3D Human Activity Recognition and Prediction using Deep Neural Networks – Literature Review

Kishalan Pather
University of Cape Town
Cape Town, South Africa
pthkis001@myuct.ac.za

ABSTRACT

Human Activity Recognition (HAR) and Human Activity Prediction (HAP) have become prominent research domains within machine learning, with applications in healthcare, surveillance, and smart homes. HAR focuses on classifying human activities from sensor-based or vision-based data, while HAP extends this by forecasting future activities based on observed behaviour.

This paper provides a comprehensive review of recent advancements in the use of deep neural networks for HAR and HAP. The discussion centres on the state-of-the-art deep neural network algorithms including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Spatio-Temporal Graph Neural Networks (ST-GNNs). Additionally, this review explores the available 3D simulation platforms that may be used to train and test HAR and HAP algorithms.

1 INTRODUCTION

Earlier approaches to Human Activity Recognition relied heavily on handcrafted feature extraction techniques combined with traditional machine learning methods, such as k-Nearest Neighbors (k-NN) and Hidden Markov Models (HMMs). While these models provided some success, they struggled to capture complex and dynamic activities, limiting their effectiveness in real-world applications. [19]

The advancement of Deep Neural Networks (DNNs) has revolutionized HAR by automating feature extraction, significantly improving classification accuracy and efficiency [19]. Deep learning models such as Convolutional Neural Networks and Long Short-Term Memory networks have outperformed traditional machine learning methods, enabling more robust and scalable activity recognition and prediction.

Despite their success, DNN-based approaches perform well primarily with Euclidean data, such as images, text, or audio, but face challenges when handling non-Euclidean data structures like graphs [51]. Representing HAR and HAP tasks as graphs provides the advantage of capturing spatial and temporal dependencies more effectively, leading to more accurate models. Spatio-Temporal Graph Neural Networks (ST-GNNs) have emerged as a promising solution, demonstrating superior performance in modeling complex human activities compared to conventional DNN-based approaches.

This literature review explores the current state of HAR and HAP, focusing on recent advancements in deep neural network architectures. Section 2 discusses state-of-the-art deep learning algorithms for HAR and HAP, while Section 3 examines the application of Spatio-Temporal Graph Neural Networks in activity recognition and prediction. Finally, Section 4 provides a comparative analysis of 3D simulation platforms used for generating, training, and evaluating datasets in HAR and HAP research.

2 CURRENT DEEP NEURAL NETWORK ALGORITHMS FOR HAR AND HAP

2.1 Mathematical Formulation

The human activity recognition task is a classification problem whereby a sequence of input observations X are mapped to a discrete activity label Y.

Given a publicly available dataset:

$$D = \{X_1^1, X_2^2, X_3^3 \ \dots, X_i^t\}$$

whereby X_i^t represents the observed values of a sensor i at the time t.

The goal is to train the model f such that:

$$Y = f(X; \theta)$$

whereby X is the input sequence of observations, Y is the corresponding activity label, and θ is the trainable parameters of the neural network.

The cross-entropy error function is used as the objective function:

$$L = -1/N \sum \sum ylog(f(x; \theta))$$

whereby $y = 1$ if it is the correct label.

The parameters θ learns by optimising the objective function using gradient descent.

$$\theta * = arg\theta minL(X, Y; \theta)$$

where L is the loss function.

2.2 Publicly Available Datasets

See attached table 1 in appendix.

2.3 Machine Learning Pipeline

In terms of the machine learning pipeline the following steps will be taken:

- Collect the data, which comes from publicly available datasets.
- Clean and format the data if needed.
- Define the inputs and outputs of the model as previously outlined.
- Choose the model. I.e. CNN's, LSTM's or STG-NN's.
- Use the cross-entropy function as the loss function.
- Minimise the loss function using gradient descent.
- Test the model using 80% of the data for training and the remaining 20% for testing.
- Evaluate the model using the relevant metrics such as accuracy, F1-score, precision, etc.

2.4 Current Algorithms for HAR

2.4.1 Convolutional Neural Networks

Convolutional Neural Networks build upon conventional DNNs by adding extra computations to handle multi-dimensional input. They are one of the most researched deep learning techniques and are especially useful for image classification [13]. They are also adept at recognizing correlations between nearby observations, which improves activity recognition accuracy, and they can handle input sequences of varying lengths. Singh et al. [42] demonstrated that a 1D CNN model achieved performance comparable to LSTM-based models while being computationally more efficient.

2.4.2 Long Short-Term Memory Networks

Long Short-Term memory networks [27] networks are a type of recurrent neural network(RNN). The LSTM layers main component is a unit called the memory block. An LSTM block has three gates which are input, output and forget gates. These gates can be seen as write, read and reset operations for the cells. They have been found to be effective at modelling time series data from sensors. [36].This characteristic makes them especially effective for Human Activity Prediction (HAP) tasks. Du et al. [17] applied an LSTM in their activity prediction model and demonstrated that LSTM-based approaches consistently outperform traditional machine learning models in terms of accuracy.

2.4.3 Hybrid

A hybrid model is when two or more ML algorithms are used together to build a model, in this case specifically two or more deep learning models. Many attempts have been made to combine algorithms, and some outperform standalone models in the literature [20]. This is done to leverage the strengths of a model and compensate for its weaknesses.

See table 2 for current studies on deep learning models in appendix.

2.5 Metrics Used For Evaluation

The most common metrics used for evaluation are accuracy, precision, F-1 score and recall.

It is helpful to note that each activity can be classified as True Positive (TP)/True Negative (TN) when correctly recognised or False Positive (FP)/False Negative (FN) when incorrectly classified.

Thus, the following metrics can be defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

The proportion of correctly classified instances out of all instances.

$$Precision = \frac{TP}{TP + FP}$$

The proportion of correctly predicted positive instances out of all predicted positives. This measures how many of the positive predictions were correct.

$$Recall = \frac{TN}{TN + FP}$$

The proportion of correctly predicted positive instances out of all actual positives. This measures how well the model captures actual positives.

$$F-1 = 2\times\frac{Precision \times Recall}{Precision + Recall}$$

Combination of precision and recall. It useful when there is an imbalance between false positives and false negatives.

2.6 Current Work on HAP

Most HAR systems rely on recorded sensor data to recognize human activities. However, research on Human Activity Prediction (HAP), which leverages future sensor data to anticipate human actions, remains relatively scarce [29]. Despite this, several studies have successfully applied deep learning techniques for activity prediction, notably the works of Jaramillo et al. [29] and Du et al. [17].

Du et al. [17] developed an activity prediction model for household environments. It utilised an "activity chain" approach where both current and previous activities were utilized to forecast future actions. Their model, based on LSTMs, outperformed traditional machine learning models such as Naïve Bayes. The system could incorporate up to three past activities to enhance prediction accuracy, achieving a peak accuracy of 78.3%. However, the framework struggled with activities that had strong temporal dependencies—such as those influenced by time and duration—because the recognition method primarily relied on spatial information. This limitation suggests that an approach incorporating Spatio-Temporal Graph Neural Networks (STG-NNs) could potentially enhance predictive accuracy.

Jaramillo et al. [29] developed a HAP system that utilised activity signals from accelerometers placed on the chest, arm, and ankle. Their model predicted one of five movements: walking, running, Nordic walking, stair ascent, and stair descent. They experimented with multiple hybrid models, including a CNN combined with an LSTM (Conv2LSTM). Their best-performing model achieved an average accuracy of 97.96%. However, the authors noted that additional sensors on the wrist or head could improve recognition of more complex activities. This suggests that incorporating a 3D simulation environment or a digital twin could further enhance model accuracy.

2.7 Mathematical formulation for HAP

The human activity prediction task involves forecasting future activities based on current and past activities. Given a sequence of past activity data X_i^t ,The goal is to predict the next Y future activity.

$$Y = f(X_i^t;\theta)$$

Whereby $X_i^t = \{x_1, x_2, \ldots, x_i\}$ is the observed sequence from time i up to time t.

The same cross entropy error function as used before in HAR, will be used as the objective function

$$L = -1/N \sum\sum ylog(f(X_i^t;\theta))$$

whereby $y = 1$ if it is the correct label.

The parameters θ learns by optimising the objective function using gradient descent.

$$\theta *= arg\theta minL(X_i^t, Y;\theta)$$

where L is the loss function.

3 SPATIAL TEMPORAL GRAPH NEURAL NETWORKS APPLIED TO HAR AND HAP

Spatial Temporal Graph Neural Networks (STG-NN) are an emerging set of DNNs. STGNNs extend Graph Neural Networks (GNNs) by incorporating temporal modeling techniques, such as recurrent layers, attention mechanisms, or transformer-based architectures, to effectively capture dynamic changes in graph-structured data.

Traditional DNN approaches can learn temporal dependencies and treat time series as sequential, but they are not designed to deal with spatial dependencies. STGNNs explicitly capture both spatial and temporal dependencies making them more suitable for multi sensor applications [51].

They have been most notably used in predicting traffic flow at different points in a city traffic network [48] but their utility extends beyond this and may be used in the HAR/HAP space.

This is because STGNN's are effective at capturing body part relationships in wearable sensor or skeletal data [48], and since they consider spatial and temporal aspects, they have better recognition and accuracy [1].

See table 3 in appendix for STG-NN studies.

4 COMPARISION OF 3D SIMULATION PLATFORMS

Efficiently training a deep learning model requires a dataset with minimal noise and variability [7]. However, finding such datasets is challenging, often leading to a lack of data that accurately represents the intended application environment. 3D simulation platforms address this issue by digitally recreating environments, such as a home setting, where data can be captured in a controlled manner. These simulations model real-world systems, allowing users to test various scenarios and generate high-quality training data.

Although similar, a digital twin is a real-time virtual representation of a physical entity, continuously updated with live data. In HAR, a digital twin could represent a smart home with virtual sensors or an avatar mimicking human activities in real time. These avatars simulate physical movements, which are then captured by virtual sensors to generate training data for deep learning algorithms. This approach accelerates data collection while reducing the costs associated with physical sensors.

See table 4 for 3D simulation platforms in appendix.

5 CONCLUSIONS

This literature review examined advancements in deep neural networks for human activity recognition and prediction, emphasizing the strengths and limitations of existing approaches. While CNNs and LSTMs have demonstrated strong performance, Spatio-Temporal Graph Neural Networks (STG-NNs) offer a promising alternative due to their ability to capture complex spatial and temporal dependencies.

Furthermore, the integration of 3D simulation platforms presents a valuable solution to the challenges posed by noisy or variable datasets, enhancing the reliability of training models. Future research should prioritize the implementation of STG-NNs while leveraging synthetic data from digital twin simulations to refine recognition and prediction accuracy.

REFERENCES

[1] Hosam Abduljalil, Ahmed Elhayek, Abdullah Marish Ali, and Fawaz Alsolami. 2024. Spatiotemporal Graph Autoencoder Network for Skeleton-Based Human Action Recognition. AI 5, 3: 1695–1708. https://doi.org/10.3390/ai5030083

[2] Emre Aksan, Manuel Kaufmann, Peng Cao, Otmar Hilliges. 2020. A Spatio-temporal Transformer for 3D Human Motion Prediction. https://doi.org/10.48550/arXiv.2004.08692

[3] Nasser Alshammari, Talal Alshammari, Mohamed Sedky, Justin Champion, and Carolin Bauer. 2017. OpenSHS: Open smart home simulator. Sensors (Switzerland) 17, 5. https://doi.org/10.3390/s17051003

[4] Ibrahim Armac, and Daniel Retkowitz. 2007. Simulation of Smart Environments. 322-331. 10.1109/PERSER.2007.4283934.

[5] Oresti Banos, Rafael Garcia, Alejandro Saez. 2014MHEALTH [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5TW22

[6] Oresti Banos. Mate Toth. Oliver Amft. 2014. REALDISP Activity Recognition Dataset [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5GP6D.

[7] Damien Bouchabou, Juliette Grosset, Sao Mai Nguyen, Christophe Lohr, and Xavier Puig. 2023. A Smart Home Digital Twin to Support the Recognition of Activities of Daily Living. Sensors 23, 17. https://doi.org/10.3390/s23177586

[8] Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit Leduc, and Ioannis Kanellos. 2021. A survey of human activity recognition in smart homes based on iot sensors algorithms: Taxonomies, challenges, and opportunities with deep learning. Sensors 21. https://doi.org/10.3390/s21186037

[9] Kévin Bouchard, Amir Ajroud, Bruno Bouchard, K Bouchard, A Ajroud, B Bouchard, A Bouzouane, and {kevin Bouchard. 2012. SIMACT: a 3D Open Source Smart Home Simulator for Activity Recognition with Open Database and Visual Editor. Retrieved from https://www.researchgate.net/publication/230646951

[10] Mario Buchmayr, Werner Kurschl, and Josef Küng. 2011. A simulator for generating and visualizing sensor data for ambient intelligence environments. In Procedia Computer Science, 90–97. https://doi.org/10.1016/j.procs.2011.07.014

[11] Mohammud J. Bocus, Wenda Li, Shelly Vishwakarma, Roget Kou, Chong Tang, Karl Woodbridge, Ian Craddock, Ryan McConville, Raul Santos-Rodriguez, Kevin Chetty, Robert Piechocki. 2021. OPERAnet: A Multimodal Activity Recognition Dataset Acquired from Radio Frequency and Vision-based Sensors. https://arxiv.org/abs/2110.04239?

[12] Joao Carreira, Eric Noland, Andras Banki-Horvath, Chloe Hillier, Andrew Zisserman. 2018. A Short Note about Kinetics-600. arXiv 2018, arXiv:1808.01340.

[13] Kaixuan Chen, Lina Yao, Dalin Zhang, Bin Guo, and Zhiwen Yu. 2019. Multi-agent Attentional Activity Recognition.

[14] Chen Chen, Roozbeh Jafari, and Nasser Kehtarnava. 2015.UTD-MHAD: A Multimodal Dataset for Human Action Recognition Utilizing a Depth Camera and a Wearable Inertial Sensor.

[15] Roggen Daniel, Förster Kilian, Calatroni Alberto. 2009. OPPORTUNITY: Towards opportunistic activity and context recognition systems.https://infoscience.epfl.ch/entities/publication/8e3caf0d-4cc8-4046-af4d-edb3d5407dd4

[16] Dima Damen, Hazel Doughty, Giovanni Maria Farinella, Antonino Furnari, Evangelos Kazakos, Jian Ma, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, Michael Wray. 2022. Rescaling Egocentric Vision: Collection, Pipeline and Challenges for Epic-Kitchens-100. Int. J. Comput. Vis. 2022, 130, 33–55.

[17] Yegang Du, Yuto Lim, and Yasuo Tan. 2019. A novel human activity recognition and prediction in smart home based on interaction. Sensors (Switzerland) 19, 20. https://doi.org/10.3390/s19204474

[18] Haodong Duan, Jiaqi Wang, Kai Chen, and Dahua Lin. 2022. DG-STGCN: Dynamic Spatial-Temporal Modeling for Skeleton-based Action Recognition. Retrieved from http://arxiv.org/abs/2210.05895

[19] Mohammed Elnazer, Abazar Elmamoon, and Ahmad Abubakar Mustapha. A Comparative Study of Deep Learning Models for Human Activity Recognition. Cloud Computing and Data Science 6. https://doi.org/10.37256/ccds.6120256264

[20] Xiaoyi Fan, Fangxin Wang, Feng Wang, Wei Gong, and Jiangchuan Liu. 2019. When RFID Meets Deep Learning: Exploring Cognitive Intelligence for Activity Identification. IEEE Wireless Communications 26, 3: 19–25. https://doi.org/10.1109/MWC.2019.1800405
[21] Yannick Francillette, Eric Boucher, Abdenour Bouzouane, and Sébastien Gaboury. 2017. The virtual environment for rapid prototyping of the intelligent environment. Sensors (Switzerland) 17, 11. https://doi.org/10.3390/s17112562
[22] Kristen Grauman, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson Hamburger, Hao Jiang, Miao Liu, Xingyu Liu, Miguel Martin, Tushar Nagarajan, Ilija Radosavovic, Santhosh Kumar Ramakrishnan, Fiona Ryan, Jayant Sharma, Michael Wray, Mengmeng Xu, Eric Zhongcong Xu, Chen Zhao. 2022. Around the World in 3000 Hours of Egocentric Video. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition
[23] Yu Guan and Thomas Ploetz. 2017. Ensembles of Deep LSTM Learners for Activity Recognition using Wearables. https://doi.org/10.1145/3090076
[24] A. Helal, K. Cho, W. Lee, Y. Sung, J. W. Lee, and E. Kim. 2012. 3D modeling and simulation of human activities in smart spaces. In Proceedings - IEEE 9th International Conference on Ubiquitous Intelligence and Computing and IEEE 9th International Conference on Autonomic and Trusted Computing, UIC-ATC 2012, 112–119. https://doi.org/10.1109/UIC-ATC.2012.35
[25] Mohammadreza Heydarian, Thomas E. Doyle. 2023. rWISDM: Repaired WISDM, a Public Dataset for Human Activity Recognition
[26] Brandon Ho, Dieter Vogts, and Janet Wesson. 2019. A smart home simulation tool to support the recognition of activities of daily living. In ACM International Conference Proceeding Series. https://doi.org/10.1145/3351108.3351132
[27] Hochreiter Sepp and Schmidhuber Jurgen. 1997. Long Short-Term Memory.
[28] Masaya Inoue, Sozo Inoue, and Takeshi Nishida. 2018. Deep recurrent neural network for mobile human activity recognition with high throughput. Artificial Life and Robotics 23, 2: 173–185. https://doi.org/10.1007/s10015-017-0422-x
[29] Ismael Espinoza Jaramillo, Channabasava Chola, Jin Gyun Jeong, Ji Heon Oh, Hwanseok Jung, Jin Hyuk Lee, Won Hee Lee, and Tae Seong Kim. 2023. Human Activity Prediction Based on Forecasted IMU Activity Signals by Sequence-to-Sequence Deep Neural Networks. Sensors 23, 14. https://doi.org/10.3390/s23146491
[30] Nobuo Kawaguchi, Nobuhiro Ogawa, Yohei Iwasaki, Katsuhiko Kaji, Tsutomu Terada, Kazuya Murao, Sozo Inoue, Yoshihiro Kawahara, Yasuyuki Sumi, and Nobuhiko Nishio. 2011. HASC Challenge: gathering large scale human activity corpus for the real-world activity understandings. In Proceedings of the 2nd Augmented Human International Conference (AH '11). Association for Computing Machinery, New York, NY, USA, Article 27, 1–5. https://doi.org/10.1145/1959826.1959853
[31] Tianjiao Li, Jun Liu, Wei Zhang, Yun Ni, Wenqian Wang, Zhiheng Li. 2021. A Large Benchmark for Human Behavior Understanding with Unmanned Aerial Vehicles. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition
[32] Meet Nagadia. 2021. Human Action Recognition (HAR) Dataset. https://www.kaggle.com/datasets/meetnagadia/human-action-recognition-har-dataset
[33] William T Ng, K Siu, Albert C Cheung, and Michael K Ng. Expressing Multivariate Time Series as Graphs with Time Series Attention Transformer. Retrieved from https://github.com/RadiantResearch/TSAT.
[34] Jorge Reyes-Ortiz, Davide Anguita. 2013 Human Activity Recognition Using Smartphones [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C54S4K.
[35] Attila Reiss. 2012. PAMAP2 Physical Activity Monitoring. https://doi.org/10.24432/C5NW2H.
[36] Pienaar Schalk Willhelm and Malekian Reza. 2019. Human Activity Recognition using LSTM-RNN Deep Neural Network Architecture.
[37] Oumaima Saidani, Majed Alsafyani, Roobaea Alroobaea, Nazik Alturki, Rashid Jahangir, and Leila Jamel. 2023. An Efficient Human Activity Recognition Using Hybrid Features and Transformer Model. IEEE Access 11: 101373–101386. https://doi.org/10.1109/ACCESS.2023.3314492
[38] L SaiRamesh, B Dhanalakshmi, and Selvakumar K. 2024. Human Activity Recognition Through Images Using a Deep Learning Approach. https://doi.org/10.21203/rs.3.rs-4443695/v1
[39] Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, Devi Parikh, and Dhruv Batra. 2019. Habitat: A Platform for Embodied AI Research. Retrieved from http://arxiv.org/abs/1904.01201
[40] Niloy Sikder, Md Sanaullah Chowdhury, Abu Shamim Mohammad Arif, and Abdullah Al Nahid. 2019. Human activity recognition using multichannel convolutional neural network. In 2019 5th International Conference on Advances in Electrical Engineering, ICAEE 2019, 560–565. https://doi.org/10.1109/ICAEE48663.2019.8975649
[41] Niloy Sikder. 2021. KU-HAR: Human Activity Recognition Dataset (v 1.0). https://www.kaggle.com/datasets/niloy333/kuhar?utm_source=chatgpt.com
[42] Deepika Singh, Erinc Merdivan, Ismini Psychoula, Johannes Kropf, Sten Hanke, Matthieu Geist, and Andreas Holzinger. 2018. Human Activity Recognition using Recurrent Neural Networks. https://doi.org/10.1007/978-3-319-66808-6_18
[43] Jonathan Synnott, Chris Nugent, and Paul Jeffers. 2015. Simulation of smart home activity datasets. Sensors (Switzerland) 15, 6: 14162–14179. https://doi.org/10.3390/s150614162

[44] Yansong Tang, Zian Wang, Jiwen Lu, Jianjiang Feng, and Jie Zhou. 2019. Multi-Stream Deep Neural Networks for RGB-D Egocentric Action Recognition. IEEE Transactions on Circuits and Systems for Video Technology 29, 10: 3001–3015. https://doi.org/10.1109/TCSVT.2018.2875441

[45] Md Ashraf Uddin, Md Alamin Talukder, Muhammad Sajib Uzzaman, Chandan Debnath, Moumita Chanda, Souvik Paul, Md Manowarul Islam, Ansam Khraisat, Ammar Alazab, and Sunil Aryal. 2024. Deep learning-based human activity recognition using CNN, ConvLSTM, and LRCN. International Journal of Cognitive Computing in Engineering 5: 259–268. https://doi.org/10.1016/j.ijcce.2024.06.004

[46] Xiaohan Wang, Yu Wu, Linchao Zhu, and Yi Yang. Symbiotic Attention with Privileged Information for Egocentric Action Recognition. Retrieved from aaai.org

[47] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. 2019. Graph WaveNet for Deep Spatial-Temporal Graph Modeling. Retrieved from http://arxiv.org/abs/1906.00121

[48] Sijie Yan, Yuanjun Xiong, and Dahua Lin. 2018. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. Retrieved from http://arxiv.org/abs/1801.07455

[49] Sijie Yan, Yuanjun Xiong, and Dahua Lin. 2018. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. Retrieved from http://arxiv.org/abs/1801.07455

[50] Piero Zappi, Clemens Lombriser, Thomas Stiefmeier, Elisabetta Farella, Daniel Roggen, Luca Benini & Gerhard Tröster. 2008. Activity recognition from on-body sensors: Accuracy-power trade-off by dynamic sensor selection. In Wireless sensor networks (pp. 17–33). Springer

[51] Assaad Zeghina, Aurélie Leborgne, Florence Le Ber, and Antoine Vacavant. 2024. Deep learning on spatiotemporal graphs: A systematic review, methodological landscape, and research opportunities. Neurocomputing 594. https://doi.org/10.1016/j.neucom.2024.127861

APPENDIX

Reference	Dataset	Type	Year	Sample Size
[14]	UTD-MHAD	RGB images, sensor signals, skeletal joint locations	2015	8
[34]	UCI-HAPT	accelerometer, gyroscope signals	2012	30
[25]	WISDM	Wireless sensor	2023	36
[35]	PAMAP-2	Accelerator, gyroscope	2012	9
[44]	THU-READ	RGB	2017	8
[30]	HASC	Smartphone sensors	2011	7
[32]	HAR	RGB	2021	15
[15]	Opportunity	accelerometer, gyroscope, magnetometer, and ambient sensors;	2010	4
[50]	Skoda	accelerometer, gyroscope	-	1
[16]	EPIC-Kitchens-100	RGB, Acceleration	2022	45
[5]	MHEALTH,	accelerometer, gyroscope, and magnetometer	2014	10
[41]	KU-HAR	Smartphone sensors	2021	90
[11]	OPERAnet	radio-Frequency devices, vision-based sensors	2021	6
[12]	Kinetics-600	RGB	2017	-
[6]	REALDISP	accelerometer, gyroscope, and magnetometer	2014	17
[22]	Ego4D	RGB, Acceleration	2022	923
[31]	UAV-Human	RGB,Skeleton,depth,infrared	2021	119

Table 1: Publicly available datasets

Study	Year	Algorithm	Type	Dataset/s	Metric	Result	Insight
[38]	2024	CNN	RGB images	UTD-MHAD	Accuracy	86.7%	Required significant computational resources and access to large datasets.
[37]	2023	Transformer model	Wireless sensor data	WISDM, PAMAP-2, UCI-HART	Accuracy (WISDM)	97.3%	The transformer model outperformed other DL models

							in these 3 datasets.
[44]	2019	CNN	Daily activity	THU-READ,WCVS	Accuracy	91.72% and 67.04% respectively.	-
[28]	2018	LSTM	Locomotion Activity	HASC	Accuracy	95.4%	Recommends trying the reLU function.
[23]	2017	LSTM (ensemble)	Daily Activity	Opportunity, PAMAP2, Skoda	F-1 score	72.6%,85.4%,92.4% respectively	One of the challenges is noisy/erroneous data, imbalanced data. Placing the data in an ensemble improved recognition accuracies.
[46]	2020	CNN	Daily Activity	EGTEA,EPIC-Kitchens	Accuracy	40.5%(top 5), 62.7% respectively	-
[20]	2019	CNN + LSTM	Locomotion Activity	Self collected dataset[180]	Accuracy	94%	-
[13]	2019	LSTM	Locomotion Activity	MHEALTH, PAMAP2, UCI HAR,	Accuracy, Precision, Recall, F-1	96.12%,90.33%,85.72% (accuracy) respectively	-
[19]	2025	CNN	Surveillance	HAR dataset	Accuracy, Precision,Recall,F-1 score, loss	80.16% (accuracy)	Most recent HAR study that deals with state of the art pretrained CNNs. Thus InceptionV3 may be the best pre-trained CNN model as of right now.

Table 2: Studies using Deep Neural Networks for HAR/HAP

Study	Year	STGNN Algorithm Applied	Potential for HAR/HAP
[47]	2019	GraphWaveNet	Very effective at learning spatial dependencies and capturing long term temporal patterns for HAR and HAP.
[46]	2020	MTGNN(Multivariate Time-series Graph Neural Network)	Especially good at time series forecasting. Effective at learning temporal patterns and can thus model the complex dependencies between body parts for HAP.
[48]	2018	ST-GCN (Spatial-Temporal Graph Convolutional Network)	Effective for skeleton-based HAR as it outperforms previous state of the art skeleton-based models.
[2]	2020	ST-Transformer	Especially good at complex tasks HAR. i.e. tasks that last up to 20 seconds long.
[18]	2022	DG-STGCN (Dynamic Graph STGCN)	Able to adapt to personalised HAR scenarios as it can recognise joint sets that are entirely different to what it was trained on.

Table 3: Studies showing the potential of STG-NNs for HAR and HAP.

Platform	Open Source?	Environment	API
AI Habitat [39]	Yes	C++	Python
Open SHS [3]	Yes	Blender	Python
SESim [26]	Yes	Unity	NA
IE Sim [43]	No	NA	NA
SIMACT [9]	Yes	JME	Java
Francillette et al. [21]	Yes	Unity	NA
Buchmayr et al. [10]	No	NA	NA
Armac et al. [4]	No	NA	NA
VirtualSmartHome[7]	No	Unity	NA

Table 4: Table of 3D simulation platforms