

Article

Not peer-reviewed version

A Structured and Methodological Review on Multi-View Human Activity Recognition for Ambient Assisted Living

[Fahmid Al Farid](#) ^{*}, [Md Ahsanul Bari](#), [Abu Saleh Musa Miah](#), [Sarina Mansur](#) ^{*}, [Jia Uddin](#), [Prabha Kumaresan](#)

Posted Date: 26 March 2025

doi: [10.20944/preprints202503.1926.v1](https://doi.org/10.20944/preprints202503.1926.v1)

Keywords: ambient assisted living; lightweight deep learning; activity recognition; machine learning; wearable sensors; smartphones; context-aware; deep learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Structured and Methodological Review on Multi-View Human Activity Recognition for Ambient Assisted Living

Fahmid Al Farid ^{1,*}, Ahsanul Bari ¹, Abu Saleh Musa Miah ², Sarina Mansor ^{1,*}, Jia Uddin ³ and Prabha Kumaresan ⁴

¹ Faculty of Engineering, Multimedia University, Cyberjaya 63100, Malaysia.

² Department of Computer Science and Engineering, Bangladesh Army University of Science and Technology (BAUST), Saidpur, Bangladesh.

³ Technology Studies Department, Endicott College, Woosong University, Daejeon, Korea 32820.

⁴ Faculty of Computing and Informatics, Multimedia University, Cyberjaya 63100, Malaysia.

* Correspondence: fahmid.farid@gmail.com (F.A.F.); sarina.mansor@mmu.edu.my (S.M.)

Abstract: Ambient Assisted Living (AAL) leverages technology to support the elderly and individuals with disabilities. A key challenge in AAL systems is efficient human activity recognition (HAR), yet no study has systematically compared single-view (SV) and multi-view (MV) HAR. This review addresses this gap by analyzing the evolution from SV to MV-HAR, covering benchmark datasets, feature extraction methods, and classification approaches. We examine how HAR systems have transitioned to MV with advanced deep learning architectures optimized for AAL, improving accuracy and robustness. Additionally, we explore machine learning and deep learning models—including CNNs, RNNs, LSTMs, TCNs, and GCNs—as well as lightweight transfer learning techniques for resource-constrained environments. Key challenges such as data remediation, privacy, and generalization are discussed alongside potential solutions like sensor fusion and advanced learning methods. Our study provides insights into advancements and future directions, guiding the development of intelligent, efficient, and privacy-compliant HAR systems for AAL.

Keywords: ambient assisted living; lightweight deep learning; activity recognition; machine learning; wearable sensors; smartphones; context-aware; deep learning

1. Introduction

Human Activity Recognition (HAR) is a fundamental research area, particularly in Ambient Assisted Living (AAL) environments where safety, independence, and well-being of elderly and disabled people are of major concern. The continuous evolution of sensor technologies, along with the power of deep learning, has revolutionized the area, making it possible to monitor and classify human activities with high accuracy in real time. Multi-view HAR (MV-HAR), which integrates data from multiple sensors, has emerged as a robust solution to overcome the limitations of single-view approaches. This review focuses on the advancements in lightweight deep learning models that optimize HAR systems, allowing them to operate efficiently in resource-constrained AAL environments. By synthesizing recent research, this paper identifies gaps and suggests future directions for enhancing HAR systems' scalability and computational efficiency.

1.1. Applications in Ambient Assisted Living (AAL)

Precise, versatile HAR solutions are of key relevance to AAL systems. Lightweight deep learning models are particularly well suited to AAL applications. MV-HAR has a lot to offer:

- **Fall Detection:** Fall detection is required to ensure the safety of the elderly living alone.



- **Activity Monitoring:** Provides a means of monitoring Activities of Daily Living (ADLs), which can detect physical and cognitive health changes when there are abnormalities detected in MV-HAR.
- **Anomaly Detection:** Detection of significant deviations from normal patterns can be reflective of health events, e.g., falls.
- **By LIFEBytes:** MV-HAR monitors exercises and guides rehabilitation progress monitoring.

1.2. *Integration in AAL Systems*

The synergy of multi-view HAR and lightweight deep learning models provides an end-to-end solution for developing ubiquitous and real-time Ambient Assisted Living (AAL) systems. We present in this paper case studies and applications that have been favorably affected using this synergistic framework as evidenced through high inference accuracy for daily activity and system performance [1]. The potential of such technologies for assisting daily activity and emergent event detection (e.g., falls and medical events) is particularly highlighted.

1.3. *Multi-View Human Activity Recognition*

To counter the limitation of single-view Human Activity Recognition systems, one of the solutions is MV-HAR. MV-HAR uses multiple synchronized high-resolution wide-angle cameras of various views to fuse information from multiple views to provide a complete and detailed description of complex activities. Even though hand-crafted features dominated early research of MV-HAR, deep learning-based approaches now prevail over traditional ones.

1.4. *Background*

HAR involves using sensors and algorithms to detect and classify human activities based on data streams, such as those generated by accelerometers, gyroscopes, and cameras. This technology is crucial in diverse fields, including healthcare, smart environments, and industrial monitoring. In the context of AAL, HAR plays a vital role in tracking daily activities, predicting health risks, and providing timely assistance to the elderly and individuals with disabilities.

Historically, HAR systems relied on single-view sensors, which posed challenges due to limited perspectives and occlusions. To address these shortcomings, MV-HAR has been introduced, allowing for more comprehensive monitoring by incorporating data from multiple sources, such as fixed and wearable cameras. The challenge lies in the computational complexity of these systems, which impedes their real-time deployment in environments like AAL, where resources are often limited. Lightweight deep learning models, designed for efficiency and accuracy, have shown promise in tackling this challenge by optimizing sensor data processing without compromising performance.

1.5. *Motivation*

The rapid increase in global life expectancy has created a pressing need for technologies that support independent living for the elderly and disabled. AAL systems, powered by HAR, can monitor individuals' daily activities and provide crucial insights into their well-being, offering a safety net in scenarios like fall detection and health monitoring. However, the high computational demands of traditional HAR systems, particularly when using deep learning models, present significant challenges in deploying these technologies in real-time, resource-constrained environments.

This review is motivated by the need to explore recent advancements in lightweight deep learning models that address these computational challenges. By examining the integration of multi-view data and deep learning techniques, we aim to identify solutions that can enhance the efficiency, scalability, and accuracy of HAR systems in AAL environments. Moreover, this paper highlights the potential for lightweight models, such as Mobile LeNet and Mobile Neural Architecture Search Network (MnasNet), to outperform more complex systems, making them ideal for real-time applications in AAL settings.

In this review, we make the following key contributions to the field of multi-view human activity recognition (MV-HAR) for Ambient Assisted Living (AAL):

1. **Comprehensive Analysis of MV-HAR Methods:** We provide a detailed survey of the state-of-the-art MV-HAR methods, exploring the transition from single-view to multi-view approaches. This includes a comparative evaluation of existing models and a discussion on how multi-view architectures enhance the accuracy and robustness of HAR systems in AAL environments.
2. **Focus on Lightweight Deep Learning Models:** We highlight the advantages of lightweight deep learning architectures tailored for resource-constrained AAL settings. Our review covers novel techniques that balance computational efficiency with accuracy, making them suitable for real-time applications in AAL.
3. **Examination of Key Challenges and Solutions:** We address several pressing challenges in the field, including issues related to data heterogeneity, privacy concerns, and the need for generalizable HAR models. In response, we explore solutions such as sensor fusion, transfer learning, and privacy-preserving techniques that improve the efficacy of HAR systems.
4. **Guideline for Future Research:** By synthesizing the current progress and identifying gaps in the literature, we offer a set of guidelines and future directions aimed at developing more intelligent, scalable, and privacy-compliant AAL systems. This serves as a foundation for researchers looking to innovate in the intersection of MV-HAR and AAL.

The rest of the paper is organized as follows: Section 3 explains the dataset in detail. Section 4 explains the proposed methodology in detail, such as data preprocessing, techniques, positional activity and skeleton extraction, segmentation, and classification approaches. Section 5 provides the experimental analysis, such as performance and accuracy. Section 6 concludes the paper.

2. Related Work

Recent research on Human Activity Recognition (HAR) has extensively explored the use of machine learning and deep learning techniques to achieve significant improvements in both accuracy and computational efficiency [2–13]. Deep learning models have revolutionized HAR, particularly through the use of lightweight architectures designed for resource-constrained environments like Ambient Assisted Living (AAL) systems [14,15]. These models, such as Mobile LeNet (M-LeNet) and Mobile Neural Architecture Search Network (MnasNet), offer superior performance in AAL scenarios due to their ability to handle multi-view data from various sensors, including fixed cameras and wearable devices [16]. Several studies have demonstrated the effectiveness of lightweight CNN architectures for improving HAR systems' computational efficiency and scalability [17]. For instance, the integration of multi-view data collection and lightweight deep learning models has been shown to significantly enhance the ability to monitor activities in AAL environments. Using the Robot Human Monitoring-Human Activity Recognition-Skeleton (RHM-HAR-SK) dataset, lightweight models such as M-LeNet and MnasNet have outperformed more complex models by offering a balance between accuracy and computational resource demands [18]. In addition, HAR research has expanded to include hybrid deep learning models that incorporate both conventional and emerging methods, addressing challenges such as sensor data synchronization and privacy concerns [17,19]. These hybrid models leverage multi-modal data sources, such as accelerometer data, video sequences, and audio signals, which improve HAR system robustness and generalization across different environments and activities [20]. The integration of skeleton information has also been explored to enhance activity recognition, particularly in detecting critical events like falls, which are vital in AAL systems [18]. Other studies have explored the potential of large language models to automate the filtering and taxonomy creation of relevant academic papers in the HAR domain, which has accelerated the pace of research [17]. These taxonomies categorize deep learning models into two main types: conventional models, which rely on traditional feature extraction techniques, and hybrid models, which combine

multiple modalities to enhance HAR system performance [15]. This approach has allowed researchers to identify gaps in the literature, particularly in terms of data diversity, privacy preservation, and computational limitations in real-time applications [21]. Further, the rise of multimodal HAR systems, which integrate information from various sensor types, including video and inertial data, has proven effective in increasing the accuracy and applicability of HAR in real-world scenarios. For example, studies using the UP-Fall detection dataset have demonstrated that integrating multi-head CNNs with LSTM networks yields superior performance over traditional HAR methods, particularly in healthcare applications [22]. Challenges such as computational intensity and reliance on multimodal data remain, but recent innovations in deep learning architectures have shown promise in addressing these issues [23].

3. Benchmark Dataset

To compare deep learning and machine learning model employment on HAR, researchers have proposed various benchmark datasets [24]. Motion dataset contains motion signals of embedded wearable sensors of wearables or smartphone multi-sensors on body positions: chest, forearm, head, pocket, or wrist. Wearables can be worn on the head, shin, forearm, chest, upper arm, thigh, waist, and legs. If you like wearing your smartphone in your pocket and your smartwatch on your right wrist (because you are right-handed), you may do so. Sensors found in the datasets There are accelerometers (A), gyroscopes (G), magnetometers (M) and others (e. g., object, temperature, ambient [24]. We have also recorded the ages and weights heights etc., of the participants varied in each of the datasets. Simple (walk, run, lie down), complex (cook, clean), postural transitions (sit to stand) Detailed descriptions and information regarding the benchmark datasets are presented in Table 1 [25]

Table 1. List of Publicly Available HAR for AAL Related Benchmark Datasets

Dataset Name	Number of Activities	Activity Type	Number of Subjects	Sampling Frequency	Devices Used	Device Position	Latest Performances
PAM-AP2 [26]	12	ADLs, Postural, Complex	9	100Hz	3 IMU units, 1 Heart rate monitor	Wrist, chest, dominant ankle	
HHAR [27]	6	ADLs	9	Highest Available	Smartwatches, Smartphones	Smartphone on waist, pouch	
MHEA-LTH [28]	12	ADLs	10	50Hz	Shimmer2 sensors	Right wrist, left ankle, chest	
UCI-HAR [29]	6	ADLs	30	50Hz	Smartphone	Left belt, no specific position	
OPPORTUNITY [30]	6	ADLs, Complex	4	-	Body-worn, Object, Ambient Sensors	Upper body, hip, leg, shoes	
WISDM [31]	18	ADLs, Postural, Complex	51	20Hz	Smartphone, Smartwatch	Right pant pocket, dominant hand (watch)	
UniMiB-SHAR [32]	17	ADLs, Falls	30	50Hz	Smartphone	Left & right trouser pockets	
Mobi-Act [33]	16	ADLs, Falls	66	100Hz	Smartphone	Trouser pocket	
HHAR [27]	6	Daily living activity, Sports fitness activity	9	Highest Available	Accelerometer, Gyroscope	Smartphone on waist, pouch	99.99%
MHEA-LTH [34]	12	Atomic activity, Daily living activity, Sports fitness activity	10	50Hz	Accelerometer, Gyroscope, Magnetometer, Electrocardiogram	Right wrist, left ankle, chest	97.83%
OPPORTUNITY [35]	17	Daily living activity, Composite activity	4	-	Acceleration, Rate of Turn, Magnetic field, Reed switches	Upper body, hip, leg, shoes	100%
WISDM [36]	6	Daily living activity, Sports fitness activity	33	20Hz	Accelerometer, Gyroscope	Right pant pocket, dominant hand (watch)	97.8%
UCI-HAR [37]	6	Daily living activity	30	50Hz	Accelerometer, Gyroscope	Left belt, no specific position	95.90%

Table 1. cont. List of Publicly Available HAR for AAL Related Benchmark Datasets

Dataset Name	Number of Activities	Activity Type	Number of Subjects	Sampling Frequency	Devices Used	Device Position	Latest Performances
PAMA-P2 [38]	18	Daily living activity, Sports fitness activity, Composite activity	9	100Hz	Accelerometer, Gyroscope, Magnetometer, Temperature	Wrist, chest, dominant ankle	94.72%, 82.12%, 90.27%
DSADS [39]	19	Daily living activity, Sports fitness activity	8	45Hz	Accelerometer, Gyroscope, Magnetometer	-	99.48%
Real-World [40]	8	Daily living activity, Sports fitness activity	15	7	Acceleration	-	95%
Exer. Activity [41]	10	Sports fitness activity	20	-	Accelerometer, Gyroscope	-	-
UTD-MHAD [42]	27	Daily living activity, Sports fitness activity, Composite activity, Atomic activity	8	-	Accelerometer, Gyroscope, RGB camera, depth camera	-	76.35%
TUD [43]	34	Daily living activity, Sports fitness, Composite activity	1	-	Accelerometer	-	-
SHAR [44]	17	Daily living activity, Sports fitness activity, Atomic activity	30	-	Accelerometer	Left & right trouser pockets	82.79%
USC-HAD [45]	12	Daily living activity, Sports fitness activity	14	-	Accelerometer, Gyroscope	-	97.25%
Mobi-Act [46]	13	Daily living activity, Atomic activity	50	100Hz	Accelerometer, Gyroscope, Orientation sensors	Trouser pocket	75.87%
Motion Sense [47]	6	Daily living activity	24	50Hz	Accelerometer, Gyroscope	-	95.35%
CASAS [48]	7	Daily living activity, Composite activity	1	-	Temperature, Infrared motion/light sensor	-	88.4%
Skoda [49]	10	Daily living activity, Composite activity	1	-	Accelerometer	-	97%
Wida-r3.0 [50]	6	Atomic activity	1	-	Wi-Fi	-	82.18%
HAPT [51]	12	Human activity	30	50Hz	Accelerometer, Gyroscope	-	92.14%, 98.73%

Table 2. Sensor data modality-based HAR models and performance.

Author	Year	Dataset Name	Modality Sensor Name	Methods	Classifier	Accuracy %
Ignatov et al. [52]	2018	WISDM UCI HAR	IMU Sensor	CNN	SoftMax	93.32 97.63
Jain et al. [53]	2018	UCI HAR	IMU Sensor	Fusion based	SVM, KNN	97.12
Chen et al. [54]	2019	MHEALTH PAMAP2 UCI HAR	IMU	CNN	SoftMax	94.05, 83.42 81.32
Alawneh et al. [55]	2020	UniMib Shar WISDM	Accelerometer IMU Sensor	Bi-LSTM	SoftMax	99.25 98.11
Lin et al. [56]	2020	Smartwach	Accelerometer gyroscope	Dilated CNN	SoftMax	95.49
Zhang et al. [57]	2020	WaFi CSI	WaFi signal	Dense-LSTM	SoftMax	90.0
Nadeem et al. [58]	2021	WISDM PAMAP2 USC-HAD	IMU	HMM	SoftMax	91.28 91.73 90.19
kavuncuoug et al. [59]	2021	Fall and ADLs	Accelerometer Gyroscope Magnetometer	ML	SVM, K-NN	99.96 95.27
Lu et al. [60]	2022	WISDM, PAMAP2 UCI-HAR	IMUs Accelerometers Accelerometers	CNN-GRU	SoftMax	96.41 96.25 96.67
Kim et al. [61]	2022	WISDM USC-HAR	IMUs	CNN-BiGRU	SoftMax	99.49 88.31
Sarkar et al. [62]	2023	UCI-HAR WISDM, MHEALTH PAMAP2 HHAR	IMUs Accelerometers Accelerometers	CNN with GA	SVM	98.74 98.34 99.72 97.55 96.87
Semwal et al. [63]	2023	WISDM PAMAP2 USC-HAD	IMUs	CNN and LSTM	SoftMax	95.76 94.64 89.83
Yao et al. [64]	2024	PAMAP2 USC-HAD, UniMib-SHAR OPPORTUNITY	IMUs Accelerometers	ELK ResNet	SoftMax	95.53 97.25 82.79 87.96
Wei et al. [65]	2024	WISDM PAMAP2 USC-HAD	IMU	TCN-Attention	SoftMax	99.03 98.35 96.32
El-Adawi et al. [66]	2024	MHEALTH	IMU	GAF+DenseNet169	SoftMax	97.83
Ye et al. [67]	2024	OPPT, PAMAP2	IMU	VAE-USM	GMM	100 82.12
Kaya et al. [68]	2024	UCI-HAPT WISDM, PAMAP2	IMU	Deep CNN	SoftMax	98 97.8 90.27
Zhang et al. [69]	2024	Shoaib, SisFall HCIHAR, KU-HAR	IMU	1DCNN-Att -BiLSTM	SVM	99.48 91.85 96.67 97.99
Zhang et al. [70]	2024	DSADS HAPT	IMU	Multi-STMT	SoftMax	99.86 98.73
Saha et al. [71]	2024	UCI HAR Motion-Sense	IMU	FusionActNet	SoftMax	97.35 95.35

4. Feature Extraction and Machine Learning Based Approach

4.1. Multi-View HAR for AAL

The presence of such datasets such as the RHM dataset is strongly favorable for multi-view Human Activity Recognition (HAR). These datasets all record an enormously large variety of activities from multiple views so that the models can learn to recognize human activities in three-dimensional space. Multi-view data eliminates the limitations of single-view systems such as occlusion and perspective distortion and thus enhances the ability of the system to recognize complex activities [25]. Such data like the RHM dataset are an important contribution towards the creation of multi-view HAR. They capture diverse activities from many different views so that models can be trained better to represent human actions in 3D. Multi-view data can provide relief from single-view system limitations such as occlusion and perspective distortion for the best opportunity for accurate recognition of complex activities [72]. Through the exploitation of sensor data from different views and perspectives, multi-view HAR facilitates the ability of the system to identify activities more accurately. It covers different applications like environmental sensors, wearable sensors, and vision-based systems and targets the

requirement for the unification of data sources across the applications. HAR system design in complex AAL environments comprises the combination of advanced methods like sensor fusion consisting of data-level, feature-level, and decision-level fusion.

4.2. Feature Extraction in Human Activity Recognition

The data segments obtained from the previous step are used for feature extraction, the most important stage in the Human Activity Recognition (HAR) pipeline. Feature extraction may be manual or automatic. Machine learning-based methods adopt these pipelines of data acquisition, data preprocessing and segmentation, feature extraction with hand-crafted features, feature selection, and finally, classification [73] depicts in Figure 1.

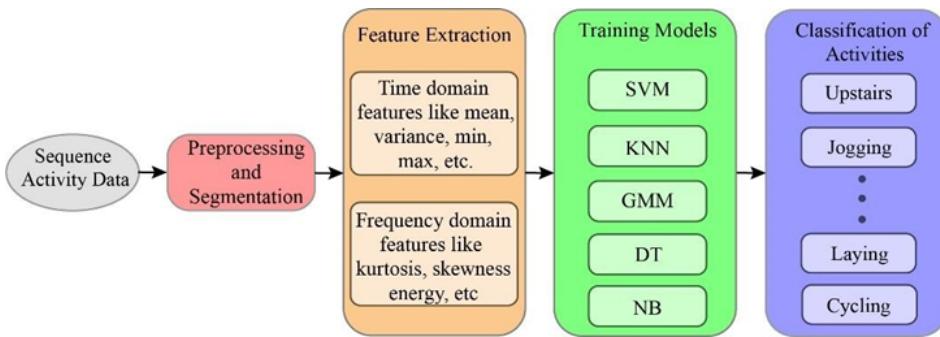


Figure 1. Human Activity Recognition using conventional ML-based techniques

4.3. Manual Feature Extraction

Domain knowledge is required to hand-engineer frequency domain features, time domain features, and other information features of signals. The majority of conventional ML classifiers like Naïve Bayes (NB), Random Forest (RF), and Support Vector Machines (SVM) have provided high accuracy in human activity recognition. Nevertheless, such conventional pattern recognition methods are feature-based and require domain knowledge. In addition, the learned features after embracing these procedures are heavily application-dependent and do not scale up.

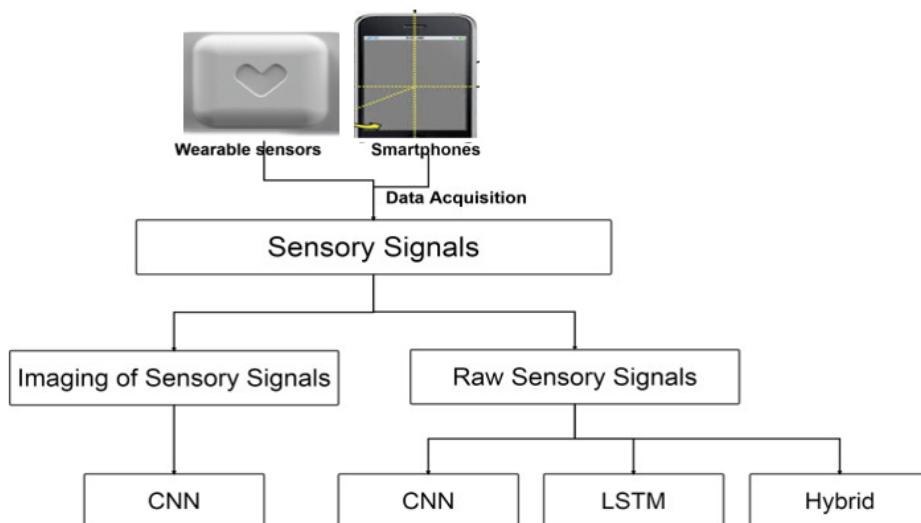


Figure 2. Categorization of the Deep Learning model

4.4. Feature Learning

DL-based approaches can learn features automatically without any manual feature engineering.

5. Deep Learning Models

Machine learning has attracted researchers to explore deep learning models for Human Activity Recognition (HAR) due to their potential for high reward. Figure 3 illustrates the HAR using DL approaches for feature extraction and classification. These models are usually categorized based on the input data they receive [24]. The HAR pipeline can be simplified with Deep Learning (DL) methods. DL has been shown to achieve outstanding empirical performance in many applications such as image synthesis, image segmentation, image classification, object detection, activity recognition, and human gait analysis [74]. DL algorithms are composed of multiple abstraction levels constructed through the neurons. Feature maps are obtained from the input feature maps of the preceding layer through the non-linear functions of each layer. Hierarchical abstraction enables DL algorithms to learn one level-specific feature for each domain of application. Deep Learning uses a Deep Neural Network (DNN) architecture where a specific loss function is optimized for feature extraction as well as determination of the classification boundaries [75].

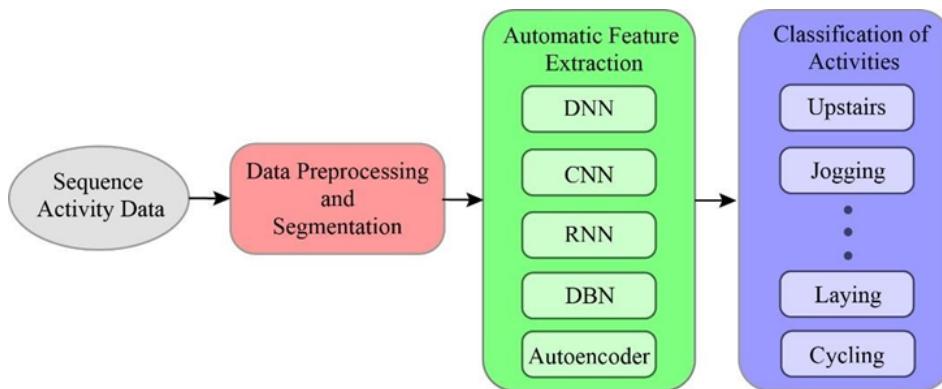


Figure 3. Flowchart of the DL based systems.

5.1. DL Categories Based on Sensory Cues

On the basis of the imagery signal type it can be divided into the following two categories:

5.1.1. Imaging of Sensory Signals

Convolutional Neural Networks (CNNs) are most appropriate for extracting the features of images and classifying large amounts of image data. The technique is also effective for time-series analysis where sensor data is converted into time-series images and passed through the CNN. A smartphone-based platform was shown by Alemayoh et al., where raw sensor data is accepted and 14×60 virtual images are constructed. A CNN (1C-1P-2C-2P-1DL) is employed to classify images into eight activities [76].

Some of the multimodal approaches, such as that of Lawal et al., incorporate wearable sensor signals into frequency images. Their effectiveness with the fusion technique is demonstrated with their two-stream CNN model [77]. Qin et al. propose heterogeneously processing sensor data, modeling the signals as images with two channels and propose residual fusion using the Lawal et al.'s layers [78]. It has been quantitatively evaluated and compared with state of the art on the HHAR and MHEALTH datasets [79].

5.2. Raw Sensory Data

Deep learning methodologies with their local dependency modeling and preservation of scaling-invariance are of immense usefulness for feature learning and classification of time series data. They are thus gaining increasing popularity for human activity recognition systems. Interdisciplinary research on CNNs, LSTMs, and hybrid deep learning techniques for improving the performance of HAR systems is an active research field [80].

5.3. Multi-View HAR with CNN-Based Methods in AAL

Convolutional Neural Networks (CNNs) have played a major role in Human Activity Recognition (HAR) modeling due to their capability of efficiently classifying images as well as raw time-series data (low-level information) of various sensors. CNNs are efficient due to their capability of discovering local dependencies, scale-invariant features, and learning complicated non-linear relationships within the input data—desired characteristics making CNNs of great utility for HAR applications [81]. CNNs were utilized to transform raw sensor data for HAR. Novel strategies are founded upon novel CNN structures for sensor data types, i.e., tri-axial accelerometer data typically used in smartphones. Small pool sizes of convolutional layers are best suited for extracting and discriminating valuable features with better activity recognition accuracy [82,83]. There has also been research that has combined CNNs with statistical features to learn local as well as global features of the signal. A hybrid approach like this maintains detailed information of the signal at various scales. Multi-parallel CNN systems have been suggested for user-dependent activity recognition to emphasize local feature extraction and combination of features. However, CNNs are computationally demanding, something often referred to as "frugality." This has led to efforts at optimization and hence efficient CNN models that are effective but less demanding. Techniques like Lego filters and Conditionally Parameterized Convolutions optimize the efficiency of the HAR process [84]. Also, applying pre-trained CNNs with dimensionality reduction like PCA and then hybrid classifiers like SVMs can optimize runtime—a major concern on smartphones where battery power is precious. Fair enough, CNNs are good at recognizing simple actions but discriminating more complex and related actions is not easy. Ensemble CNNs with different architecture are being suggested as the answer to make HAR systems more sensitive and specific [85]. Transfer learning schemes, for instance, the proposal of Heterogeneous Deep CNN (HDCNN) models, generalize across different sensor domains. Since HAR detectors are position-dependent on sensor placement, other studies have explored determinant-agnostic methods [86]. These systems attempt to work regardless of sensor placement on the body but lose accuracy at certain instances. Deep learning models are now being constructed to be sensor position robust with satisfactory performances on practical data. The "cold-start" problem or vanishing gradients in CNNs remains even now and has led to even more complex network architecture such as residual networks in CNNs. These are aimed towards more detailed activity-type recognition of varied dynamic activities for enhancing the applicability of HAR in Ambient Assisted Living (AAL) environments [87]. The future of HAR with CNNs in AAL environments requires continued research into architecture designs, hybrid models, and deep learning strategies. This will upgrade the perception, interaction, and functionality of HAR systems in AAL and promote the well-being and independence of users. Recent advancements in Human Activity Recognition (HAR) have seen the ubiquitous adoption of Convolutional Neural Networks (CNNs) for sensor data analysis of wearables and mobile devices. Ignatov et al. [52] employed a deep learning technique that takes advantage of CNNs for sensor data-based HAR in real-time. Their proposed technique aims at local feature extraction using CNNs with the assistance of simple-to-compute statistical features that detect global time series patterns. They tested their technique using UCI and WISDM datasets with high accuracy for various users and datasets. This is proof of their deep learning technique being effective without requiring complex computational hardware and hand-crafted feature engineering. Taking advantage of the power of CNNs, Chen et al. [54] proposed the semi-supervised deep learning method optimized for imbalanced HAR using multimodal wearable sensor data. Their model addresses the shortage of fewer labeled samples and class imbalance using the pattern-balanced model that finds heterogenous patterns. With the recurrent convolutional attention networks, their model is capable of finding the most significant features across sensor modalities, which highly empowers the model to work in imbalanced data environments. Kaya et al. [68] proposed the 1D-CNN-based method for accurate HAR from sensor data with special focus on raw accelerometer and gyroscope sensor data. Their model was tested on three publicly released datasets—UCI-HAPT, WISDM, and PAMAP2—demonstrating the strength and flexibility of the model

towards various data sources. The method simplifies feature extraction without accuracy degradation and thus is highly suitable for AAL applications in real-time. In another recent research work, Zhang et al. [88] proposed ConvTransformer, which is a novel method that combines CNNs, Transformer networks, and attention mechanisms to address the dual issue of extracting detailed and general features of sensor data. The hybrid method takes the power of CNNs in extracting local patterns while Transformers and attention mechanisms enable the model to learn global context, making it highly suitable for complex HAR in multi-view AAL environments.

5.4. RNN, LSTM, and Bi-LSTM in HAR for AAL

Recurrent Neural Networks (RNNs) have been at the core of recent Human Activity Recognition (HAR) studies due to their capacity in dealing with temporal dependencies in sensor data. For instance, Ordóñez et al. [89] demonstrated how RNNs play a crucial role in capturing the sequentiality of sensor data, which is critical in guaranteeing activity recognition accuracy. Traditional RNNs, however, are prone to suffering from gradient vanishing issues, a problem that necessitated the development of Long Short-Term Memory (LSTM) networks [90]. LSTM networks retain information for longer sequences, thus addressing the issues brought about by long-range dependencies in HAR applications.

LSTMs are suitable for Human Activity Recognition (HAR) since they are capable of processing sequential time-series data without vanishing or exploding gradients by maintaining long-term dependencies. While CNNs excel at image recognition through discovering spatial correlations, LSTM models use feedback connections to discover time patterns. A review of LSTM-based HAR systems, challenges, and future research directions is given. Early studies on on-body sensor placements explored LSTM models, but most did not crowdsource enough data from a diverse range of individuals to generalize to more than a few basic activities. On the computation limitation side, Agarwal et al. proposed an edge-device-friendly lightweight LSTM model but did not test it on complex activities. Extending this idea, Rashid et al. focused on energy efficiency and sensor data processing [91]. To overcome the vanishing gradient issue, Zhao et al. suggested a residual bi-directional LSTM architecture that concatenates forward and backward states to improve performance. Their approach maintained high performance in the dynamic and complex setting of HAR [91]. In [89], Wang and Liu suggested a Hierarchical Deep Long-Short Term Memory (H-LSTM) model, reducing the interference of noise and making feature learning easier in time-frequency domains. Similarly, Ashry et al. suggested a cascaded LSTM model that fused raw signal information and feature-based information, reducing the need for large-scale training datasets.

To address the local feature analysis weaknesses in traditional HAR approaches, a new hybrid deep learning framework was proposed. It took advantage of the convolutional capability of CNNs (layer-wise) for feature extraction, while LSTM units processed sequential information. The LSTM performance was also enhanced with the use of Extreme Learning Machine (ELM) classifiers. On the other hand, Zhou et al. addressed the problem of insufficient sensor labels. Their approach was to train a semi-supervised LSTM model combined with a Deep Q-Network, which annotates data automatically, improving performance in noisy and well-labeled dataset settings.

Researchers have also explored the use of attention-based Bidirectional LSTM (Bi-LSTM) models to further improve the performance of HAR systems [55,57,92,93]. Such models have been shown to achieve superior performance compared to other deep learning-based methods, as validated by experimental comparisons on several benchmark datasets. The results, as indicated in Table 2, show that attention-based Bi-LSTM models can achieve high accuracy for different user groups and datasets, while requiring relatively modest computational resources and without large-scale manual feature engineering.

Saha et al. [94] suggested Fusion ActNet, a new method for HAR that uses sensor data to differentiate between static and dynamic actions. The model includes individual residual networks for both actions and a decision guidance module. The method is trained in two stages and has been

thoroughly experimented upon using benchmark datasets, and it has been shown to be efficient in complex HAR scenarios, particularly in Ambient Assisted Living (AAL) environments where a wide range of activities have to be accurately detected. Murad et al. [90] have highlighted the merits of using Deep Recurrent Neural Networks (DRNNs) for HAR, specifically in capturing long-range dependencies in variable-length input sequences from body-worn sensors. Unlike traditional methods that fail to consider temporal correlations, DRNNs—unidirectional, bidirectional, and cascaded LSTM networks—yield state-of-the-art performance on several benchmark datasets. They compared DRNNs to traditional machine learning techniques like Support Vector Machines (SVM) and k-Nearest Neighbors (KNN) and other deep learning methods like Deep Belief Networks (DBNs) and CNNs, consistently demonstrating the superior performance of DRNNs in human activity recognition.

5.5. Integration of CNN and LSTM-Based Techniques

In recent the last several years, there has been significant progress development in the development design of hybrid deep learning models that combine different architectures to achieve high accuracy performance in Human Activity Recognition (HAR). One popular approach involves integrating to integrate Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks to leverage the strengths of both models. For Hybrid example, hybrid CNN-LSTM models, for example, have demonstrated been improved found to achieve superior performance in various several applications, including sleep-wake detection using heterogeneous sensors [63,95]. These models benefit exploit from the ability of CNNs' ability to extract local spatial features from sensor data and the ability of LSTMs' capability to capture model temporal dependencies, making them particularly extremely effective in HAR scenarios applications where both spatial and temporal information are crucial relevant.

Another notable design is the TCCSNet, which, along combined with CSNet, enhances human behavior detection recognition by leveraging both temporal and channel dependencies [96]. These architectures effectively combine the strengths of both CNNs and LSTMs, allowing for more accurate recognition of complex human activities by modeling both spatial and temporal aspects of the data. Ordóñez et al. further took explored this integration one step further by developing presenting a CNN-LSTM-based model for HAR [89]. Their approach involves relies extracting on features feature from extraction over raw sensor data by using CNNs, which are then additionally processed fed by to LSTM recurrent units to model complex temporal dynamics. This model also supports multimodal sensor fusion, enabling meaning it to can handle process data from multiple sensor types without of the sensors need without for manual feature design engineering. The evaluation on benchmark datasets such as Opportunity and Skoda revealed showed significant performance improvements over traditional methods, underscoring showcasing the effectiveness of this hybrid approach in HAR applications.

Zhang et al. introduced suggested a multi-channel deep learning network, called the 1DCNN-Att-BiLSTM hybrid model, which that combines 1D CNNs with attention mechanisms and Bidirectional LSTM units [57]. This The model was evaluated using publicly accessible available datasets, and the results performance metrics showed improved recognition performance compared to both machine learning and deep learning models. The Hybridization hybrid approach of integrating CNNs and LSTMs, along with attention mechanisms, provides presents a robust promising framework for to handle solve the complexities problems of multi-view HAR in Ambient Assisted Living (AAL) settings.

El-Adawi et al. developed suggested a HAR model within in the context of a Wireless Body Area Network (WBAN) context, environment utilizing with a novel approach that of integrates integrating the Gramian Angular Field (GAF) and DenseNet [66]. In this model, time series data is converted into to 2D images using the GAF technique, which are is then input processed to by DenseNet in order to achieve high accuracy in performance accuracy. This innovative novel method approach demonstrates

the potential of combining integrating different deep learning techniques approaches and data transformation methods to enhance HAR, particularly in environments like AAL where sensor data from body-worn wearable devices sensors is prevalent.

5.6. TCN Approaches for Multi-View Human Activity Recognition in Ambient-Assisted Living

Recent advancements in Human Activity Recognition (HAR), particularly within the context of Ambient-Assisted Living (AAL) paradigm, have been increasingly focused towards leveraging deep learning models to process the processing and analyze the analysis of data collected from various sensors. These sensors, which can be categorized as user-driven, environment-driven, or object-driven, play a crucial role in capturing the intricacies dynamics of human activity across different contexts environments [97]. Traditional methods have struggled to be effective in capturing the temporal dependencies inherent in sensor-driven time series data. However, recent progress in deep learning, particularly through the use of Transformer models founded on multi-head attention mechanisms, has significantly enhanced the ability to model these temporal relationships dependencies [98]. This capability is essential for accurately identifying, recognizing, and predicting activities in dynamic AAL environments with precision, where the activities of interest may change over time.

One of the key challenges in HAR is the making adaptation of these systems adaptable to recognize new activities in ever-changing environments. To address this, researchers have emphasized the importance of incorporating information about sensor frequency and conducting thorough analyses of both time and frequency domains. This approach not only improves the understanding of sensor-driven time series data, but also facilitates the development of more robust and efficient HAR systems [99]. In response to these challenges, Kim et al. [97] introduced a Contrastive Learning-based Novelty Detection (CLAN) method for HAR that utilizes sensor data. The CLAN method is particularly effective in handling temporal and frequency features, as well as the complex dynamics of activities and variations in sensor modalities diversity. By leveraging data augmentation to create diverse negative pairs, this method enhances the ability to detect novel activities even when they share common features with known activities. The two-tower model employed by CLAN extracts invariant representations of known activities, thus improving the system's ability to recognize new activities with similar characteristics.

Extending on these concepts, Wei et al. [65] proposed a Time Convolutional Network with Attention Mechanism (TCN-Attention-HAR) specifically designed to optimize HAR using wearable sensor data. The TCN-Attention-HAR model addresses critical fundamental challenges in temporal feature extraction and the gradient issues commonly encountered by deep networks. By optimizing the optimization of temporal convolution sizes and incorporating the use of attention mechanisms, the model effectively can prioritize and concentrate on important information, resulting in more accurate activity recognition. This approach is particularly advantageous in AAL scenarios applications, where the need for real-time, and lightweight processing is paramount. Zhang et al. [93] further advanced the field with the development of Multi-STMT, a multilevel model that integrates spatiotemporal attention and multiscale temporal embedding to enhance HAR using wearable sensors. The Multi-STMT model combines Convolutional Neural Networks (CNNs) with Bidirectional Gated Recurrent Units (BiGRUs) and attention mechanisms to capture subtle differences between human activities. This integration makes the model capable of distinguishing between similar activities effectively, making it particularly valuable in multi-view HAR applications within AAL environments. Additionally, the Conditional Variational Autoencoder with Universal Sequence Mapping (CVAE-USM) proposed by Zhang et al.

[93] addresses the challengeproblem of non-independent and identically distributed (non-i.i.d.) data in cross-user scenarios. By leveraging the temporal relationshipsdependencies withinin time-series data and combining Variational Autoencoder (VAE) withand Universal Sequence Mapping (USM) techniques, CVAE-USM effectively aligns user data distributions effectively, and capturingextracts common temporal patterns acrossamong differentvarious users. This approach significantly enhances the accuracy of activity recognition accuracy, particularly in multi-view AAL settingsscenarios where between-user data variabilityvariation across users is a majorserious concern.

5.6.1. TCN-Based Methods in Multi-View HAR for AAL

The adoption of Temporal Convolutional Networks (TCNs) in multi-view HAR for AAL environments offers unique advantages, particularly in terms of efficient temporal modeling. TCNs are specifically designed to handle sequential data by applying convolutional operations along the temporal dimension. This structure allows TCNs to effectively capture long-range dependencies and patterns in activity sequences, which are crucial for accurate recognition in multi-view settings. Moreover, the integration of attention mechanisms within TCN frameworks, as seen in the TCN-Attention-HAR model, allows for dynamic weighting of important temporal features, ensuring that the most relevant information is emphasized during the recognition process. This capability is particularly beneficial in AAL environments, where real-time decision-making is critical, and the system must be lightweight enough to operate efficiently on limited hardware.

The combination of TCNs with other deep learning models, such as CNNs and BiGRUs, further enhances their effectiveness in multi-view HAR. These hybrid models, like the Multi-STMT, can simultaneously capture spatial and temporal dependencies, providing a comprehensive understanding of human activities across different views. This multi-view capability is essential in AAL environments, where activities may be observed from various angles and perspectives.

In summary, the integration of TCN-based methods into multi-view HAR systems for AAL represents a significant advancement in the field. These methods not only improve the accuracy and efficiency of activity recognition but also ensure that the systems remain lightweight and responsive—key requirements for real-world applications in Ambient-Assisted Living.

5.7. GCN-Based Multi-View HAR for AAL

Graph Convolutional Neural Networks (GCNNs) have emerged as a crucial tool in the field of Human Action Recognition (HAR), especially within the context of multi-view analysis for Ambient-Assisted Living (AAL) environments. Unlike traditional deep learning models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which are designed to handle Euclidean data structures such as images, text, and sequential data, GCNNs excel at processing non-Euclidean data where the relationships between data points are more complex and graph-structured. This unique capability makes GCNNs particularly well-suited for HAR in AAL settings, where understanding the nuanced movements of individuals across multiple views is critical.

5.7.1. The Role of GCNs in Multi-View HAR

GCNs, first introduced by Franco Scarselli, have since evolved into a powerful framework for learning from graph-structured data [100]. In the context of HAR, human skeleton data, which consist of joint coordinates and their interconnections, can be naturally represented as a graph. Each joint is modeled as a node, while the connections between joints (e.g., bones) are treated as edges. This graph-based representation enables GCNNs to capture the intricate spatial dependencies between joints, which is essential for recognizing complex actions, especially when viewed from multiple angles.

In multi-view HAR, the challenge is further compounded by the need to integrate information from different perspectives. GCNs are uniquely positioned to address this challenge due to their ability

to aggregate and process information across multiple views, ensuring that the resulting models can effectively generalize across varying conditions typical of AAL environments.

5.7.2. Spectral GCN-Based Methods

Spectral GCNs leverage the principles of spectral graph theory, utilizing the eigenvalues and eigenvectors of the Graph Laplacian Matrix (GLM) to transform graph data from the spatial domain into the spectral domain [101]. This transformation facilitates the application of convolutional operations on graph-structured data. However, while effective, spectral GCNs have historically faced limitations in computational efficiency—a significant consideration in lightweight models for AAL applications. Kipf et al. [102] addressed this limitation by introducing a simplified version of spectral GCNs, where the filter operation is constrained to only one-hop neighbors. This enhancement significantly reduces computational overhead, making spectral GCNs more feasible for real-time HAR applications, including those required in AAL settings. Nevertheless, the fixed nature of the graph structure in spectral GCNs can be a limitation when dealing with the dynamic and variable nature of human actions viewed from multiple perspectives.

5.7.3. Spatial GCN-Based Methods

Spatial GCNs have become the focal point of research in GCN-based HAR, particularly for multi-view scenarios. Unlike spectral GCNs, spatial GCNs operate directly on the graph in the spatial domain, allowing for more intuitive and flexible modeling of dynamic human actions. This flexibility is especially crucial in AAL environments, where actions must be recognized accurately across different views and conditions.

A significant advancement in this domain was the introduction of the Spatio-Temporal Graph Convolutional Network (ST-GCN) by Yan et al. [103]. The ST-GCN model is designed to handle both spatial and temporal aspects of human motion, capturing the evolving relationships between joints over time. In the context of multi-view HAR, ST-GCNs can effectively integrate information from various viewpoints, providing a holistic understanding of the action being performed.

Further enhancing the flexibility of GCNs in multi-view HAR, Shi et al. [104] developed the Two-Stream Adaptive Graph Convolutional Network (2s-AGCN). This model introduces an adaptive mechanism that allows the network to learn the graph topology dynamically, making it highly adaptable to diverse datasets and varying viewpoints typical in AAL settings. The inclusion of an attention mechanism further improves the model's ability to focus on the most critical joints and their connections, ensuring robustness in action recognition across multiple views.

Shiraki et al. [105] advanced this concept with the Spatiotemporal Attentional Graph Convolutional Network (STA-GCN), which specifically addresses the varying importance of joints in different actions. STA-GCN not only considers the spatial relationships but also the temporal significance of joints, making it particularly effective for multi-view HAR where the importance of certain joints may vary depending on the action and the viewpoint.

5.7.4. Recent Innovations of GCN

Recent innovations, such as the Shift-GCN model introduced by Shi et al. [106], have pushed the boundaries of multi-view HAR further by expanding the receptive field of spatiotemporal graphs and employing lightweight techniques to reduce computational costs [107]. This approach aligns well with the goals of AAL, where lightweight models must balance efficiency and accuracy. Other approaches, such as the Partial-Based Graph Convolutional Network (PB-GCN) developed by Thakkar et al. [108] and Li et al. [56], focus on segmenting the human skeleton into distinct parts, allowing the network to learn more focused and specialized representations of human actions. These methods are particularly useful in multi-view HAR, where the ability to isolate and analyze specific body parts can lead to more accurate recognition of complex actions.

The use of GCN-based methods in multi-view HAR for AAL is unique due to their ability to capture and integrate spatial-temporal dependencies from multiple perspectives. These methods allow for the development of lightweight, yet highly accurate models that are essential for real-time applications in AAL. The ongoing advancements in spatial GCNs, particularly with the integration of attention mechanisms and adaptive learning techniques, are setting new standards in the field, enabling more robust and flexible human action recognition systems. By leveraging these innovations, your survey paper aims to highlight the cutting-edge developments in GCN-based multi-view HAR, offering insights into how these methods can be optimized for AAL applications. This focus on lightweight, efficient models makes your work particularly relevant for practical implementations, ensuring that the systems are both scalable and effective in real-world AAL environments. In summary, the unique advantages of GCN-based methods, particularly the flexibility and efficiency offered by spatial GCNs, make them highly suitable for multi-view human activity recognition in ambient-assisted living scenarios. By continuously refining these models and integrating innovative techniques like attention mechanisms and residual connections, researchers are pushing the boundaries of what is possible in HAR, making systems more accurate, adaptable, and efficient.

Table 3. Summary of challenges and future enhancements in HAR

Challenge	Description
Scarcity of annotated data	Deep learning models are trained and evaluated on labeled data. However, all sensory data need to be collected and labeled, which makes it costly and time-consuming.
Limited performance in outdoor environments	HAR models are known to have poor performance in outdoor environments due to variable lighting, noise, and different weather conditions.
Difficulty recognizing complex activities and postural transitions	Simple activities and postural transitions may not be properly recognized by many HAR models.
Hyperparameter optimization	HAR models can particularly affect accuracy if the requisite hyperparameters are not properly tuned.
Fusion of multimodal sensor data	By fusing data from different sensors, data of various formats and resolutions can be combined to enhance the information available to HAR systems.
Performance limitations in unsupervised learning	One of the most common problems with unsupervised HAR models is their accuracy compared to supervised models.

5.8. Transfer Learning-Based Lightweight Deep Learning

Traditional deep learning models are computationally intensive and challenging to deploy in real-time edge-based Ambient Assisted Living (AAL) systems due to their high computational demands. The previous chapter discussed the architecture and functioning of lightweight models such as MobileNet, SqueezeNet, and EfficientNet, along with approaches to reduce computational requirements, including depth-wise separable convolutions and squeeze-and-excitation blocks [109–111]. Furthermore, model compression and quantization techniques are reviewed to enhance the deployability of deep learning models in AAL systems [112].

Many researchers have introduced lightweight deep learning models that balance computational efficiency and model performance. These models, optimized for real-time processing, are crucial for AAL applications, where precise and fast activity recognition is essential [113,114]. Techniques such as model pruning, quantization, and knowledge distillation enable the derivation of compact and efficient models without significantly compromising accuracy [115,116].

This section focuses on the evolution of Multi-View Human Activity Recognition (MV-HAR) beyond traditional methodologies and explores how integrating multiple viewpoints enhances performance. By leveraging multiple sensors or cameras positioned at different angles, MV-HAR improves

visibility and ensures highly accurate activity detection [117]. The significance of energy-efficient processing in real-time applications, such as mobile devices and large-scale smart environments, is underscored. In the context of assistive living, personalized service options and proactive care strategies play a crucial role [101,118].

By integrating multiple camera perspectives, we demonstrate significant improvements in recognition accuracy, offering a scalable and practical solution for AAL systems with constrained computational resources. This initiative aims to enhance the quality and customization of support services for residents, including fall detection and home security. Leveraging advancements in machine learning, the proposed approach promotes safer and more comfortable living environments.

5.8.1. Resource Efficiency with Lightweight Deep Learning

Deep learning models scale out at the expense of higher computational complexity and storage. To mitigate these and improve model size and speed, lightweight deep learning techniques have been developed, including model compression, pruning, quantization, and knowledge distillation. These techniques aim for smaller and faster models with as minimal compromises in accuracy as possible and are thus suitable for power-constrained devices typical of most AAL settings.

6. Problems, Challenges, and Current/Future Directions

Although multi-view Human Activity Recognition (HAR) has made significant advances, it is still faced with many challenges, such as data synchronization, privacy protection, and the need for universal models. Solving these challenges will require new solutions, such as new synchronization techniques, secure data processing, and adaptive models—models that do not require large amounts of data to train and can simply adapt to changes in the environment. With the rapid development of micro-electromechanical systems and sensor technology, the recognition task of HAR has gained increasing attention in recent years. It is a crucial problem for a variety of applications such as ambient assisted living, smart homes, sports, and work injury detection. HAR is a new field where recent development driven by deep learning has opened up new possibilities. However, these solutions are still limited by problems that restrict their practical applicability.

Labeled data is a precious commodity in NLP and one of the largest bottlenecks in HAR. Deep learning models require vast quantities of labeled data to train and test. However, acquiring and labeling this data is a time-consuming and costly process. This shortage of data can lead to models overfitting and failing to generalize well to real-world situations [?]. HAR models perform poorly in outdoor environments due to parameter variation when counting specific movements, i.e., push-ups and squats. Most HAR models are trained and tested in controlled laboratory environments. Outdoor environments are uncontrollable and contain several dozen factors—i.e., light, noise, and weather—that greatly affect sensor data. Therefore, high-performance HAR models in controlled environments may not offer the same accuracy outdoors. In addition, HAR models are challenged when they have to handle complex activities and postural transitions. Most HAR tasks consider simple activities such as walking, sitting, and sleeping. Daily activities also involve cooking, cleaning, and dressing. Moreover, HAR models are not effective in identifying falls and other postural transitions. One of the pitfalls involves the need for meticulous hyperparameter tuning in order to realize the best model performance. Hyperparameters are settings of learning algorithms that one may adjust to derive optimum results. However, the vast number of hyperparameters can render it challenging to determine the most suitable settings for any problem. One other challenge is presented by the fusion of multimodal sensor data. Most HAR systems utilize only one type of sensor, e.g., an accelerometer. However, fusion of data from different sensors—e.g., gyroscopes and magnetometers—can enhance the performance of HAR. This leads to the challenge of how to fuse data from such heterogeneous sources optimally. The final limitation of HAR models is with regard to their expected performance in unsupervised learning. Supervised machine learning requires human labeling, while clustering is akin

to supervised learning without the human element. Unsupervised learning differs from supervised learning because the latter requires labeled data, which is typically expensive and time-consuming to create. Yet, the majority of unsupervised HAR models are inferior to even the poorer performing supervised models. In summary, there are some challenges for HAR that need to be considered to advance the field and improve its applications. Addressing these challenges will enable researchers to develop more precise and suitable HAR models for various environments. Although MV-HAR, when combined with lightweight models, holds promise for ambient assisted living (AAL), there remain several challenges to be tackled, including:

- **Limited Multi-view AAL Datasets:** There is a shortage of multi-view datasets for AAL, making it difficult to develop and compare models.
- **Real-World Deployment:** Privacy concerns, environmental robustness, and typical real-world AAL settings must be taken into account.

In summary, HAR for AAL using data acquired from diverse angles and leveraging low-cost deep learning approaches has a promising future. New directions such as sensor fusion, transfer learning, and latent domain adaptation offer potential for enhancing HAR systems. To solve these issues, the research community must adopt an interdisciplinary approach to developing systems that are not just accurate, but also privacy-compliant, and adaptable to individual needs. According to the insights provided in these researches, this review will synthesize the knowledge that is available, enlighten the gaps that are present, and provide future research directions.

7. Conclusions

Lightweight Deep Learning approach technique cited referenced synergizes multi-view HAR and is shown proven to realize have high potential in accuracy and efficiency of the implementation of Activity Recognition in Ambient Assisted Living (AAL). Significant Great progress efforts have been made by researchers towards the direction creation of building reliable HAR systems that with translated direct translation to improvement in of the quality of life of assisted living individuals users. Limitations Constraints in related terms to computational requirements, sensor fusion combination, and robustness are overcome, ultimately leading the way to more effective and userfriendly-friendly to-use AAL systems. What are some promising directions for future research directions? Future advancements of HAR systems can be achieved by through further development evolutions of lightweight architectures architecture, exploration of transfer learning for personalization, and robust privacy-preserving techniques mechanisms. These efforts are accompanied with the development production of multi-view datasets with a large range activity of variety activities and are essential required for the advancement advancements of AAL to address reach the elderly and disabled people in order to help allow them attain to achieve acceptable standards levels of safety, health and quality of life.

Author Contributions: Conceptualization, Fahmid Al Farid; Ahsanul Bari; Abu Saleh Musa Miah; Tanjil Sarkar; Zulfadzli Yusoff; Jia Uddin; Hezerul Abdul Karim; Software, Fahmid Al Farid; Ahsanul Bari; Prabha Kumaresan; Validation, Fahmid Al Farid; Ahsanul Bari; Abu Saleh Musa Miah; Formal analysis, Fahmid Al Farid; Ahsanul Bari; Tanjil Sarkar; Zulfadzli Yusoff; Sarina Mansur; Hezerul Abdul Karim; Investigation, Fahmid Al Farid; Tanjil Sarkar; Zulfadzli Yusoff; Sarina Mansur; Data curation, Fahmid Al Farid; Ahsanul Bari; Abu Saleh Musa Miah; Writing—original draft preparation, Fahmid Al Farid; Ahsanul Bari; Abu Saleh Musa Miah; Jia Uddin; Prabha Kumaresan; Hezerul Abdul Karim; Writing—review and editing, Fahmid Al Farid; Ahsanul Bari; Tanjil Sarkar; Zulfadzli Yusoff; Sarina Mansur; Jia Uddin; Prabha Kumaresan; Visualization, Fahmid Al Farid; Ahsanul Bari; Abu Saleh Musa Miah; Hezerul Abdul Karim; Supervision, Tanjil Sarkar; Zulfadzli Yusoff; Sarina Mansur; Prabha Kumaresan; Funding acquisition, Tanjil Sarkar; Zulfadzli Yusoff; Sarina Mansur; Jia Uddin; All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information Communications Technology Planning Evaluation (IITP) grant funded by the Korea government (MSIT) (IITP-2024-2020-0-01791).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

LSTM	Long-Short-Term-Memory
BiLSTM	Bi-Directional long short-term Memory
CNN	Convolutional Neural Networks
GMM	Gaussian Mixture Model
DD-Net	Double-feature Double-motion Network
GAN	Generative Adversarial Network
SOTA	State-of-the-Art
MobileNetV2	Mobile Network Variant 2
ReLU	Rectified Linear Unit
DCNN	Dilated Convolutional Neural Network
GRNN	General regression neural network
KFDI	Key Frames Dynamic Image (
QST	Quaternion Spatial-Temporale
RNNs	Recurrent Neural Networks
BT-LSTM	Block-Term Long-Short-Memory
KF	Kalman filter
KM-Model	Keypoint mesh model
SRGB-Model	Segmented RGB model

References

1. Chen, T.; Zhou, D.; Wang, J.; Wang, S.; Guan, Y.; He, X.; Ding, E. Learning Multi-Granular Spatio-Temporal Graph Network for Skeleton-based Action Recognition, 2021. <https://doi.org/10.1145/3474085.3475574>.
2. Miah, A.S.M.; Islam, M.R.; Molla, M.K.I. Motor imagery classification using subband tangent space mapping. In Proceedings of the 2017 20th International Conference of Computer and Information Technology (ICCIT). IEEE, 2017, pp. 1–5. <https://doi.org/10.1109/ICCITECHN.2017.8281828>.
3. Miah, A.S.M.; Islam, M.R.; Molla, M.K.I. EEG classification for MI-BCI using CSP with averaging covariance matrices: An experimental study. In Proceedings of the 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2). IEEE, 2019, pp. 1–5. <https://doi.org/10.1109/IC4ME247184.2019.9036591>.
4. Tusher, M. M. R., F.F.A.K.H.M.M.A.S.M.R.S.R.I.M.R.M.A.M.S..K.H.A. BanTrafficNet: Bangladeshi Traffic Sign Recognition Using A Lightweight Deep Learning Approach. *Computer Vision and Pattern Recognition*. <https://doi.org/10.21203/rs.3.rs-4216970/v1>.
5. Zobaed, T.; Ahmed, S.R.A.; Miah, A.S.M.; Binta, S.M.; Ahmed, M.R.A.; Rashid, M. Real time sleep onset detection from single channel EEG signal using block sample entropy. In Proceedings of the IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2020, Vol. 928, p. 032021. <https://doi.org/10.1088/1757-899X/928/3/032021>.
6. Ali, M.S.; Mahmud, J.; Shahriar, S.M.F.; Rahmatullah, S.; Miah, A.S.M. Potential Disease Detection Using Naive Bayes and Random Forest Approach. *BAUST Journal* **2022**.
7. Hossain, M.M.; Chowdhury, Z.R.; Akib, S.M.R.H.; Ahmed, M.S.; Hossain, M.M.; Miah, A.S.M. Crime Text Classification and Drug Modeling from Bengali News Articles: A Transformer Network-Based Deep Learning Approach. In Proceedings of the 2023 26th International Conference on Computer and Information Technology (ICCIT). IEEE, 2023, pp. 1–6. <https://doi.org/10.1109/ICCIT60459.2023.10441195>.

8. Rahim, M.A.; Farid, F.A.; Miah, A.S.M.; Puza, A.K.; Alam, M.N.; Hossain, M.N.; Karim, H.A. An Enhanced Hybrid Model Based on CNN and BiLSTM for Identifying Individuals via Handwriting Analysis. *CMES-Computer Modeling in Engineering and Sciences* **2024**, *140*. <https://doi.org/10.32604/cmes.2024.048714>.
9. Miah, A.S.M.; Ahmed, S.R.A.; Ahmed, M.R.; Bayat, O.; Duru, A.D.; Molla, M.K.I. Motor-Imagery BCI task classification using riemannian geometry and averaging with mean absolute deviation. In Proceedings of the 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT). IEEE, 2019, pp. 1–7. <https://doi.org/10.1109/EBBT.2019.8741603>.
10. Kibria, K.A.; Norman, A.S.; Hossain, M.A.; Bulbul, M.S.I.; Rashid, M.M.; Miah, A.S.M. Creation of a Cost-Efficient and Effective Personal Assistant Robot using Arduino Machine Learning Algorithm. In Proceedings of the 2020 IEEE Region 10 Symposium (TENSYMP). IEEE, 2020, pp. 477–482. <https://doi.org/10.1109/TENSYMP50017.2020.9230773>.
11. Hossain, M.M.; Norman, A.S.; Begum, M.M.; Warka, W.A.; Hossain, M.M.; Miah, A.S.M. Exploring Bangladesh's Soil Moisture Dynamics via Multispectral Remote Sensing Satellite Image. *European Journal of Environment and Earth Sciences* **2023**, *4*, 10–16. <https://doi.org/https://doi.org/10.24018/ejgeo.2023.4.5.415>.
12. Rahman, M.R.; Hossain, M.T.; Nawal, N.; Sujon, M.S.; Miah, A.S.M.; Rashid, M.M. A Comparative Review of Detecting Alzheimer's Disease Using Various Methodologies. *BAUST Journal* **2020**.
13. Md. Mahbubur Rahman Tusher, Fahmid Al Farid, M.A.H.A.S.M.M.S.R.R.M.H.J.S.M.M.A.R.H.A.K. Development of a Lightweight Model for Handwritten Dataset Recognition: Bangladeshi City Names in Bangla Script. *Computers, Materials & Continua* **2024**, *80*, 2633–2656. <https://doi.org/10.32604/cmc.2024.049296>.
14. Obinata, Y.; Yamamoto, T. Temporal Extension Module for Skeleton-Based Action Recognition. In preparation, <https://doi.org/10.1109/ICPR48806.2021.9412113>.
15. Lecun, Y.; Bottou, E.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition 1998. <https://doi.org/10.1109/5.726791>.
16. Chen, K.; Zhang, D.; Yao, L.; Guo, B.; Yu, Z.; Liu, Y. Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Computing Surveys (CSUR)* **2021**, *54*, 1–40. <https://doi.org/https://doi.org/10.1145/3447744>.
17. Xiao, Y.; Chen, J.; Wang, Y.; Cao, Z.; Zhou, J.T.; Bai, X. Action Recognition for Depth Video using Multi-view Dynamic Images. *arXiv preprint arXiv:1806.11269* **2018**. <https://doi.org/10.1016/j.ins.2018.12.050>.
18. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep High-Resolution Representation Learning for Human Pose Estimation. *arXiv preprint arXiv:1902.09212* **2019**. <https://doi.org/10.48550/arXiv.1902.09212>.
19. Luvizon, D.C.; Picard, D.; Tabia, H. 2D/3D Pose Estimation and Action Recognition using Multitask Deep Learning. *arXiv preprint arXiv:1802.09232* **2018**. <https://doi.org/10.1016/j.ins.2018.12.050>.
20. Liu, J.; Shahroudy, A.; Perez, M.; Wang, G.; Duan, L.Y.; Kot, A.C. NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2020**, *42*, 2684–2701. <https://doi.org/https://doi.org/10.1109/TPAMI.2019.2916873>.
21. Yang, D.; Li, M.M.; Fu, H.; Fan, J.; Zhang, Z.; Leung, H. Unifying Graph Embedding Features with Graph Convolutional Networks for Skeleton-based Action Recognition. *arXiv preprint arXiv:2003.03007* **2020**. <https://doi.org/10.48550/arXiv.2003.03007>.
22. Guerra, B.M.V.; Torti, E.; Marenzi, E.; Schmid, M.; Ramat, S.; Loporati, F.; Danese, G. Ambient assisted living for frail people through human activity recognition: state-of-the-art, challenges and future directions. *Frontiers in Neuroscience* **2023**, *17*. <https://doi.org/https://doi.org/10.3389/fnins.2023.1256682>.
23. Duan, H.; Zhao, Y.; Chen, K.; Lin, D.; Dai, B. Revisiting Skeleton-based Action Recognition, 2022, [arXiv:cs.CV/2104.13586]. <https://doi.org/10.48550/arXiv.2104.13586>.
24. Zhao, Z.; Zhang, L.; Shang, H. A Lightweight Subgraph-Based Deep Learning Approach for Fall Recognition. *Sensors* **2022**, *22*, 1–14. <https://doi.org/10.3390/s22155482>.
25. Action, S.G.H. Skeleton Graph-Neural-Network-Based Human Action **2022**. <https://doi.org/10.3390/s22062091>.
26. Reiss, A. PAMAP2 Physical Activity Monitoring. UCI Machine Learning Repository, 2012. <https://doi.org/10.24432/C5NW2H>.
27. Blunck, Henrik, B.S.P.T.K.M.; Dey, A. Heterogeneity Activity Recognition. UCI Machine Learning Repository, 2015. <https://doi.org/10.24432/C5689X>.
28. Banos, Oresti, G.R.; Saez, A. MHEALTH. UCI Machine Learning Repository, 2014. <https://doi.org/10.24432/C5TW22>.

29. Reyes-Ortiz, Jorge, A.D.G.A.O.L.; Parra, X. Human Activity Recognition Using Smartphones. UCI Machine Learning Repository, 2013. <https://doi.org/10.24432/C54S4K>.

30. Roggen, Daniel, C.A.N.D.L.V.C.R.; Sagha, H. OPPORTUNITY Activity Recognition. UCI Machine Learning Repository, 2010. <https://doi.org/10.24432/C5M027>.

31. Weiss, G. WISDM Smartphone and Smartwatch Activity and Biometrics Dataset . UCI Machine Learning Repository, 2019. <https://doi.org/10.24432/C5HK59>.

32. Micucci, D.; Mobilio, M.; Napoletano, P. UniMiB SHAR: a new dataset for human activity recognition using acceleration data from smartphones, 2017, [arXiv:cs.CV/1611.07688]. <https://doi.org/10.3390/app7101101>.

33. Chatzaki, C.; Pediaditis, M.; Vavoulas, G.; Tsiknakis, M. Human Daily Activity and Fall Recognition Using a Smartphone's Acceleration Sensor. In Proceedings of the Information and Communication Technologies for Ageing Well and e-Health; Röcker, C.; O'Donoghue, J.; Ziefle, M.; Helfert, M.; Molloy, W., Eds., Cham, 2017; pp. 100–118. https://doi.org/10.1007/978-3-319-62704-5_7.

34. Baños, O.; García, R.; Terriza, J.A.H.; Damas, M.; Pomares, H.; Rojas, I.; Saez, A.; Villalonga, C. mHealth-Droid: A Novel Framework for Agile Development of Mobile Health Applications. In Proceedings of the International Workshop on Ambient Assisted Living and Home Care, 2014. https://doi.org/10.1007/978-3-319-13105-4_14.

35. Roggen, D.; Calatroni, A.; Rossi, M.; Holleczek, T.; Förster, K.; Tröster, G.; Lukowicz, P.; Bannach, D.; Pirk, G.; Ferscha, A.; et al. Collecting complex activity datasets in highly rich networked sensor environments. *2010 Seventh International Conference on Networked Sensing Systems (INSS) 2010*, pp. 233–240. <https://doi.org/10.1109/INSS.2010.5573462>.

36. Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter* **2011**, *12*, 74–82. <https://doi.org/10.1145/1964897.1964918>.

37. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. A Public Domain Dataset for Human Activity Recognition using Smartphones. In Proceedings of the The European Symposium on Artificial Neural Networks, 2013.

38. Reiss, A.; Stricker, D. Introducing a New Benchmarked Dataset for Activity Monitoring. *2012 16th International Symposium on Wearable Computers 2012*, pp. 108–109. <https://doi.org/10.1109/ISWC.2012.13>.

39. Barshan, B.; Altun, K. Daily and Sports Activities. UCI Machine Learning Repository, 2010. <https://doi.org/10.24432/C5C59F>.

40. Sztylet, T.; Stuckenschmidt, H. On-body localization of wearable devices: An investigation of position-aware activity recognition. In Proceedings of the 2016 IEEE international conference on pervasive computing and communications (PerCom). IEEE, 2016, pp. 1–9. <https://doi.org/10.1109/PERCOM.2016.7456521>.

41. betul oktay.; Sabir, M.; Tuameh, M. Fitness Exercise Pose Classification. <https://kaggle.com/competitions/fitness-pose-classification>, 2022. Kaggle.

42. Chen, C.; Jafari, R.; Kehtarnavaz, N. UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In Proceedings of the 2015 IEEE International conference on image processing (ICIP). IEEE, 2015, pp. 168–172. <https://doi.org/10.1109/ICIP.2015.7350781>.

43. Morris, C.; Krieg, N.M.; Bause, F.; Kersting, K.; Mutzel, P.; Neumann, M. Tudataset: A collection of benchmark datasets for learning with graphs. *arXiv preprint arXiv:2007.08663* **2020**. <https://doi.org/10.48550/arXiv.2007.08663>.

44. Micucci, D.; Mobilio, M.; Napoletano, P. Unimib shar: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences* **2017**, *7*, 1101. <https://doi.org/10.3390/app7101101>.

45. Zhang, M.; Sawchuk, A.A. USC-HAD: A Daily Activity Dataset for Ubiquitous Activity Recognition Using Wearable Sensors. In Proceedings of the ACM International Conference on Ubiquitous Computing (Ubicomp) Workshop on Situation, Activity and Goal Awareness (SAGAware), Pittsburgh, Pennsylvania, USA, September 2012. <https://doi.org/10.1145/2370216.2370438>.

46. Vavoulas, G.; Chatzaki, C.; Malliotakis, T.; Pediaditis, M.; Tsiknakis, M. The mobiact dataset: Recognition of activities of daily living using smartphones. In Proceedings of the International conference on information and communication technologies for ageing well and e-health. SciTePress, 2016, Vol. 2, pp. 143–151. <https://doi.org/10.5220/0005792401430151>.

47. Malekzadeh, M.; Clegg, R.G.; Cavallaro, A.; Haddadi, H. Mobile Sensor Data Anonymization. In Proceedings of the Proceedings of the International Conference on Internet of Things Design and Implementation, New York, NY, USA, 2019; IoTDI '19, pp. 49–58. <https://doi.org/10.1145/3302505.3310068>.

48. Patiño-Saucedo, J.A.; Ariza-Colpas, P.P.; Butt-Aziz, S.; Piñeres-Melo, M.A.; López-Ruiz, J.L.; Morales-Ortega, R.C.; De-la-hoz Franco, E. Predictive Model for Human Activity Recognition Based on Machine Learning and Feature Selection Techniques. *International Journal of Environmental Research and Public Health* **2022**, *19*. <https://doi.org/10.3390/ijerph191912272>.

49. Zappi, P.; Roggen, D.; Farella, E.; Tröster, G.; Benini, L. Network-Level Power-Performance Trade-Off in Wearable Activity Recognition: A Dynamic Sensor Selection Approach. *ACM Trans. Embed. Comput. Syst.* **2012**, *11*. <https://doi.org/10.1145/2345770.2345781>.

50. Yang, Z.; Zhang, Y.; Zhang, G.; Zheng, Y. Widar 3.0: WiFi-based activity recognition dataset. *IEEE Dataport* **2020**. <https://doi.org/10.21227/7znf-qp86>.

51. Reyes-Ortiz, J.L.; Oneto, L.; Samà, A.; Parra, X.; Anguita, D. Transition-aware human activity recognition using smartphones. *Neurocomputing* **2016**, *171*, 754–767. <https://doi.org/10.1016/j.neucom.2015.07.085>.

52. Ignatov, A. Real-time human activity recognition from accelerometer data using convolutional neural networks. *Applied Soft Computing* **2018**, *62*, 915–922. <https://doi.org/10.1016/j.asoc.2017.09.027>.

53. Jain, A.; Kanhangad, V. Human activity classification in smartphones using accelerometer and gyroscope sensors. *IEEE Sensors Journal* **2017**, *18*, 1169–1177. <https://doi.org/10.1109/JSEN.2017.2782492>.

54. Chen, K.; Yao, L.; Zhang, D.; Wang, X.; Chang, X.; Nie, F. A semisupervised recurrent convolutional attention model for human activity recognition. *IEEE transactions on neural networks and learning systems* **2019**, *31*, 1747–1756. <https://doi.org/10.1109/TNNLS.2019.2927224>.

55. Alawneh, L.; Mohsen, B.; Al-Zinati, M.; Shatnawi, A.; Al-Ayyoub, M. A comparison of unidirectional and bidirectional LSTM networks for human activity recognition. In Proceedings of the 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). IEEE, 2020, pp. 1–6. <https://doi.org/10.1109/PerComWorkshops48775.2020.9156264>.

56. Lin, Y.; Wu, J. A novel multichannel dilated convolution neural network for human activity recognition. *Mathematical Problems in Engineering* **2020**. <https://doi.org/10.1155/2020/5426532>.

57. Zhang, J.; Wu, F.; Wei, B.; Zhang, Q.; Huang, H.; Shah, S.W.; Cheng, J. Data augmentation and dense-LSTM for human activity recognition using Wi-Fi signal. *IEEE Internet of Things Journal* **2020**, *8*, 4628–4641. <https://doi.org/10.1109/JIOT.2020.3026732>.

58. Nadeem, A.; Jalal, A.; Kim, K. Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy Markov model. *Multimedia Tools and Applications* **2021**, *80*, 21465–21498. <https://doi.org/10.1007/s11042-021-10687-5>.

59. Kavuncuoğlu, E.; Uzunhisarcıklı, E.; Barshan, B.; Özdemir, A.T. Investigating the performance of wearable motion sensors on recognizing falls and daily activities via machine learning. *Digital Signal Processing* **2022**, *126*, 103365. <https://doi.org/10.1016/j.dsp.2021.103365>.

60. Lu, L.; Zhang, C.; Cao, K.; Deng, T.; Yang, Q. A multichannel CNN-GRU model for human activity recognition. *IEEE Access* **2022**, *10*, 66797–66810. <https://doi.org/doi=10.1109/ACCESS.2022.3185112>.

61. Kim, Y.W.; Cho, W.H.; Kim, K.S.; Lee, S. Oversampling technique-based data augmentation and 1D-CNN and bidirectional GRU ensemble model for human activity recognition. *Journal of Mechanics in Medicine and Biology* **2022**, *22*, 2240048. <https://doi.org/10.1142/S0219519422400486>.

62. Sarkar, A.; Hossain, S.S.; Sarkar, R. Human activity recognition from sensor data using spatial attention-aided CNN with genetic algorithm. *Neural Computing and Applications* **2023**, *35*, 5165–5191. <https://doi.org/10.1007/s00521-022-07911-0>.

63. Semwal, V.B.; Jain, R.; Maheshwari, P.; Khatwani, S. Gait reference trajectory generation at different walking speeds using LSTM and CNN. *Multimedia Tools and Applications* **2023**, *82*, 33401–33419. <https://doi.org/10.1007/s11042-023-14733-2>.

64. Yao, M.; Zhang, L.; Cheng, D.; Qin, L.; Liu, X.; Fu, Z.; Wu, H.; Song, A. Revisiting Large-Kernel CNN Design via Structural Re-Parameterization for Sensor-Based Human Activity Recognition. *IEEE Sensors Journal* **2024**. <https://doi.org/10.1109/JSEN.2024.3371462>.

65. Wei, X.; Wang, Z. TCN-Attention-HAR: Human activity recognition based on attention mechanism time convolutional network. *Scientific Reports* **2024**, *14*, 7414. <https://doi.org/10.1038/s41598-024-57912-3>.

66. El-Adawi, E.; Essa, E.; Handosa, M.; Elmougy, S. Wireless body area sensor networks based human activity recognition using deep learning. *Scientific Reports* **2024**, *14*, 2702. <https://doi.org/10.1038/s41598-024-53069-1>.

67. Ye, X.; Wang, K.I.K. Deep Generative Domain Adaptation with Temporal Relation Knowledge for Cross-User Activity Recognition. *arXiv preprint arXiv:2403.14682* **2024**. <https://doi.org/10.48550/arXiv.2403.14682>.

68. Kaya, Y.; Topuz, E.K. Human activity recognition from multiple sensors data using deep CNNs. *Multimedia Tools and Applications* **2024**, *83*, 10815–10838. <https://doi.org/10.1007/s11042-023-15830-y>.

69. Zhang, H.; Xu, L. Multi-STMT: multi-level network for human activity recognition based on wearable sensors. *IEEE Transactions on Instrumentation and Measurement* **2024**. <https://doi.org/10.1109/TIM.2024.3365155>.

70. Zhang, L.; Yu, J.; Gao, Z.; Ni, Q. A multi-channel hybrid deep learning framework for multi-sensor fusion enabled human activity recognition. *Alexandria Engineering Journal* **2024**, *91*, 472–485. <https://doi.org/10.1016/j.aej.2024.01.030>.

71. Saha, U.; Saha, S.; Kabir, M.T.; Fattah, S.A.; Saquib, M. Decoding human activities: Analyzing wearable accelerometer and gyroscope data for activity recognition. *IEEE Sensors Letters* **2024**. <https://doi.org/10.1109/LSENS.2024.3423340>.

72. Qi, W.; Su, H.; Yang, C.; Ferrigno, G.; De Momi, E.; Aliverti, A. A fast and robust deep convolutional neural networks for complex human activity recognition using smartphone. *Sensors (Switzerland)* **2019**, *19*, 1–15. <https://doi.org/10.3390/s19173731>.

73. Plizzari, C.; Cannici, M.; Matteucci, M. Skeleton-based Action Recognition via Spatial and Temporal Transformer Networks. *arXiv preprint 2020*. <https://doi.org/10.1016/j.cviu.2021.103219>.

74. Jin, S.; Xu, L.; Xu, J.; Wang, C.; Liu, W.; Qian, C.; Ouyang, W.; Luo, P. Whole-Body Human Pose Estimation in the Wild, 2020, [<http://arxiv.org/abs/2007.11858>]. <https://doi.org/10.48550/arXiv.2007.11858>.

75. Qin, Z.; Liu, Y.; Ji, P.; Kim, D.; Wang, L.; McKay, B.; Anwar, S.; Gedeon, T. Fusing Higher-order Features in Graph Neural Networks for Skeleton-based Action Recognition. *arXiv preprint 2021*. <https://doi.org/10.1109/TNNLS.2022.3201518>.

76. Mitra, S.; Kanungoe, P. Smartphone based Human Activity Recognition using CNNs and Autoencoder Features. In Proceedings of the 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), 2023, pp. 811–819. <https://doi.org/10.1109/ICOEI56765.2023.10126051>.

77. Badawi, A.A.; Al-Kabbany, A.; Shaban, H. Multimodal Human Activity Recognition From Wearable Inertial Sensors Using Machine Learning. In Proceedings of the 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2018, pp. 402–407. <https://doi.org/10.1109/IECBES.2018.8626737>.

78. Pan, D.; Liu, H.; Qu, D. Heterogeneous Sensor Data Fusion for Human Falling Detection. *IEEE Access* **2021**, *9*, 17610–17619. <https://doi.org/10.1109/ACCESS.2021.3051899>.

79. Kim, C.; Lee, W. Human Activity Recognition by the Image Type Encoding Method of 3-Axial Sensor Data. *Applied Sciences* **2023**, *13*. <https://doi.org/10.3390/app13084961>.

80. Vijeta Sharma, Manjari Gupta, A.K.P.D.M.; Kumar, A. A Review of Deep Learning-based Human Activity Recognition on Benchmark Video Datasets. *Applied Artificial Intelligence* **2022**, *36*, 2093705, [<https://doi.org/10.1080/08839514.2022.2093705>]. <https://doi.org/10.1080/08839514.2022.2093705>.

81. Peng, W.; Hong, X.; Chen, H.; Zhao, G. Learning graph convolutional network for skeleton-based human action recognition by neural searching. In Proceedings of the AAAI 2020 - 34th AAAI Conference on Artificial Intelligence, 2020, pp. 2669–2676. <https://doi.org/10.1609/aaai.v34i03.5652>.

82. Albaigar, M. A Survey on Deep Learning and Its Impact on Agriculture: Challenges and Opportunities. *Agriculture (Switzerland)* **2023**, *13*. <https://doi.org/https://doi.org/10.3390/agriculture13030540>.

83. Bai, Y.; Tao, Z.; Wang, L.; Li, S.; Yin, Y.; Fu, Y. Collaborative Attention Mechanism for Multi-View Action Recognition, 2020, [<http://arxiv.org/abs/2009.06599>]. <https://doi.org/10.48550/arXiv.2009.06599>.

84. Madokoro, H.; Nix, S.; Woo, H.; Sato, K. A mini-survey and feasibility study of deep-learning-based human activity recognition from slight feature signals obtained using privacy-aware environmental sensors. *Applied Sciences (Switzerland)* **2021**, *11*, 1–31. <https://doi.org/10.3390/app112411807>.

85. Garcia-Ceja, E.; Galván-Tejada, C.E.; Brena, R. Multi-view stacking for activity recognition with sound and accelerometer data. *Information Fusion* **2018**, *40*, 45–56. <https://doi.org/https://doi.org/10.1016/j.inffus.2017.06.004>.

86. Ramanujam, E.; Perumal, T.; Padmavathi, S. Human Activity Recognition with Smartphone and Wearable Sensors Using Deep Learning Techniques: A Review. *IEEE Sensors Journal* **2021**, *21*, 1309–13040. <https://doi.org/10.1109/JSEN.2021.3069927>.

87. Dua, N.; Singh, S.N.; Challa, S.K.; Semwal, V.B.; Kumar, M.L.S. A Survey on Human Activity Recognition Using Deep Learning Techniques and Wearable Sensor Data. In Proceedings of the Communications in

Computer and Information Science, 2022, Vol. 1762, pp. 52–71. https://doi.org/https://doi.org/10.1007/978-3-031-24352-3_5.

- 88. Zhang, Z.; Wang, W.; An, A.; Qin, Y.; Yang, F. A human activity recognition method using wearable sensors based on convtransformer model. *Evolving Systems* **2023**, *14*, 939–955. <https://doi.org/10.1007/s12530-022-09480-y>.
- 89. Ordóñez, F.J.; Roggen, D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors (Switzerland)* **2016**, *16*, 115. <https://doi.org/10.3390/s16010115>.
- 90. Murad, A.; Pyun, J.Y. Deep recurrent neural networks for human activity recognition. *Sensors* **2017**, *17*, 2556. <https://doi.org/10.3390/s17112556>.
- 91. Islam, M.M.; Nooruddin, S.; Karray, F. Multimodal Human Activity Recognition for Smart Healthcare Applications. In Proceedings of the Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics, 2022, Vol. 2022-Octob, pp. 196–203. <https://doi.org/10.1109/SMC53654.2022.9945513>.
- 92. Gupta, S. Deep learning based human activity recognition (HAR) using wearable sensor data. *International Journal of Information Management Data Insights* **2021**, *1*, 100046.
- 93. Zhang, L.; Yu, J.; Gao, Z.; Ni, Q. A multi-channel hybrid deep learning framework for multi-sensor fusion enabled human activity recognition. *Alexandria Engineering Journal* **2024**, *91*, 472–485. <https://doi.org/10.1016/j.aej.2024.01.030>.
- 94. Shahabian Alashti, M.R.; Bamorovat Abadi, M.; Holthaus, P.; Menon, C.; Amirabdollahian, F. Lightweight human activity recognition for ambient assisted living. In Proceedings of the ACHI 2023: The Sixteenth International Conference on Advances in Computer-Human Interactions. IARIA, April 2023.
- 95. Chen, Z.; Wu, M.; Cui, W.; Liu, C.; Li, X. An attention based CNN-LSTM approach for sleep-wake detection with heterogeneous sensors. *IEEE Journal of Biomedical and Health Informatics* **2020**, *25*, 3270–3277. <https://doi.org/10.1109/JBHI.2020.3006145>.
- 96. Essa, E.; Abdelmaksoud, I.R. Temporal-channel convolution with self-attention network for human activity recognition using wearable sensors. *Knowledge-Based Systems* **2023**, *278*, 110867. <https://doi.org/10.1016/j.knosys.2023.110867>.
- 97. Kim, H.; Lee, D. CLAN: A Contrastive Learning based Novelty Detection Framework for Human Activity Recognition. *arXiv preprint arXiv:2401.10288* **2024**. <https://doi.org/0.48550/arXiv.2401.10288>.
- 98. Wang, Y.; Liu, Z. Human Activity Recognition Using Wearable Sensors and Deep Learning: A Review. *IEEE Xplore* **2023**.
- 99. Madsen, H. *Time series analysis*; Chapman and Hall/CRC, 2007. <https://doi.org/10.1201/9781420059687>.
- 100. Gori, L.R.; Tapaswi, M.; Liao, R.; Jia, J.; Urtasun, R.; Fidler, S. Situation recognition with graph neural networks. In Proceedings of the Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4173–4182. <https://doi.org/10.48550/arXiv.1708.04320>.
- 101. Li, F.; Shirahama, K.; Nisar, M.A.; Köping, L.; Grzegorzek, M. Comparison of feature learning methods for human activity recognition using wearable sensors. *Sensors (Switzerland)* **2018**, *18*. <https://doi.org/https://doi.org/10.3390/s18020679>.
- 102. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* **2016**. <https://doi.org/10.48550/arXiv.1609.02907>.
- 103. Yan, S.; Xiong, Y.; Lin, D. Spatial temporal graph convolutional networks for skeleton-based action recognition. In Proceedings of the Proceedings of the AAAI conference on artificial intelligence, 2018, Vol. 32. <https://doi.org/10.1609/aaai.v32i1.12328>.
- 104. Shi, L.; Zhang, Y.; Cheng, J.; Lu, H. Two-stream adaptive graph convolutional networks for skeleton-based action recognition. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 12026–12035. <https://doi.org/10.48550/arXiv.1805.07694>.
- 105. Shiraki, K.; Hirakawa, T.; Yamashita, T.; Fujiyoshi, H. Spatial temporal attention graph convolutional networks with mechanics-stream for skeleton-based action recognition. In Proceedings of the Proceedings of the Asian Conference on Computer Vision, 2020. https://doi.org/10.1007/978-3-03-69541-5_21.
- 106. Shi, L.; Zhang, Y.; Cheng, J.; Lu, H. Skeleton-based action recognition with multi-stream adaptive graph convolutional networks. *IEEE Transactions on Image Processing* **2020**, *29*, 9532–9545. <https://doi.org/10.1109/TIP.2020.3028207>.

107. Huang, J.; Xiang, X.; Gong, X.; Zhang, B.; et al. Long-short graph memory network for skeleton-based action recognition. In Proceedings of the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 645–652. <https://doi.org/10.1109/WACV45572.2020.9093598>.
108. Thakkar, K.; Narayanan, P. Part-based graph convolutional network for action recognition. *arXiv preprint arXiv:1809.04983* **2018**. <https://doi.org/10.48550/arXiv.1809.04983>.
109. Howard, A.G. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* **2017**.
110. Iandola, F.N. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size. *arXiv preprint arXiv:1602.07360* **2016**.
111. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International conference on machine learning. PMLR, 2019, pp. 6105–6114.
112. Han, S.; Pool, J.; Tran, J.; Dally, W. Learning both weights and connections for efficient neural network. *Advances in neural information processing systems* **2015**, 28.
113. Awais, M.; Raza, M.; Ali, K.; Ali, Z.; Irfan, M.; Chughtai, O.; Khan, I.; Kim, S.; Ur Rehman, M. An internet of things based bed-egress alerting paradigm using wearable sensors in elderly care environment. *Sensors* **2019**, 19, 2498. <https://doi.org/10.3390/s19112498>.
114. Chen, C.; Jafari, R.; Kehtarnavaz, N. A real-time human action recognition system using depth and inertial sensor fusion. *IEEE Sensors Journal* **2015**, 16, 773–781. <https://doi.org/10.1109/JSEN.2015.2487358>.
115. Hinton, G. Distilling the Knowledge in a Neural Network. *arXiv preprint arXiv:1503.02531* **2015**. <https://doi.org/10.48550/arXiv.1503.02531>.
116. Polino, A.; Pascanu, R.; Alistarh, D. Model compression via distillation and quantization. *arXiv preprint arXiv:1802.05668* **2018**. <https://doi.org/10.48550/arXiv.1802.05668>.
117. Yan, S.; Xiong, Y.; Lin, D. Spatial temporal graph convolutional networks for skeleton-based action recognition. In Proceedings of the Proceedings of the AAAI conference on artificial intelligence, 2018, Vol. 32. <https://doi.org/10.1609/aaai.v32i1.12328>.
118. Xue, H.; Jiang, W.; Miao, C.; Ma, F.; Wang, S.; Yuan, Y.; Yao, S.; Zhang, A.; Su, L. DeepMV: Multi-view deep learning for device-free human activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **2020**, 4, 1–26. <https://doi.org/10.1145/3380980>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.