets or environments because of differences in sensor type, point of view and action style. The models that are constructed using cross-dataset generalization

and domain adaptation strategies are needed to develop robust models that can be reliably used in different settings. Adversarial domain adaptation, feature normalization, and cross-view training are among the techniques that can be used to reduce the biases of the dataset and enhance the ability to work with unexamined data.

All these future directions are intended to improve the applicability, the efficiency and generalization of the skeleton-based HAR systems. The incorporation of multi-modal data, the fashioning of sparse structures, the utilization of unlabeled data and overcoming the cross-dataset challenges will serve as a further step toward the enhancement of the GNN-based methods in practice.

## **7. Conclusion**

Human Action Recognition based on skeletons (HAR) with Graph Neural Networks (GNNs) has emerged to be one of the leading trends in the field because it can effectively represent the structural and the temporal features of human movement. Incorporating the human body into a graph of joints and bones, GNNs offer a versatile and explainable form of representation, which breaks dozens of constraints of traditional RGB-based approaches, such as sensitivity to occlusion, change of light, and cluttered backgrounds. The first GCN architectures like ST-GCN were the basis to model spatial-temporal dependencies but over time, it developed to become more accurate and stronger in complex actions by adding two stream architectures, attention, adaptive graph learning, and hierarchical model improvements.

Although these have been made, there are still challenges. Noise, missing joints, and small dataset sizes are likely to affect skeleton data, whereas over-smoothing and high costs of computations are likely to affect deep GCNs. Cross-view generalization and domain adaptation is still a necessity in the real world implementation, especially in cases where there are differences between training and testing environment. To solve these limitations, data augmentation, multi-view training, lightweight architectures, and effective temporal modeling are the mitigation strategies.

Skeleton-based HAR has many applications which include but are not limited to surveillance and security systems, healthcare monitoring and rehabilitations, human-computer interaction, gaming, and robotics. The fusion of the skeleton representations and GNNs in these fields offers real-time, robust and understandable recognition.

In the future, there are a number of avenues of research with potential. A combination of skeleton data with other complementary modalities RGB and optical flow may enhance the process of recognizing fine or contextual actions. The lightweight GNN models are needed to implement on edge devices and real-time applications. Unlabeled skeleton data can be used to advantage self-supervised learning, which can subsequently be used to reduce reliance on labeled datasets. Lastly, domain adaptation and cross-dataset generalization will improve the model strength and will be used in different settings. Altogether, GNN-based skeleton HAR is a potent vision-based action recognition paradigm because it is efficient, interpretable and accurate. Further studies in the described directions will bring the state-of-the-art to the next level and allow more useful and intelligent human centred systems in the real world.

## **References**

- [1] P. Pareek and A. Thakkar, "A survey on video-based human action recognition: Recent updates, datasets, challenges, and applications," *Artif. Intell. Rev.*, vol. 54, no. 3, pp. 2259–2322, 2021.
- [2] A. Lentzas and D. Vrakas, "Non-intrusive human activity recognition and abnormal behavior detection on elderly people: A review," *Artif. Intell. Rev.*, vol. 53, no. 3, pp. 1975–2021, 2020.
- [3] U. E. Ogenyi et al., "Physical human–robot collaboration: Robotic systems, learning methods, collaborative strategies, sensors, and actuators," *IEEE Trans. Cybern.*, vol. 51, no. 4, pp. 1888–1901, 2019.
- [4] M. B. Shaikh and D. Chai, "RGB-D data-based action recognition: A review," *Sensors*, vol. 21, no. 12, Art. no. 4246, 2021.
- [5] R. Abdulghafor, S. Turaev, and M. A. H. Ali, "Body language analysis in healthcare: An overview," *Healthcare*, vol. 10, no. 7, 2022.
- [6] Z. Wang et al., "Skeleton-based human pose recognition using channel state information: A survey," *Sensors*, vol. 22, no. 22, Art. no. 8738, 2022.
- [7] S. Ghidoni and M. Munaro, "A multi-viewpoint feature-based re-identification system driven by skeleton keypoints," *Robot. Auton. Syst.*, vol. 90, pp. 45–54, 2017.
- [8] N. Baka et al., "2D–3D shape reconstruction of the distal femur from stereo X-ray imaging using statistical shape models," *Med. Image Anal.*, vol. 15, no. 6, pp. 840–850, 2011.
- [9] M. Feng and J. Meunier, "Skeleton graph-neural-network-based human action recognition: A survey," *Sensors*, vol. 22, no. 6, Art. no. 2091, 2022.
- [10] L. Shi et al., "Skeleton-based action recognition with directed graph neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, 2019, pp. 7912–7921.
- [11] F. Wu et al., "Simplifying graph convolutional networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Long Beach, CA, USA, 2019, pp. 6861–6871.

- [12] Y. Hu et al., "Graph-MLP: Node classification without message passing in graph," arXiv preprint, arXiv:2106.04051, 2021.
- [13] P. V. V. Kishore et al., "Spatial joint features for 3D human skeletal action recognition system using spatial graph kernels," *Int. J. Eng. Technol.*, vol. 7, pp. 489–493, 2018.
- [14] D. Jani et al., "Repositioning the knee joint in human body FE models using a graphics-based technique," *Traffic Injury Prevention*, vol. 13, no. 6, pp. 640–649, 2012.
- [15] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018, pp. 7444–7452.
- [16] Q. Wang, K. Zhang, and M. A. Asghar, "Skeleton-based ST-GCN for human action recognition with extended skeleton graph and partitioning strategy," *IEEE Access*, vol. 10, pp. 41403–41410, 2022.
- [17] X. Ding, K. Yang, and W. Chen, "An attention-enhanced recurrent graph convolutional network for skeleton-based action recognition," in *Proc. 2nd Int. Conf. Signal Process. Mach. Learn.*, 2019, pp. 1–6.
- [18] L. Dang et al., "MSR-GCN: Multi-scale residual graph convolution networks for human motion prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, QC, Canada, 2021, pp. 11467–11476.
- [19] S. Sarker et al., "Skeleton-based activity recognition: Preprocessing and approaches," in *Contactless Human Activity Analysis*, Cham, Switzerland: Springer, 2021, pp. 43–81.
- [20] S. Song et al., "Learning to recognize human actions from noisy skeleton data via noise adaptation," *IEEE Trans. Multimedia*, vol. 24, pp. 1152–1163, 2021.
- [21] M. Kan, S. Shan, and X. Chen, "Multi-view deep network for cross-view classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 4847–4856.
- [22] F. Angelini, *Novel Methods for Posture-Based Human Action Recognition and Activity Anomaly Detection*, Ph.D. dissertation, Newcastle Univ., Newcastle upon Tyne, U.K., 2020.
- [23] M. S. Kibbanahalli Shivalingappa and M. Swamy, "Real-time human action and gesture recognition using skeleton joints information towards medical applications," 2020.
- [24] Y. Ji et al., "A survey of human action analysis in HRI applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 7, pp. 2114–2128, 2019.