algorithms

MDPI

Review

A Reviewof Recent Techniques for Human Activity Recognition: Multimodality, Reinforcement Learning, and Language Models

Ugonna Oleh *, Roman Obermaisser * and Abu Shad Ahammed

Chair of Embedded System, University of Siegen, 57076 Siegen, Germany; abu.ahammed@uni-siegen.de
* Correspondence: ugonna.oleh@uni-siegen.de (U.O.); roman.obermaisser@uni-siegen.de (R.O.); Tel.: +49-271-740-3350

Abstract: Human Activity Recognition (HAR) is a rapidly evolving field with the potential to revolutionise how we monitor and understand human behaviour. This survey paper provides a comprehensive overview of the state-of-the-art in HAR, specifically focusing on recent techniques such as multimodal techniques, Deep Reinforcement Learning and large language models. It explores the diverse range of human activities and the sensor technologies employed for data collection. It then reviews novel algorithms used for Human Activity Recognition with emphasis on multimodality, Deep Reinforcement Learning and large language models. It gives an overview of multimodal datasets with physiological data. It also delves into the applications of HAR in healthcare. Additionally, the survey discusses the challenges and future directions in this exciting field, highlighting the need for continued research and development to fully realise the potential of HAR in various real-world applications.

Keywords: HAR; multimodal; LLMs; deep reinforcement learning

Citation: Oleh, U.; Obermaisser, R.; Ahammed, A.S. A Review of Recent Techniques for Human Activity Recognition: Multimodality, Reinforcement Learning, and Language Models. *Algorithms* **2024**, *17*, 434. https://doi.org/10.3390/a17100434

Academic Editor: Ulrich Kerzel

Received: 1 September 2024
Revised: 25 September 2024
Accepted: 26 September 2024
Published: 28 September 2024

1. Introduction

Human Activity Recognition (HAR) involves using different techniques to determine what a person is doing from observations of sensory information of the person's body and conditions of the surrounding environment [1] and also inferring the person's goal and mental status [2]. It is the automatic detection and recognition of human activities using data collected from various sensors. These activities can range from simple actions like standing or sitting to more complex activities like eating, cooking, or commuting to work [3].

HAR is a fast-evolving field that can change how we monitor and understand human behaviour. It is used in multiple domains such as sports, gaming, health monitoring, robotics, human–computer interaction, security, surveillance, and entertainment [4]. The monitoring of vital health parameters like blood pressure, respiration rate, electrocardiograph (ECG) patterns, and pulse rate is crucial to personal health systems and telemedicine techniques [5]. These data are important in the application of HAR in health-based use cases.

This survey aims to provide an overview of the state-of-the-art in HAR, focusing on new techniques such as multimodality, Deep Reinforcement Learning (DRL) and the use of large language models (LLMs) in HAR. It explores the wide range of human activities and the various sensor technologies used for data collection. It examines the various recent algorithms for activity recognition and available datasets with physiological data. Also, the applications of HAR in healthcare are discussed as well as the challenges and future directions.

This survey paper uses a systematic literature review methodology to identify and analyse recent and relevant research on HAR with a focus on multimodality, DRL, and LLMs. For this survey, three primary data sources were used: IEEE Xplore, SCOPUS, and

Google Scholar. The search was limited to articles published in English between 2019 and 2024, which captured novel and recent advancements in the topic area. The decision to limit the survey to publications between 2019 and 2024 was to capture the most recent algorithms in HAR, especially ones that have not been captured by already existing surveys. Publications not accessible either due to the website being unavailable or being part of paid content were excluded from this study. Redundant publications (for instance conference proceedings already extended in a journal article) were also excluded.

Publications with novel algorithms were selected for this survey with emphasis on multimodality, DRL, and LLMs. Multimodal HAR algorithms are important for capturing the complex nature of human activities, and there are various research works that have attempted to harness its potential. Traditionally, HAR models use supervised learning and require large and annotated datasets which are computationally expensive to acquire. While some of the sensors can acquire a ton of data, researchers are limited to only the data they can afford to label. With DRL and LLM, such large data can be labelled with less computational cost and even unlabelled and semi-labelled datasets can be used in building HAR models. This is important because HAR models are limited to their training datasets and the more robust their dataset, the more robust the model would be. For future research in the area of HAR, these three methodologies will aid in the development of robust, user-centred, and more accurate HAR systems which can be personalised especially for use in healthcare applications. The publications reviewed are grouped firstly based on the type of sensors used and then based on the algorithm used.

1.1. Review of Related Works

There have been survey papers on HAR with a focus on different areas like Wearable Sensor-based HAR [6], HAR in ambient smart living environments [2,7] and deep learning-based HAR [4,8,9]. This section analyses some of those survey papers and highlights the gap this survey paper aims to bridge.

Wang et al. [6] discussed sensor modalities in HAR, classifying them as Wearable Sensor-based HAR (WSHAR), Ambient Sensor-based HAR (ASHAR) and Hybrid Sensor-based HAR (HSHAR). The survey focused on wearable sensor modality-centred HAR in healthcare and discussed the various steps involved in WSHAR, including sensor selection and placement, data collection and preprocessing, feature extraction and selection, and classification algorithms. Manoj and Thyagaraju [7] reviewed the application of deep learning approaches to HAR and vital health sign monitoring in ambient assisted living environments. They highlighted the importance of ambient sensors in recognising complex human activities, especially in ambient assisted living environments. The paper also discussed different health vitals like respiration rate, body temperature, heart rate, etc., and the use of wearable and ambient sensors to monitor them.

Nweke et al. [10] provided an analysis of data fusion techniques and classifier systems for HAR with a focus on wearable and mobile devices. The authors highlighted the importance of integrating data from multiple sensors to reduce uncertainty and enhance the accuracy of HAR systems. They discussed different data fusion techniques and highlighted the advantages of deep learning in automatic feature extraction. Hussain et al. [3] presented an overview of HAR research conducted between 2010 and 2018 with a focus on device-free solutions. In the survey paper, a new taxonomy is proposed to classify HAR into three main categories: action based, motion based, and interaction based [3]. The authors discussed the latest research in each category and highlighted key attributes and design approaches. They also highlighted some applications of HAR in various areas and discussed future research issues and challenges such as recognising complex activities, addressing environmental interference, security, and privacy issues, and so on.

Chen et al. [8] conducted a comprehensive overview of deep learning techniques in sensor-based HAR and proposed a new taxonomy of deep learning methods given the challenges of HAR. The paper discussed various deep learning architectures, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), and their

applications in addressing challenges like feature extraction, annotation scarcity, and class imbalance. They emphasised the growing importance of HAR due to the increasing prevalence of smart devices and the Internet of Things (IoT) [8]. The paper also outlined the typical HAR process and identified challenges in HAR and compared different solutions for these challenges. The authors also suggested some directions for future research such as the use of deep unsupervised transfer learning and the need for a standardisation of the state of the art.

Li et al. [9] reviewed deep learning-based human behaviour recognition methods. Their survey was mainly on human behaviour recognition based on two-stream, 3D convolutional and hybrid networks. They highlighted the importance of the availability of large-scale annotated datasets and reviewed some of the available datasets. They also pointed out some issues with the available datasets such as occlusion, which is when other people or objects obstruct the person being recognised, and the large amount of time and resources needed to properly label a large dataset.

Diraco et al. [2] provided an overview of current HAR research in smart living and also provided a road-map for future advances in HAR. The authors delved into the multifaceted domain of HAR within the context of smart living environment. The paper identified five key domains important for the successful deployment of HAR in smart living: sensing technology, real-time processing, multimodality, resource-constrained processing, and interoperability [2]. It emphasised the importance of HAR in enabling a seamless integration of technology into our daily lives and enhancing our overall quality of life.

Nikpour et al. [4] discussed the advantages of Deep Reinforcement Learning (DRL) in HAR, such as its ability to adapt to different data modalities and learn without explicit supervision. The authors explained that DRL, a machine learning technique where agents learn through interaction with an environment, can improve the accuracy and efficiency of HAR systems by identifying the most informative parts of sensor data, such as video frames or body joint locations. The authors also acknowledged the challenges of DRL, like the need for large amounts of interaction data and the high computational cost. They also suggested future research directions such as multiagent reinforcement learning, unsupervised and few-shot activity recognition and computational cost optimisation [4].

While some survey papers point out the potential of HAR in healthcare [2,6] and even the use of sensors to monitor health vitals [7], they do not discuss the importance of health vitals datasets nor the combination of other sensor data and health vital data in HAR. With the recent advances in technology, it is important to stay up to date, and some of the available surveys do not cover these recent research areas [3,6,7]. More recent survey papers are focused on specific areas like the application of HAR in smart homes [2] or the use of deep learning models [4,8,9]. This survey paper aims to bridge that gap.

1.2. Contributions of This Review Paper

This survey paper aims to close these research gaps by focusing on more recent papers and more recent technologies like the use of multimodal techniques, DRL, and LLMs, and exploring the integration of physiological data in HAR and its application in healthcare. It also specifically reviews the available datasets with physiological and inertial data that can be used for future work in this field.

The main contributions of this review paper are described below:

1. Analysis of human activities and sensors used for the recognition of human activities.
2. Review of recent vision and non-vision HAR publications.
3. Review of recent HAR algorithm with focus on multimodal techniques, DRL, and LLMs.
4. Review of multimodal datasets with physiological data.
5. Discussion on the applications of HAR in healthcare.
6. Discussion on the challenges and future directions for HAR.

Table 1 shows a comparison of our review paper and other review papers. The comparison is based on six factors: the range of years that the papers reviewed cover, the datasets

reviewed (if any), the review of recent methodologies (DRL, multimodality, and LLMs) in HAR, and whether HAR applications in healthcare were analysed.

Our review paper provides a very recent review of new methodologies (DRL, multimodality, and LLMs) in HAR. It highlights the role these new methodologies can play in advancing the task of HAR. In the detection of complex human activities, multimodal HAR models are shown to demonstrate improved performance over the single-modal ones, and DRL can be used to improve the result of multimodal fusion. LLMs can be used in the annotation of large datasets and in developing of personalised HAR systems and models that better understand and predict human behaviour using contextual information and user preferences. DRL is also shown to improve feature extraction and human–robot interactions. It can also be used to adjust the movement of robots in homes for better HAR. In addition, our paper provides a review of multimodal datasets with physiological data which can be further utilised in the application of multimodal HAR in healthcare. We also look at applications of HAR in healthcare and directions for future research.

Table 1. Comparison with other review papers.

Paper	Year of Most Recent Reviewed Paper	Dataset	DRL	Multimodality	LLMs	Applications in Healthcare
Wang et al. [6]	2019	Reviewed datasets	No	No	No	Yes
Manoj and Thyagaraja [7]	2019	No review of datasets	No	No	No	Yes
Nweke et al. [10]	2018	No review of datasets	No	Yes	No	No
Hussain et al. [3]	2010–2019	A brief review of datasets	No	No	No	No
Chen et al. [8]	2020	Reviewed datasets	No	Yes	No	No
Li et al. [9]	2021	Reviewed vision-based HAR Datasets	No	Vision-based multimodal-ity reviewed	No	Yes
Diraco et al. [2]	2023	Reviewed datasets	No	Yes	No	No
Nikpour et al. [4]	2023	Reviewed datasets	Focused on DRL in HAR	No	No	No
Our Review Paper	2019–2024	Reviewed multimodal Datasets in detail	Yes	Yes	Yes	Yes

1.3. Organisation of This Review Paper

This review paper has ten sections. The first section gives an introduction to the review paper and includes a review of related papers and the contributions of our review paper. Section 2 provides a structured representation of human activities and sensors used in HAR. In Section 3, a review of recent research on vision-based HAR is carried out highlighting the most recent techniques like the skeleton data-based HAR. Section 4 contains a review of non-vision-based HAR and covers recent research for wearable sensors and ambient sensors. Section 5 covers a review of multimodal, DRL, and LLM techniques in HAR. Section 5.1 highlights the different ways multimodality has been proposed in HAR, including fusion techniques. Multimodal datasets with physiological data are reviewed comprehensively in Section 6. Section 7 highlights the applications of HAR in healthcare.

Section 8 discusses the challenges faced by HAR and future research directions. Finally, Section 9 concludes the review paper.

2. Human Activities and HAR Sensors

2.1. Human Activities

This section provides a structured representation of different human activities, enabling better analysis of these activities. A Unified Modelling Language (UML) diagram is used to illustrate the relationship and attributes of these activities. Figure 1 shows the UML diagram of the different types of activities and their attributes. Human activities can be simple or complex. Complex human activities (CHA) are made up of multiple simultaneous or overlapping actions like cooking or cleaning, while simple human activities (SHA) are single repeated actions such as walking or sitting [11]. Complexity is used as an attribute in the UML diagram to show whether an activity is complex or simple. These activities can also be carried out alone or with other people. This is important to consider when discussing social activities and is represented by the attribute no_of_persons, which is an integer showing the number of people involved in that particular activity. Other attributes that human activities have include location, duration, start, and end time.

There are five basic types of human activities: physical activities, intellectual activities, social activities, recreational activities [12], and daily activities. Intellectual activities include activities like reading, writing, working at a desk, and most other activities around school or office work. Physical activities include exercises and other activities that require a lot of energy expenditure [13]. Social activities are activities that require communicating and interacting with others [14]. Recreational activities are activities undertaken for pleasure, leisure, or relaxation. Daily activities are activities carried out regularly that do not fit into the other four types such as cooking, cleaning, and eating.

Each type of activity has attributes that are relevant to them based on the task of HAR. With physical activities, health vitals such as heart rate and pulse rate are relevant. Specific activities like swimming have attributes such as speed, stroke rate [15], and body temperature. Activities like jogging, running, biking and swimming all have the attribute speed. Some activities can be classified under more than one type, like walking, which is a physical activity but can also be classified as a daily activity. For the human activity UML model, activities are classified based on the type they best fit into.

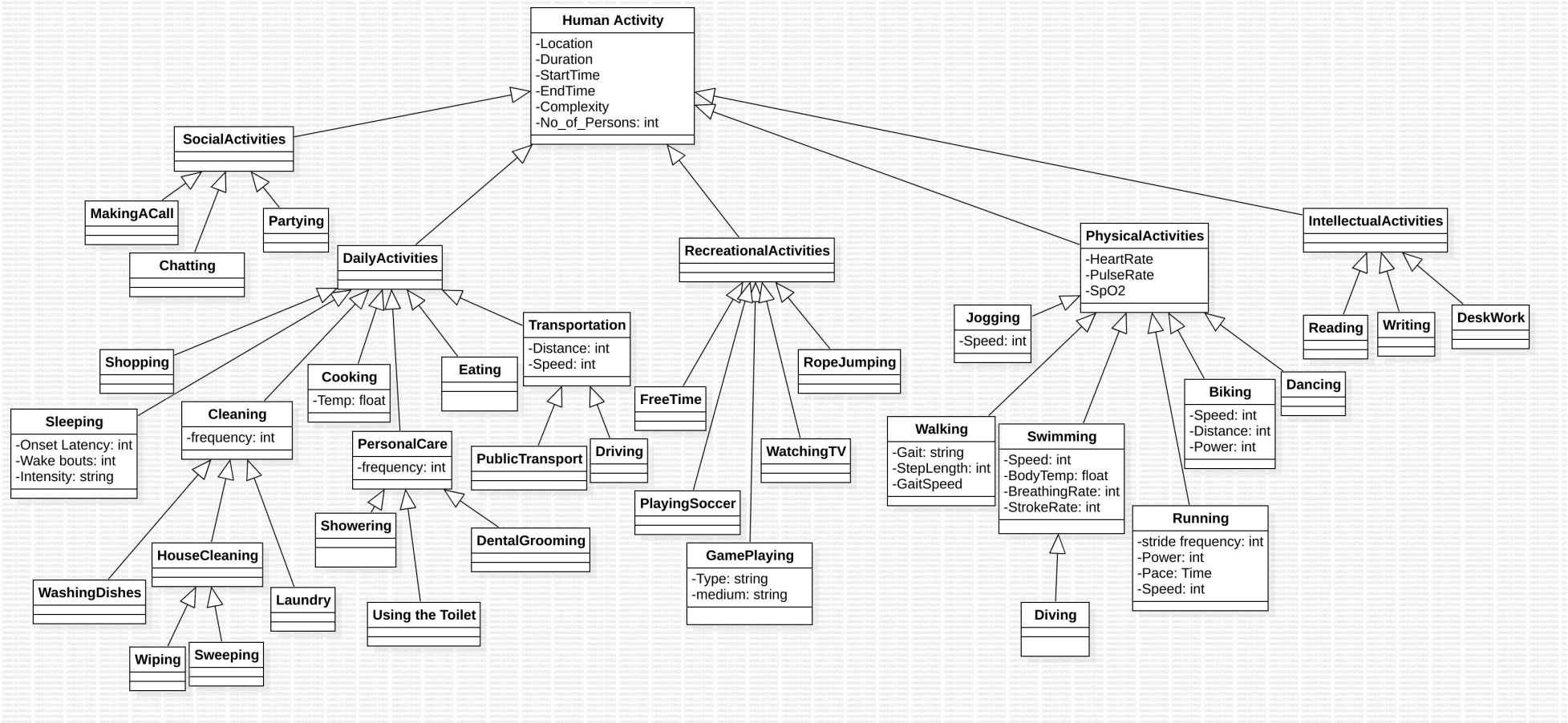

Figure 1. UML diagram of human activities.

2.2. HAR Sensors

This subsection provides a structured framework to categorise and understand the range of sensors employed in HAR. The type of sensor used in HAR is important, as it influences the kind of data collected and the activities that can be recognised. The initial focus of HAR studies was on activity recognition from images and videos, but later researchers started to explore tracking human behaviour using ambient and wearable sensors [6]. HAR sensors can be grouped into ambient, movement/inertial, and health/physiological sensors. These sensors have attributes of sensitivity, accuracy, range, resolution, and precision. Figure 2 shows the UML diagram of sensors used in the task of HAR.

Movement sensors are used to sense a person's body motion and return values for x, y, and z axis. Movement sensors include accelerometers, gyroscopes, and magnetometers [16]. Most smartphones and smartwatches contain an accelerometer and gyroscope and have been used in different HAR research [17]. These movement sensors are the basis for a lot of non-vision HAR tasks and are generally considered non-invasive because they can be found in smartwatches or smartphones, and other wearable devices.

Ambient sensors gather information about the environment like humidity, sound, temperature, etc. In HAR, ambient sensors are used in smart home environments. Sensors like pressure sensors are used to detect when the participant sits or lies down, and proximity sensors are used to determine the relative position of the occupants in a smart home environment. Vision sensors are also ambient sensors because they gather information about the environment. Location sensors such as GPS are also found in this category.

Physiological sensors measure health vitals such as heart rate, blood pressure, body temperature, etc. [16]. Photoplethysmography (PPG) sensors are used to detect blood volume through the skin and can be used to estimate blood pressure [18]. They are used in smartwatches to estimate the heart rate of the wearer [19]. Electrocardiogram (ECG) sensors measure the electrical activity of the heart and can also be found on smartwatches. Electroencephalogram (EEG) sensors are used to monitor brain signals, and there are wearable devices with EEG such as Emotiv insight, mindwave, Interaxon Muse and Neuro-Tracker [20]. Some of these wearable devices with physiological sensors are designed to be worn as headsets, caps, or over-the-ear devices, which makes them non-invasive and easy to use.

Manufacturers of smartwatches have started incorporating additional sensors such as physiological sensors in these devices [21]. Smartphones too now have a wide array of sensors. There are also other wearable devices with some of these sensors like over-the-ear wearable devices or chest straps used to measure health vitals. The availability of these non-invasive wearable devices with the necessary sensors has made the deployment of HAR applications more feasible.

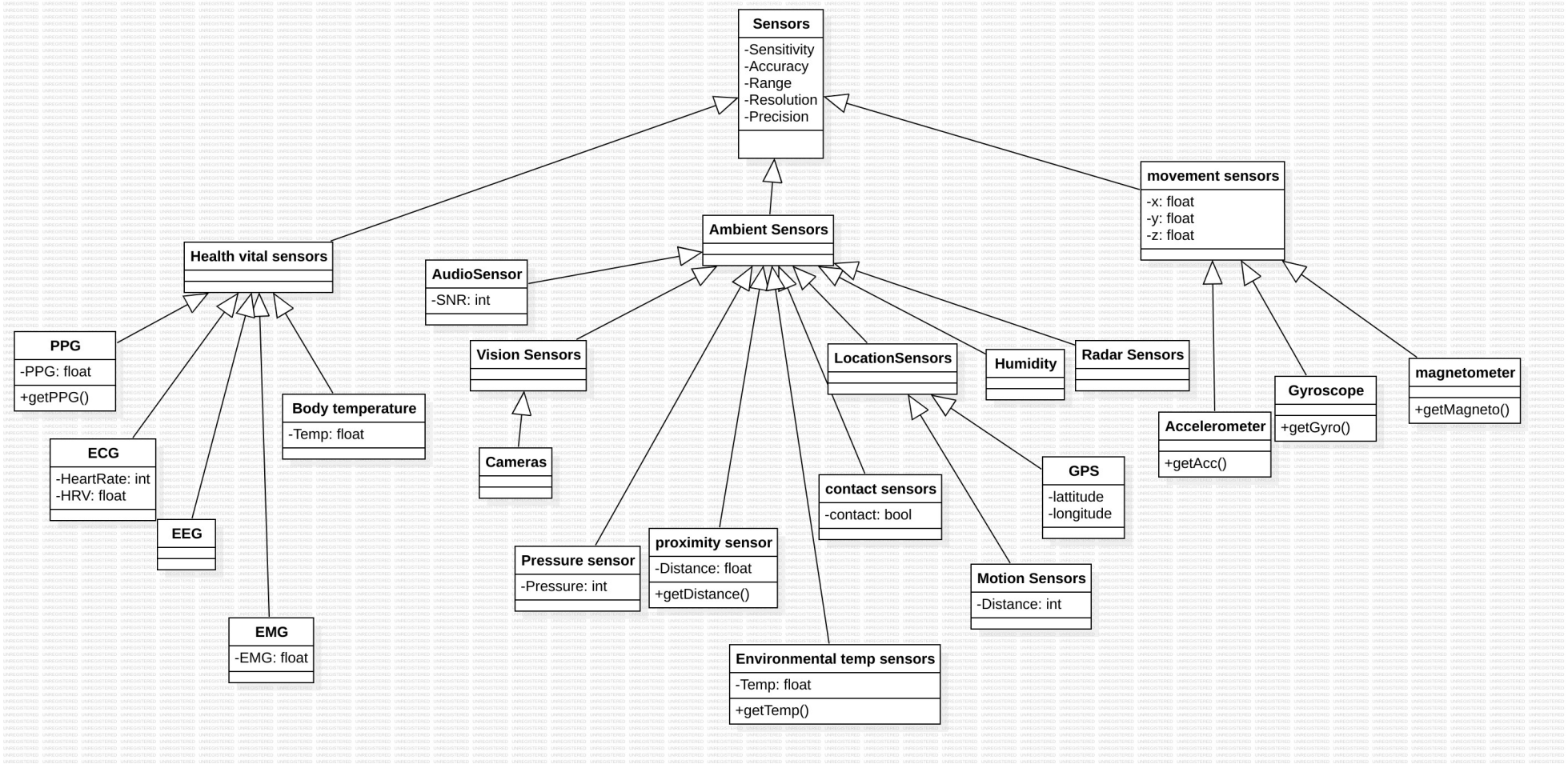

Figure 2. UML diagrams of sensors used in HAR.

3. Vision-Based HAR

Vision-based HAR is one of the earlier approaches to HAR [3]. It involves applying computer vision techniques to the task of HAR. It usually requires the use of images acquired by visual sensing technologies. It focuses on recognising appearance-based behaviour [22]. With the advancement in computer vision, research on vision-based HAR has also become popular. This section will review recent HAR papers pertaining to computer vision.

The use of the Convolutional Neural Network (CNN) is popular in computer vision problems, and this is no different in HAR. Researchers have proposed different versions of CNN models and also in combination with other models. In 2020, Basavaiah and Mohan proposed a feature extraction approach that combines Scale-Invariant Feature Transform (SIFT) and optical flow computation, and then used a CNN-based classification approach [23]. They analysed the model's performance on the Weizmann and KTH datasets, and it had an accuracy of 98.43% and 94.96% respectively.

Andrade-Ambriz et al. proposed using a Temporal Convolutional Neural Network (TCNN) which leverages spatiotemporal features for vision-based HAR using only short videos [24]. Their proposed model is based on a feature extraction module that contains a three-dimensional convolutional layer and a two-dimensional LSTM layer and then two fully connected layers, which are used for classification. They used the Kinect Activity Recognition Dataset (KARD), Cornell Activity Dataset (CARD-60), and the MSR Daily Activity 3D Dataset to evaluate the performance of their model. They used only the RGB images from these datasets. The model achieved 100% accuracy for the KARD and CAD-60 datasets and 95.6% accuracy for the MSR Daily Activity 3D dataset. They also compared their results with those of other papers and showed that their model performed better. In 2023, Parida et al. proposed the use of a hybrid CNN with LSTM model for vision-based HAR [25]. Their proposed model uses CNN for filtering features and LSTM for sequential classification, and then fully connected layers are used for feature mapping. They trained and validated their model on the UCF50 dataset, and it had an accuracy of 90.93%, which was better than the CNN and LSTM models.

Some researchers have proposed using skeleton data for HAR due to its robustness to circumstance and illumination changes [26]. Graphical Convolutional Networks (GCNs) are favoured by researchers for this task [27]. In 2019, Liu et al. proposed the structure-induced GCN [28] which carries out spectral graph convolution on a constructed inter-graph of mini-graphs of specific parts of the human skeleton. They evaluated their model on the NTU RGB+D and HDM05 datasets, and it had an accuracy of 89.05% and 85.45%, respectively. In 2020, Cheng et al. proposed a Decoupling GCN (DC-GCN), which uses a decoupling aggregation mechanism and DropGraph instead of dropout for regularisation [26]. They validated their model on the NTU-RGBD, NTU-RGBD-120, and Northwestern-UCLA datasets and carried out ablation studies to show the efficacy of the DC-GCN. The result of their experiments showed that the DC-GCN exceeds the performance of other graph models with less computational cost.

In 2022, Jiang et al. proposed a novel Inception Spatial Temporal GCN (IST-GCN), which they trained and validated on the NTU RGB+D [29]. The model uses multiscale convolution to improve the GCN based on the inception structure. The model had an accuracy of 89.9% using x-subject validation and 96.2% using x-view, which is better than other skeleton HAR models. More research has also been performed with the combination of skeleton data with other modalities of data.

Most recently in 2024, Lovanshi et al. proposed a Dynamic Multiscale Spatiotemporal Graph Recurrent Neural Network (DMST-GRNN), which uses multiscale graph convolution units (MGCUs) to represent the interconnections of the human body [30]. Their model is an encoder–decoder framework with the MGCUs as encoders and Graph-Gated Recurrent Unit Lite (GGRU-L) used as the decoder. They trained and validated their model on the Human3.6M and CMU Mocap datasets. Average mean angle errors were used to measure the accuracy of the model, and it outperformed baseline models for both datasets.

Vision-based HAR has experienced better accuracy in recent multimodal research, which uses RGB-D data rather than just RGB data [31]. RGB-D images, which provide more data such as depth information, other data such as the skeleton image, and additional context information, have also been explored by researchers. The prevalence of depth cameras like the Microsoft Kinect has also increased the availability of multimodal vision data [32]. These multimodal approaches are reviewed in Section 5.1. Vision-based HAR has also been known to face some issues such as occlusion, privacy issues and the effect of lighting conditions, which is one of the reasons for the growth in popularity of non-vision HAR approaches.

Table 2 summarises the recent literature concerning vision-based HAR. There are some commonly used performance metrics for evaluating classification algorithms. They include confusion matrix, accuracy, recall, precision, F1-score, and specificity. Most of the papers reviewed used accuracy to measure the performance of their model. Hence, for the purpose of comparison, accuracy is used to compare the different models reviewed. The accuracy of a model measures the ratio of correctly predicted outputs to the total number of outputs. The formula for accuracy is given as

$$(Accuracy = (TP + TN)/(TP + TN + FP + FN)) \quad (1)$$

where TP is True Positives, TN is True Negatives, FP is False Positives and FN is False Negatives. Ref. [24] proposed a Temporal Convolutional Neural Network and achieved 100% accuracy on the KARD dataset, while [25] proposed a hybrid CNN with LSTM model and achieved an accuracy of 90.93%. There has also been some research on HAR using skeleton data, and one of the common datasets for evaluating these models is the NTU RGB+D. It has two validation views: x-subject and x-view. In x-subject, the validation data are collected from different subjects from those of the training data, while in x-view, the training and validation data are obtained from various camera views [33]. The DC-GCN model [26] achieved an accuracy of 90.8% on the x-subject validation, while IST-GCN model [29] obtained an accuracy of 89.9% on the same. On the x-view validation of NTU RGB+D, the DC-GCN model had an accuracy of 96.6%, while IST-GCN had an accuracy of 96.2%.

In real-world environments, vision HAR faces challenges like changing light conditions, activities that look similar (like walking vs. jogging), and actions happening at different speeds. Factors such as illumination and lighting environment can affect image quality, which is the foundation of vision-based HAR [31]. HAR can be difficult when there are distractions or movement in the background. These factors can make it hard for the models to work as well as they would in controlled settings [34]. There is also the issue of privacy and the invasive nature of vision sensors. Due to these issues, non-vision sensors have become popular for HAR tasks.

Table 2. Vision-based Human Activity Recognition papers.

S/N	Paper	Dataset	Model	Accuracy	Contribution to Knowledge
1	Human activity recognition using Temporal Convolutional Neural Network architecture [24]	1. KARD 2. CARD-60 3. MSR Daily Activity 3D	TCNN	1. 100% 2. 100% 3. 95.6%	Proposed a TCNN that uses spatiotemporal features and takes only a short video (2 s) as input.
2	A novel approach for Human Activity Recognition using vision based method [25]	UCF50	hybrid CNN with LSTM	90.93%	Integrated CNN with LSTM, where CNN extracts the spatial characteristics and LSTM is used to learn the temporal information

Table 2. *Cont.*

S/N	Paper	Dataset	Model	Accuracy	Contribution to Knowledge
3	Si-GCN: Structure-induced Graph Convolution Network for skeleton-based action recognition [28]	1. NTU RGB+D 2. HDM05	SI-GCN	1. 89.05% 2. 85.45%	Constructed inter-graph of mini-graphs of specific parts of the human skeleton to show the interactions between human parts
4	Decoupling GCN with DropGraph Module for skeleton-based action recognition [26]	1. NTU-RGBD 2. NTU-RGBD-120 3. Northwestern-UCLA	DC-GCN	1. X-sub: 90.8% and x-view: 96.6%. 2. X-sub: 86.5% and x-view: 88.1%. 3. 95.3%	Proposed an Attention-guided DropGraph (ADG) to relieve the prevalent overfitting problem in GCNs.
5	Inception Spatial Temporal Graph Convolutional Networks for skeleton-based action recognition [29]	NTU RGB+D	IST-GCN	x-sub: 89.9%, x-view: 96.2%	Improved GCN and TCN based on the Inception structure using the idea of multiscale convolution to better extract spatial and temporal features.
6	3D skeleton-based human motion prediction using dynamic multiscale spatiotemporal graph recurrent neural networks [30]	1. CMU Mocap 2. Human3.6M	DMST-GRNN	Average Mean Angle Errors (MAE): 1. 0.19 for 80 ms and 1.20 for 1000 ms 2. 0.25 for 80 ms and 1.43 for 1000 ms	Proposed a multiscale approach to spatial and temporal graphs using multiscale graph convolution units (MGCUs) to describe the human body's semantic interconnection.

4. Non-Vision-Based HAR

Recent research in HAR has shifted towards the non-vision-based approach [3], and a variety of non-vision sensors such as movement sensors are now popular for use in HAR [35]. This is because movement sensor technology is more low cost and considered less invasive [36]. Ambient sensors are also used in smart homes for HAR. There are also instances of the integration of non-vision sensors to enhance vision-based HAR [37]. These instances are also covered in Section 5.1. This section will review the recent papers focusing on non-vision sensor technology for HAR.

Long Short-Term Memory (LSTM) models are popular for non-vision HAR tasks because they are designed to capture temporal dependencies in sequential data such as accelerometer and gyroscope data. Khan et al. classified two activities (walking and brisk walking) using an LSTM model [38]. They collected their data using an accelerometer, gyroscope, and magnetometer, which were then used to train and test their model. They used different sensor combinations to train and test the model. Then, they compared the performance of the different combinations and found that the combination of acceleration and angular velocity had the highest accuracy of 96.5%. The sensor input of only magnetic field gave the lowest accuracy of 59.2%.

Hernandez et al. proposed the use of a Bidirectional LSTM (BiLSTM) model for inertial-based HAR [39]. They collected data from 30 volunteers wearing a smartphone (Samsung Galaxy S II) on the waist using the embedded inertial sensors. Their proposed BiLSTM model had the best accuracy of 100% in identifying the laying-down position and 85.8% as its worst accuracy in identifying standing. They compared their model with other existing models and it performed competitively.

In 2023, Mekruksavanich and Jitpattanakul proposed a one-dimensional pyramidal residual network (1D-PyramidNet) [21] which they based on Han's deep pyramidal residual networks (DPRNs). The model was trained on the Daily Human Activity (DHA) dataset, and the results were compared with other baseline deep learning models (CNN, LSTM, BiLSTM, GRU, and BiGRU). The authors used the Bayesian optimisation approach to fine-tune the hyper-parameters of the models being compared. The 1D-PyramidNet achieved an accuracy of 96.64% and an F1-score of 95.48%, which was higher than the other baseline models.

Following the successes in using CNN models in vision-based HAR, researchers have proposed adjusting CNN models or combining them with other models to accommodate the nature of non-vision/time series data. In 2022, Deepan et al. proposed the use of a one-dimensional Convolutional Neural Network (1D-CNN) for non-vision based HAR [37]. They collected and processed the Wireless Sensor Data Mining (WISDM) dataset using a wearable sensor. They trained and validated their model on the dataset and it showed high accuracy, precision, recall, and F1-score.

In 2020, Mekruksavanich and Jitpattanakul proposed using a CNN-LSTM model which eliminates the need for the manual extraction of features [40]. They also used Bayesian optimisation to tune the hyper-parameters and trained their model on the WISDM dataset. They compared the performance of their model with that of baseline CNN and LSTM models, and it performed better with an accuracy of 96.2%.

In 2023, Choudhury and Soni proposed a 1D Convolution-based CNN-LSTM model [41]. They used their own calibrated dataset and the MotionSense and mHealth datasets to train and test the model. They also compared their model with baseline ANN and LSTM models, and their model performed better with an accuracy of 98% on their calibrated dataset. On the MotionSense dataset, it achieved 99% accuracy and 99.2% on the mHealth dataset. Abdul et al., in 2024, modified the 1D CNN-LSTM by making it a two-stream network [42], and their hybrid model achieved 99.74% accuracy on the WISDM dataset.

In 2024, El-Adawi et al. proposed a hybrid HAR system that combines Gramian Angular Field (GAF) with DenseNet [43], specifically the DenseNet169 model. The GAF algorithm turns the time series data into two-dimensional images. Then, DenseNet is used to classify these data. They used the mHealth dataset to train and test their model. Then, they compared its performance with other models and obtained an accuracy of 97.83% and F1-score of 97.83%.

Choudhury and Soni proposed a model to recognise complex activities using electromyography (EMG) sensors [44]. They proposed a lightweight CNN-LSTM model, which they trained and tested on the EMG Physical Action Dataset [45]. They compared the performance of their model against Random Forest (RF), Extreme Gradient Boosting (XGB), Artificial Neural Network (ANN), and Convolutional Neural Network-Gated Recurrent Unit (CNN-GRU) models. Their proposed lightweight CNN-LSTM performed better with the highest accuracy of 84.12% and average accuracy of 83%. They also performed an ablation study and showed the importance of integrating CNN, LSTM, and the dropout layers, as the model performed better with these layers than without.

With the increase in the popularity of smart homes, HAR within these homes becomes an important task, especially for health monitoring [46]. Ambient non-vision sensors are generally considered more acceptable due to comfortability and privacy [47]. In 2019, Natani et al. used variations of Recurrent Neural Networks (RNNs) to analyse ambient data from two smart homes with two residents each [46]. They applied Gated Recurrent Unit (GRU) and LSTM models to the Activity Recognition with Ambient Sensing (ARAS)

dataset and used a Generative Adversarial Network (GAN) to generate more data. They compared the results of the GRU and LSTM models on 10, 30 days, and 50 days of data and found that the GRU generally performed better than the LSTM models. In 2021, Diallo and Diallo compared the performance of three models (Multilayer Perceptron (MLP), RNN and LSTM) on HAR task in an ambient home environment [48]. They evaluated their models on the ARAS dataset and found that they performed well on more frequent activities but were not efficient on rare activities. Also, their MLP model had the best performance on the dataset.

One of the problems faced by HAR in smart home environments is the labelling of the data. To solve this problem, Niu et al. propose the use of multisource transfer learning to transfer a HAR model from labelled source homes to an unlabelled target home [47]. They first generated transferable representations (TRs) of the sensors in the labelled homes. Then, an LSTM HAR model was built based on the TR and then deployed to the unlabelled home. They conducted experiments on four homes in the CASAS dataset (HH101, HH103, HH105, and HH109) to validate their proposed method. Their experiments showed that their method outperformed HAR models which are based on only common sensors or single-source homes.

HAR becomes even more complicated when there is more than one resident in the house and the activities of residents overlap. To address this issue, Jethanandani et al. proposed the classifier chain method of multilabel classification [49], which considers underlying label dependencies. They used K-Nearest Neighbour as the base classifier and evaluated the proposed method on the ARAS dataset using 10-fold cross-validation. Their method was able to recognise the 27 activities in the dataset and the resident performing it.

Table 3 shows a summary of the papers reviewed in this section. On the WISDM dataset, a movement sensor dataset, ref. [42] in 2024 achieved the highest model accuracy of 99.74% using a multi-input CNN-LSTM model showing the efficiency of multimodality. More multimodal methods are discussed in Section 5.1. For ambient sensors and the smart home environment, using the ARAS dataset, ref. [48] achieved an accuracy of between 0.91 and 0.92 using MLP, RNN, and LSTM models for both houses in the dataset, while ref. [49] achieved a higher accuracy of 0.931 in house B using the classifier chain method of the Multilabel the Classification (MLC) technique with KNN as the base classifier. Some work has also been conducted with datasets that have physiological data. On the mHealth dataset, which has movement sensors and physiological sensors (ECG), ref. [43] used GAF-DenseNet and achieved 97.83% accuracy, and ref. [41] achieved an accuracy of 99.2%. In 2023, ref. [44] collected a new EMG dataset and obtained an accuracy of 83% on the HAR task using a 1D CNN-LSTM model.

Table 3. Non-vision HAR reviewed papers.

S/N	Paper	Dataset	Sensors	Model	Accuracy	Contribution to Knowledge
1	Classification of Human Motion Activities using Mobile Phone Sensors and Deep Learning Model [38]	New dataset of walking and brisk walking	Accelerometer, gyroscope and magnetometer	DNN model (LSTM)	96.5%	Investigated the best combination of sensor data (found acceleration and angular velocity to give the highest accuracy).
2	Human Activity Recognition on Smartphones Using a Bidirectional LSTM Network [39]	UCI HAR	Accelerometer and gyroscope	BiLSTM	92.67%	Used a grid search to identify the best architecture.

Table 3. *Cont.*

S/N	Paper	Dataset	Sensors	Model	Accuracy	Contribution to Knowledge
3	Efficient Recognition of Complex Human Activities Based on Smartwatch Sensors Using Deep Pyramidal Residual Network [21]	DHA	Accelerometer	1D-PyramidNet	96.64%	Introduced the 1D-PyramidNet model which uses an incremental strategy for feature map expansion.
4	An Intelligent Robust One Dimensional HAR-CNN Model for Human Activity Recognition using Wearable Sensor Data [37]	WISDM	Accelerometer	HAR-CNN model	95.2%	Proposed a 1D HAR-CNN model and collected a new dataset
5	Smartwatch-based Human Activity Recognition Using Hybrid LSTM Network [40]	WISDM	Accelerometer and gyroscope	CNN-LSTM model	96.2%	Proposed a 2-layer CNN-LSTM model and tuned the hyper-parameters using Bayesian optimisation.
6	Enhanced Complex Human Activity Recognition System: A Proficient Deep Learning Framework Exploiting Physiological Sensors and Feature Learning [44]	New HAR dataset using EMG sensors	8-channel EMG sensors	1D CNN-LSTM	83%	Proposed a lightweight model. Used physiological sensor (EMG). Collected a new HAR dataset with EMG.
7	An Efficient and Lightweight Deep Learning Model for Human Activity Recognition on Raw Sensor Data in Uncontrolled Environment [41]	New Calibrated Dataset, MotionSense and mHealth datasets.	Accelerator and gyroscope	1D CNN-LSTM	98% on new dataset. , MotionSense dataset: 99% and mHealth: 99.2%	Proposed a Conv1D based CNN-LSTM model and developed a framework for DL based HAR on sensor data. Also collected and calibrated a new HAR dataset.
8	Compressed Deep Learning Model For Human Activity Recognition [42]	WISM	Accelerator and gyroscope	Multi-input CNN-LSTM	99.74%	Introduced a multi-input CNN-LSTM model with dual input streams
9	Wireless body area sensor networks based Human Activity Recognition using deep learning [43]	mHealth	ECG, accelerometer, gyroscope and magnetometer	GAF-DenseNet169	97.83%	Proposed using GAF to transform 1D time series data to 2D images

Table 3. *Cont.*

S/N	Paper	Dataset	Sensors	Model	Accuracy	Contribution to Knowledge
10	Deep Learning for Multiresident Activity Recognition in Ambient Sensing Smart Homes [46]	ARAS multiresident dataset	Force sensor/pressure mat, photocell, contact sensors, proximity sensors, infrared receiver, temperature sensors and sonar distance sensor	RNN models (GRU and LSTM)	88.21% for GRU and 86.55% for LSTM	Used GAN to generate more data and compared the performance of GRU and LSTM models.
11	Human Activity Recognition in Smart Home using Deep Learning Models [48]	ARAS	Force sensor/pressure mat, photocell, contact sensors, proximity sensors, infrared receiver, temperature sensors and sonar distance sensor	MLP, RNN and LSTM	LSTM 0.92 R1 and 0.91 R2 RNN 0.91 R2 and R1 MLP 0.92 R1 and 0.92 R2	Compared the performance of three models on the ARAS dataset
12	Multisource Transfer Learning for Human Activity Recognition in Smart Homes [47]	CASAS dataset (HH101, HH103, HH105 and HH109)	Motion sensor, Door sensor, Wide-area sensor	TRs-LSTM	TRs method performed better than transferring HAR model based on only common sensors	Proposed transferring HAR models from a labelled home to an unlabeled one using Transferable sensor representations
13	Multi-Resident Activity Recognition using Multilabel Classification in Ambient Sensing Smart Homes [49]	ARAS multiresident dataset	Force sensor/pressure mat, photocell, contact sensors, proximity sensors, infrared receiver, temperature sensors and sonar distance sensor	Classifier Chain method of the Multi Label Classification (MLC) technique with KNN as base classifier	0.931 in House B and 0.758 in House A	Proposed an approach which uses the correlation between activities to recognise activities and the resident carrying them out.

5. Recent HAR Algorithms

HAR has evolved significantly over time, driven by advancements in research and technology. Research on HAR dates as far back as the 1990s [50]. Early HAR research was on rule-based systems, which used hand-crafted rules for simple activity recognition [51]. Then, the early 2000s saw an increase in statistical methods such as feature engineering for HAR [52] followed by machine learning algorithms. Identifying human activities using machine learning (ML) often relies heavily on manually extracting specific features [53]. This reliance on human knowledge can be limiting. These ML models need significant preprocessing which can be time-consuming [54]. Some of the popular ML algorithms

include K-Nearest Neighbour, Random Forest, Support Vector Machine, Hidden Markov and Gaussian Mixture [55].

The 2010s saw the start of research on deep learning for HAR. Researchers turned to deep learning algorithms which automate feature extraction to reduce the time for preprocessing and improve the accuracy of HAR models. Deep learning architectures, particularly Convolutional Neural Networks, Recurrent Neural Networks, and Long Short-Term Memory networks, have gained significant traction in the field of HAR [53]. These deep learning algorithms can automatically learn abstract patterns from raw sensor data during training, eliminating the need for manual feature extraction. They are also able to handle more complex activities effectively. However, they require a large amount of data for training. Over time, different deep learning algorithms have been proposed for the task of HAR, such as DenseNet [35], TCNN [24], CNN-LSTM models [25,42,44], DC-GCN [26], IST-GCN [29], DMST-GRNN [30], and DEBONAIR [11].

In recent years, in addition to DL algorithms, there has been some research using multimodal techniques, DRL and LLMs. This section will focus on these three recent algorithms.

5.1. Multimodal HAR

Multimodal approaches to HAR stem from the need to accurately and properly track human activities, especially considering the complexity of human movements, which can pose a challenge to single-modality HAR methods [56]. Also, the advancement in smart objects, wearable sensors, and IoT technology has enabled the collection of multimodal data [57].

The different sensor outputs that can be combined include RGB-D images, skeleton data, inertial data, wearable sensor data (smartwatches, in-ear sensors, etc.), and ambient sensor data [56–61]. Some researchers have also explored combining different forms of the same modality like RGB images with skeleton images [61], acceleration with angular velocity [62], and so on. Multimodal HAR will allow for the proper integration of physiological data into HAR, which would enhance the efficiency of the application of HAR in healthcare.

Some researchers have proposed using additional sensors to refine the activities' context. Bharti et al. [58] proposed using a combination of movement sensors and ambient environment sensors, where the ambient sensors will provide context of the activities detected from the movement sensors. They leveraged movement sensors, ambient environment sensors, and location context to develop a multimodal HAR model. They used an ablation test to show that their model gave the highest accuracy with the different sensors integrated than when alone. Mekruksavanich et al. [63] also proposed adding location context data to their model to improve its accuracy. They tested their model on the DHA dataset and it outperformed other standard models.

Rashid et al. proposed combining smartwatch and earbuds for HAR [56]. They collected data from 44 subjects (24 males and 20 females) from a controlled in-lab environment and an in-home environment. Then, they trained and tested five different classifier models. Their result showed an overall improvement in the detection of the activities with a combination of both smartwatch data and earbud data. Although smartwatch data were overall more useful, earbuds data showed better performance in detecting moving data.

Hnoohom et al. proposed a model based on the combination of data from a smartwatch and smart-shoes embedded with accelerometers [64]. In their paper, they present a deep residual network model called HARNeXt, which they trained and tested on the 19NonSens dataset [65]. They ran three experiments using their proposed model and three baseline models (CNN, LSTM, and CNN-LSTM). The first experiment used only the sensor data from the smart-shoes, the second used only data from the smartwatch, and the third used data from both the smart-shoes and the smartwatch. Their proposed model performed better than the baseline models in all three experiments. It had an accuracy of 79.55%, 93.26%, and 97.11%, respectively, for the different experiments. The confusion matrix of the third experiment also showed an accuracy of more than 90% for all activities.

Due to the varied nature of the data being combined in multimodal systems, researchers have proposed the use of parallel streams of models combined using fusion. This has allowed for the creation of multimodal HAR models which can take as inputs a variety of data like video and audio, images and inertia data, RGB images and depth/skeletal images, and so on. The individual component models can then be chosen and adapted based on the type of data that will be fed into them. This accounts for the differences in the sensor data and allows for feature extraction from different dimensions [66]. Fusion is then used to combine the outputs of the component models. Fusion can happen at the data level, feature level, or classifier level [11,62].

In vision-based HAR, the use of parallel stream networks with fusion has been proposed in different ways by researchers [9,59–61,67]. In 2020, Kumrai et al. proposed a parallel stream network with one stream for the skeleton information using an LSTM model and the other for the RGB Image using a VGG-16 model. The outputs from these streams were then concatenated and processed through densely connected layers [61]. In 2021, Zehra et al. proposed the use of ensemble learning with Multiple CNNs [68]. They compared the performance of different ensembles of three 1D CNN models with the performance of the single models and found that the ensembles had a higher accuracy. Also in 2021, Das et al. proposed a Multimodal Human Activity Recognition Ensemble Network (MMHAR-EnsemNet) which uses multiple streams to process information from the skeleton and RGB data in addition to the accelerometer and gyroscope data [67]. They trained and validated their model on the UTD-MHAD and Berkeley-MHAD datasets. They compared the model's performance using different combinations of the inputs, and the model had the highest accuracy when all four modalities were used.

In 2022, Guo et al. proposed a three-stream model that uses reinforcement learning for data fusion [69]. Their model took as input data the RGB image, skeleton and depth data from the NTU RGB and HMDB51 datasets. They proposed that each modal datum cannot be weighed equally because they have different representation abilities for various actions [69]. Thus, using reinforcement learning to determine the fusion weights would account for this variability. They trained and tested their model on the NTU RGB and HMDB51 datasets and performed ablation studies to show the effectiveness of their reinforcement learning multimodal data fusion method.

Lin et al. proposed the combination of cameras and movement sensors for the health monitoring of people with mobility disabilities [57]. Their model took as inputs the skeleton sequence from the camera and the inertial sequence from the movement sensors. Two component models, ALSTGCN and LSTM-FCN, were used to process the skeleton sequence and inertial sequence, respectively. An Adaptive Weight Learning (AWL) model was used to fuse the skeleton and inertial features optimally. They tested their model on two public datasets, C-MHAD and UTD-MHAD, and on M-MHAD, which is a novel dataset created by them. They compared the model with other state-of-the-art single-modality models, and it performed better [57].

Some researchers have also proposed combining non-vision inputs in multiple stream models. Zhang et al. proposed a multistream HAR model called 1DCNN-Att-BiLSTM, which combines a one-dimensional CNN, an attention mechanism, and a bidirectional LSTM [66]. The sensor inputs are processed in parallel and fed into the 1DCNN and BiLSTM models connected in series. An attention mechanism was used to select important features. Then, a fully connected layer was used for the fusion of the feature data obtained from each channel. A softmax layer was used for the final output. They tested their model on the Shoaib AR, Shoaib SA and HAPT datasets. Their focus was on the combination of three sensor types: accelerometer, gyroscope, and magnetometer. They compared the performance of their model to other models and demonstrated its superior performance. In their work, they also compared the performance of various sensor combinations, and showed that the accelerometer and gyroscope sensor combination performed better than the accelerometer and magnetometer or gyroscope and magnetometer combination [66].

Muhoza et al. proposed a position-aware HAR model using a multimodal Deep Convolutional Neural Network (DCNN) [70]. They proposed that forming a body network using data from sensors on different positions on the body would improve activity recognition. They treated each position independently by passing them through an ensemble of Convolutional Neural Networks (CNNs) before fusion. The fusion block is made up of fully connected layers that concatenate the outputs from the CNN blocks. Using the Leave-Subjects-Out Hold Out (LSOHO) method, they trained and validated their model on the SHO [71] and mHealth [72] datasets. Their result showed an overall improvement compared to simple multiposition and single-position models.

Shi et al. proposed a Distributed Sensors Fusion Network (DSFNet) [62]. The model has two branches: an angular velocity-based branch and an acceleration-based branch. Angular velocity and acceleration are taken from the inertial sensors on different positions on the body and fed into the corresponding branch of the model. The feature maps from both branches are expanded into a vector and fed into a classification network, which then outputs the classification score for the different categories. They evaluated their model on the Comprehensive Multimodal Human Action Dataset (CZU-MHAD) [73], and it showed competitive performance. They performed an ablation study and showed improved performance by the post-fusion model compared to the pre-fusion model [62].

Chen et al. proposed a multimodal deep learning model for recognising Complex Human Activities called DEBONAIR (deep learning-based multimodal complex human activity recognition) [11]. They classified their input data into three based on their properties: fast-changing and simple data, fast-changing and complex data and slow-changing data. They proposed using different sub-networks to process each type of data. Then, a depth concatenation operation was applied to the output of the sub-networks, and a convolutional layer was used to fuse the features. Two LSTM layers were used to learn sequential information from the fused features, and a final fully connected layer was used to generate the probability distribution for the Complex Human Activities (CHA). DEBONAIR was evaluated on the lifelog dataset and the PAMAP2 dataset using weighted F1-score. It was compared to different models (Hierarchy [74], Non-hierarchy [75], SADeepSense [76], DeepSense [77], Channel-based Late Fusion [78], and DeepConvLSTM [79]) and it scored higher.

Mahmud et al. proposed the use of a Deep Multistage LSTM model to integrate features from multiple sensors for HAR [80]. Their proposed model first extracts temporal features from each sensor datum, which is then aggregated and optimised before being fed into the final activity prediction layer. They tested the model on a dataset from Physionet [81] and compared the effect of the different sensors against the performance of integrating multiple sensors. Their result showed that the integration of multimodal features led to an improvement in accuracy.

Multimodal HAR has been shown to demonstrate improved performance over single-modality HAR [11,56,58,67]. It allows for combining different data types as each stream of the model can be adapted to a particular data type. Table 4 shows a summary of some of the multimodal HAR models that have been proposed in recent literature. Almost all of them had an accuracy of over 91% in the experiments conducted, and where ablation tests were performed, they showed that the multimodality increased the accuracy of the models. Its application in healthcare would allow for combining inertial data and vital health data for a more robust HAR solution.

However, one of the issues faced by this method is the availability of multimodal datasets. This is largely due to the novelty of the method; with an increase in interest in this methodology, there will also be an increase in the availability of datasets. Section 6 covers some of the available multimodal datasets. Quality multimodal datasets have to ensure that the data are not misaligned, which can be caused by the heterogeneity of the different data types. It is also computationally intensive to process long sequences of multimodal data, and it requires complex fusion techniques. Research in this area should focus on finding optimal solutions to these issues.

Table 4. Multimodal Human Activity Recognition papers.

S/N	Paper	Dataset	Model	Accuracy	Contribution to Knowledge
1	Human Activity Recognition Through Ensemble Learning of Multiple Convolutional Neural Networks [68]	WISDM	Ensemble of three CNN models	93.66%	Proposed an ensemble of CNN models
2	MMHAR-EnsemNet: A multimodal Human Activity Recognition Model [67]	UTD-MHAD and Berkeley-MHAD	MMHAR-EnsemNet	0.991 on UTD-MHAD and 0.996 on Berkley-MHAD	Proposed a novel deep learning based ensemble model called MMHAR-EnsemNet
3	A Deep Reinforcement Learning Method For Multimodal Data Fusion in Action Recognition [69]	NTU RGB and HMDB51	Twin Delayed Deep Deterministic (TD3) for data fusion	94.8% on NTU RGB+D and 70.3% on HMDB51 dataset	Proposed a reinforcement learning based multimodal data fusion method.
4	Adaptive multimodal Fusion Framework for Activity Monitoring of People With Mobility Disability [57]	C-MHAD, UTD-MHAD and M-MHAD	ALSTGCN and LSTM-FCN models using an Adaptive Weight Learning (AWL) features fusion.	91.18% and the recall rate of falling activity is 100%	Proposed a deep and supervised adaptive multimodal fusion method (AMFM) and collected a new multimodal human activity dataset, the H-MHAD dataset.
5	A multichannel hybrid deep learning framework for multisensor fusion enabled Human Activity Recognition [66]	Shoaib AR, Shoaib SA and HAPT datasets	1DCNN-Att-BiLSTM	99.87% on Shoaib SA, 99.42% on Shoaib AR and 98.73% on HAPT	Proposed a multistream HAR model called 1DCNN-Att-BiLSTM and also compared the performance on various sensor combinations.
6	Multiposition Human Activity Recognition using a multimodal Deep Convolutional Neural Network [70]	Shoaib and mHealth	Multichannel 1D CNN models fused using a fully connected layer.	97.84% on Shoaib and 91.77% on mHealth	Proposed a multimodal deep CNN capable of recognizing different activities using accelerometer data from several body positions.
7	DSFNet: A Distributed Sensors Fusion Network for Action Recognition [62]	CZU-MHAD	DSFNet	Ranging from 91.10% to 100% on different experimental settings	Proposed a distributed sensors fusion network (DSFNet) for multisensor data which uses one-to-many dependencies for acceleration and local–global features for angular velocity

Table 4. *Cont.*

S/N	Paper	Dataset	Model	Accuracy	Contribution to Knowledge
8	Deep learning-based multimodal complex Human Activity Recognition using wearable devices [11]	Lifelog and PAMAP2	DEBONAIR	F1-score of 0.615 on lifelog and 0.836 on PAMAP2 dataset.	Proposed using different sub-networks to process fast-changing and simple, fast-changing and complex and slow-changing data.
9	Human Activity Recognition From multimodal Wearable Sensor Data Using Deep Multistage LSTM Architecture Based on Temporal Feature Aggregation [80]	Wrist PPG data from Physionet	Deep Multistage LSTM	Average F1 score of 0.839	Proposed individual LSTM streams for temporal extraction of each data type.

5.2. HAR Using Deep Reinforcement Learning

Deep Reinforcement Learning (DRL) has recently emerged as a promising alternative or addition to traditional supervised learning approaches to HAR. DRL is a combination of deep learning and reinforcement learning. Reinforcement learning is a reward-oriented learning technique that is based on trial and error.

Guo et al. proposed using DRL for data fusion in a multimodal system and performed an ablation study to show the effectiveness of this DRL-based data fusion [69]. Dong et al. proposed using DRL in attention-aware sampling for HAR. Their proposed method involves training an attention model using DRL to identify the key frames in a video and discard the irrelevant ones [82]. They evaluated the performance of their method on the UCF101 and HMDB51 datasets and it performed competitively. In the same vein, Wu et al. proposed a Multiagent Reinforcement Learning (MARL) framework that uses DRL to identify relevant frames for the task of HAR [83]. In their paper, they proposed using multiple agents, where each agent picks a series of frames encoded into a vector which are then fed into a policy network. When a frame is selected as relevant, a classification model then classifies the activity being carried out on that frame. They examined MARL with various architectures (BN-Inception, Inception-V3, ResNet-101, ResNet-152 and C3D) and it improved the overall accuracy of all the models. They also compared its performance on the ActivityNet, YouTube Birds and YouTube Cars datasets in comparison to other state-of-the-art models. They found that the method with ResNet-152 outperformed other models.

In 2021, Nikpour and Armanfard proposed a Joint Selection Deep Reinforcement Learning (JSDRL) framework that selects the key joints in the skeleton data of a video and uses that for Human Activity Recognition [32]. It formulates the problem as a joint selection problem and uses DRL to find the best solution. It selects and filters out the relevant joints, thus enhancing the classifier's performance and reducing training time. They compared the performance of three models (BiLSTM, CNN, and decoupling graph Convolutional Neural Networks with dropGraph module (DCGCN)) with and without the JSDRL framework on the NTU-RGBD and UT-Kinect Datasets, and the models performed better with the framework than without. The framework achieved the highest accuracy when paired with DCGCN.

Zhang et al. also proposed a DRL framework called the Dynamic Key Feature Selection Network (DKFSN) for feature extraction on time-series sequential data [84]. The DKFSN uses a reinforcement agent to select the best deep features to optimise recognition and

disregard the ones that negatively affect recognition. The baseline classification model used for the framework is the BiLSTM with fully connected layers, and it was validated on the Opportunity and UCI HAR datasets. From their experiments, the baseline model had an accuracy of 89.16% on the opportunity dataset, while the DKFSN had an accuracy of 89.74%. On the UCI HAR dataset, the baseline model had an accuracy of 92.06%, while the DKFSN had an accuracy of 93.82%.

In 2019, Zhang and Li proposed a DRL-based human activity prediction algorithm in a smart home environment [85]. They proposed the use of Deep Q-Network (DQN) for recognising and predicting human activities in a smart home environment. They compared the performance of their proposed methodology to more traditional methods (bag, set and list approaches) and found that the DQN neural network performed better than the others.

In 2019, Raggioli and Rossi proposed a DRL framework that adaptively positions the home robot in a way that does not distract or discomfort the user [86]. They used a DNN with LSTM for activity recognition which had a 97.4% accuracy on the PAMAP2 dataset. Their proposed DRL framework had an 82% success rate for the simulated episodes. Kumrai et al. in 2020 also proposed using DRL to control the movement of a home robot to maximise its recognition of human activities [61]. They used a deep Q-network (DQN) to automatically control the movement of the home robot. The agent was trained to maximise the confidence value of the HAR model. The HAR model used in their work is a dual-stream network that takes skeleton data and cropped images fed into LSTM and VGG-16 streams respectively. The streams are then concatenated, and the recognised activity is outputted with a confidence value. A virtual environment was used to evaluate the proposed DRL method. In the same vein, Ghadirzadeh et al. also proposed a DRL framework that aligns the action of a robot based on the recognised action of the human [87].

DRL is a promising technique for improving HAR. It has been used to improve the results for the fusion of multistream models [69], adjust the movement of robots in smart homes for better HAR [61,86], improve human–robot interactions [87] and increase the efficiency of feature extraction [32,83,84]. There are ethical challenges to the use of DRL in real-world applications such as its compliance with legal standards and issues around privacy and algorithmic bias. It also has a black-box nature, which makes interpretability of the algorithm difficult, and this limits its clinical applicability. While DRL trained in simulated environments might perform well, they may not perform well in real-world environments. These challenges highlight the need for more research in this area to improve the applicability of DRL in HAR.

5.3. HAR Using Large Language Models

With the boom of Artificial Intelligence, LLMs have been at the forefront of more recent research in business, cyber security, finance and healthcare [88]. The task of HAR is not left out of this trend. This subsection reviews some of these papers. While supervised learning might achieve high accuracy in HAR tasks, it requires a large amount of annotated data, which is expensive and time intensive [89]; with LLMs, this can be avoided or mitigated.

Kim et al. evaluated twelve LLMs on various health prediction tasks including HAR and presented a fine-tuned LLM model, HealthAlpaca [90]. They used four public datasets—PMData, LifeSnaps, GLOBEM, and AW_FB—for prompting and fine-tuning the LLMs. Their fine-tuned model achieved the highest accuracy in 8 out of the 10 tasks considered including HAR. They also performed an ablation study that showed the importance of context enhancement strategies (combining health knowledge, user context and temporal information).

Ji et al. proposed using a zero-shot human activity recogniser, HARGPT, which takes raw sensor data with a simple prompt and yields a recognition outcome [91]. They used well-known LLMs such as ChatGPT, Google Gemini, and LLaMA2-70b with a chain-of-thought (CoT) prompt design. They validated their models on the Capture24 and HHAR datasets, and they achieved an average accuracy of 80%. The authors compared the performance of GPT4 to various baseline models (Random Forest, SVM, DCNN, and

LIMU-LSTM). The baseline models achieved over 90% accuracy on the seen datasets but on unseen data, GPT4-CoT surpassed them with F1 scores of 0.795 on Capture24 and 0.790 on the HHAR dataset.

Gao et al. proposed a model called LLMIE_UHAR that uses Iterative Evolution and LLM [89]. Their proposed model first selects valuable data from the large unlabelled dataset using a clustering algorithm. The selected data points are then changed into prompts which are then used as input into the LLM, which annotates them. The annotated data are then used to train a neural network. Their model integrates clustering algorithm, LLMs, and the neural network to enhance the HAR task. They validated their model on the ARAS dataset and it achieved an accuracy of 96%.

In 2021, Xu et al. developed LIMU-BERT (lite BERTlike self-supervised representation learning model for mobile IMU data), which makes use of unlabelled data [92]. Their model is based on BERT (Bidirectional Encoder Representations from Transformers), which is a language representation model. They validated their model on four datasets: HHAR, UCI, MotionSense and Shoaib datasets. The LIMU-BERT model achieved an average accuracy of 0.929 and F1 score of 0.921 on all datasets, which was higher than other baseline models (DCNN, DeepSense, R-GRU, and TPN). In 2024, Imran et al. developed LLaSa (Large Language and Sensor Assistant), which combines LIMU-BERT and Llama [93]. In their paper, they also introduced SensorCaps, a movement sensors activity narration dataset and OpenSQA, a question–answer dataset. These datasets were created using movement sensor data from HHAR, UCI-HAR, MotionSense and Shoaib datasets. The model was evaluated first as a closed-ended zero-shot task on five datasets (HHAR, UCI-HAR, MotionSense, Shoaib, and SHL) and then as an open-ended task on PAMAP2 dataset. Its performance was compared to that of non-fine-tuned and fine-tuned GPT-3.5-Turbo and it performed much better.

Fang et al. developed PhysioLLM, which integrates contextual data to physiological data from a wearable to provide personalised understanding of health data and suggesting actions to achieve personal health goals [94]. They used two LLMs in their system, one to generate insights from the data and another for the conversation side of the system. Both LLMs are based on OpenAI's GPT-4-turbo model. They validated their system using a user study of the sleep pattern of 24 Fitbit watch users and their system outperformed the Fitbit App and a generic LLM. Two sleep experts also carried out a preliminary valuation of the system and they concluded that the system gave valuable, actionable health advice.

LLMs can be used in the annotation of large datasets and in the developing of personalised HAR systems and models that better understand and predict human behaviour using contextual information and user preferences. However, LLMs face some issues such as ethical issues concerning bias and privacy [90]. Due to the black-box nature of LLMs, it is also difficult to assess their clinical validity. There is also the problem of hallucination, where the model makes up an answer that is incorrect or that does not exist. In order to maximise LLMs in HAR, it is necessary to explore solutions to these issues, such as incorporating explainable Artificial Intelligence (AI) methods and fine-tuning these models to specific use cases to avoid hallucination.

6. HAR Datasets with Physiological Data

The availability of high-quality datasets is important for developing and evaluating any HAR model. Traditionally, HAR relies on motion data from wearable inertial or ambient sensors or image data from vision sensors. Some of the notable datasets are Weizmann [95], KTH Action dataset [96], Kinect Activity Recognition Dataset [97], Cornell Activity Dataset [98], MSR Daily Activity 3D Dataset [99], UCF50 [100], HDM05 [101], NTU RGBD [102], NTU RGBD-120 [103], Northwestern-UCLA, Human 3.6M [104], CMU Mocap, WISDM [105], Daily Human Activity (DHA), ARAS [106], CASAS, OPPORTUNITY [107], and UCI HAR [108].

However, research has shown that physiological data can improve the accuracy of HAR models [75], especially when applied to healthcare use cases. While numerous

datasets for HAR with motion data exist, datasets incorporating physiological data are relatively limited. While these sensors offer precious insight into movement patterns, they fail to capture the physiological and contextual aspects of human activities [17]. Researchers have recently recognised this limitation and started integrating health vitals with traditional motion data. This section will review selected HAR datasets with physiological data incorporated. These datasets include PAMAP2 [109], ETRI lifelog [17], mHealth [110], EMG physical action, 19NonSense, and Wrist PPG During Exercise [81] datasets.

PAMAP2 was introduced as a new dataset for physical activity monitoring and benchmarked on several classification tasks [109]. The dataset contains over 10 h of data collected from 9 subjects (8 males and 1 female) performing 18 activities. The activities included 12 protocol activities (lie, sit, stand, walk, run, cycle, Nordic walk, iron, vacuum clean, rope jump, and ascend and descend stairs) and 6 optional activities (watch TV, computer work, drive car, fold laundry, clean house, and play soccer) [109]. Three Inertial Measurement Units (IMUs) and a heart rate monitor were used to capture data. The IMUs were attached with straps on the dominant arm and ankle and the third one was attached with the heart rate monitor on the chest. The dataset was made publicly available on the internet.

ETRI lifelog dataset was collected from 22 participants (13 males and 9 females) over 28 days using a variety of sensors, including smartphones, wrist-worn health trackers, and sleep-quality monitoring sensors [17]. The dataset includes physiological data (photoplethysmography (PPG), electrodermal activity (EDA), and skin temperature), behavioural data (accelerometer, GPS, and audio) and self-reported labels of emotional states and sleep quality. The authors demonstrated the feasibility of the dataset by using it to recognise human activities and extract daily behaviour patterns [17]. They suggested that the dataset can be used to understand the multifaceted nature of human behaviour and its relationship to physiological, emotional, and environmental factors. The dataset contains over 2.26 TB of data with 590 days of sleep quality data and 10,000 h of sensor data.

The mHealth dataset includes body motion and vital signs recordings from 10 participants performing 12 physical activities [110]. The data were collected using Shimmer2 wearable sensors placed on the chest, wrist, and ankle. The sensor on the chest contains a 2-lead ECG. The activities range from simple ones like standing still or lying down to more complex ones like cycling or running. The publicly available dataset can be used to develop and evaluate activity recognition models.

Wrist PPG During Exercise is a publicly available dataset of PPG signals collected during exercise. The dataset includes PPG signals from the wrist, electrocardiography (ECG) signals from the chest, and motion data from a 3-axis accelerometer and a 3-axis gyroscope [81]. The PPG signals were recorded during walking, running, and cycling at two different resistance levels. The dataset is intended for use in developing and validating signal processing algorithms that can extract the heart rate and heart rate variability from PPG signals during exercise.

The 19NONSens (Non-obstructive Sensing) dataset is a human activity dataset collected using a smartwatch and an e-shoe [65]. The dataset was collected from 12 subjects doing 18 activities in indoor and outdoor contexts. The smartwatch, Samsung Gear S2, was equipped with accelerometer, gyroscope, heart rate sensor, a thermal, and a light sensor, while the e-shoe had an accelerometer embedded in the sole. The subjects wore the smartwatch on their preferred hand and performed the 18 activities, which included 9 indoor activities and 9 outdoor activities.

The EMG Physical Action dataset contains a dataset of 20 physical actions performed by four subjects (3 males and 1 female) [45]. The data were collected using EMG sensors with 8 electrodes. The subjects were recorded using Delsys EMG apparatus while performing 10 aggressive and 10 normal physical actions.

PPG-DaLiA is a publicly available multimodal dataset recorded from 15 subjects wearing a wrist and a chest-worn device while performing daily life activities [111]. The chest-worn device provides ECG data, respiration data, and three-axis acceleration data, while

the wrist-worn device provides body temperature, blood volume, electrodermal activity data, and acceleration data. Table 5 provides an overview of the datasets reviewed.

Table 5. HAR datasets with physiological data.

S/N	Dataset	No. of Participants	Sensors Used	Activities
1	PAMAP2	9	accelerometer, gyroscope, heart rate, magnetometer	lying, sitting, standing, walking, running, cycling, Nordic walking, watching TV, computer work, car driving, ascending stairs, descending stairs, vacuum cleaning, ironing, folding laundry, house cleaning, playing soccer, rope jumping
2	ETRI lifelog Dataset	22	GPS, PPG, accelerometer, gyroscope, heart rate, magnetometer, skin temp	Sleep, personal care, work, study, housework, caregiving, media, entertainment, sports, hobby, free time, shopping, regular activity, transport, meal, social
3	Wrist PPG During Exercise	23	ECG, PPG, accelerometer, gyroscope, magnetometer	walking, running, easy bike riding and hard bike riding
4	PPG-DaLiA	15	ECG, PPG, accelerometer	Sitting still, Ascending/Descending stairs, table soccer, cycling, driving car, lunch break, walking, working
5	EMG Physical Action Dataset	4	EMG	Bowing, Clapping, Handshaking, Hugging, Jumping, Running, Seating, Standing, Walking, Waving, and aggressive actions, such as Elbowing, Front kicking, Hammering, Headering, Kneeing, Pulling, Punching, Pushing, Side-kicking, and Slapping.
6	19NonSense dataset	12	accelerometer, gyroscope, heart rate, light sensor, themal sensor	Brushing, Washing hand, slicing, peeling, upstair, downstair, mixing, wiping, sweeping floor, turning shoulder, turning knee, turning haunch, turning ankle, walking, kicking, running, cycling
7	mHealth dataset	10	accelerometer, gyroscope, 2-lead ECG	Standing still, Sitting and relaxing, Lying down, Walking, Climbing stairs, Waist bends forward, Frontal elevation of arms, Knees bending (crouching), Cycling, Jogging, Running, Jump front and back.

7. Applications of HAR in Healthcare

HAR can be applied in different sectors; however, healthcare research is of paramount importance [112]. Human health and well-being can be greatly improved through the recognition of human activities. In healthcare, there are numerous applications of HAR. Some of these applications are highlighted in this section.

7.1. Assistive Healthcare for Elderly People

HAR is valuable in providing assistive healthcare for elderly or vulnerable groups especially those living alone. There is an increasing population of elderly people [113], and they are vulnerable to diseases and home accidents. However, access to healthcare services requires them going to hospitals, which might not always be feasible. With the incorpora-

tion of health vital sensors in wearable devices, their health vitals can be monitored in the comfort of their homes without interrupting their everyday routine.

Wearable HAR devices can also be used to detect potential issues such as a fall or a decline in physical activities for this vulnerable group. In the event of a fall, caregivers and the necessary healthcare personnel can be notified for prompt intervention. It can also be used to identify and mitigate potential fall risks using gait patterns and balance. This is especially important because falls are one of the main causes of injuries and even death in the elderly [114].

However, these wearable sensors would depend on the elderly to wear them always, which might not always be convenient [115]. Also there is a question of the accuracy of some of these sensors, especially with healthcare applications where there is very little room for error. Generally, the adoption of new technology is not always easy for the elderly; using HAR systems that rely less on them operating it might be the best course of action. Also with the increase in more comfortable wearable systems, some of which are integrated into smartwatches, adoption by the elderly population should also increase.

Smart homes with cameras or other ambient sensors, like those modelled in the ARAS and CASAS datasets, can be used by caregivers and healthcare practitioners to monitor the health status of the elderly and detect abnormalities in their pattern of activities. This can be used to monitor adherence to medication schedule and even in the early detection of cognitive decline (Alzheimer's and dementia). This method would eradicate having to depend on the elderly to wear the sensors; however, there is still the issue of privacy and data protection. With the HAR system monitoring their every move, it could be considered invasive, and some might feel like their independence is being taken away.

For independent elderly people, quick intervention is important in the event of an accident, and this is dependent on quick reporting or the detection of issues, and HAR-based systems can help achieve this. However, the adoption of HAR in the assistive healthcare of the elderly depends on issues such as privacy and data protection being adequately handled.

7.2. Mental Health Issues

HAR has the potential to detect mental health issues (anxiety, depression, and dysthymia) and even monitor symptoms of mental health issues by analysing daily activity and sleeping patterns. Mental health issues is usually characterised by a change in the pattern of daily activities, with the patient becoming less interested in certain activities. Less physical activity has generally been associated with people dealing with mental health issues. People with anxiety are likely to experience or carry out certain anxiety behaviours such as biting their nails, pulling their hair, pacing, and so on. They might also experience diarrhoea or constipation. Insomnia is also one of the symptoms of mental health issues. These activities can be detected by HAR systems, and this can be the foundation of a HAR-based mental health support system.

People with major depressive disorder and who are at risk of self-harm can also be monitored using HAR-based systems. While current mental health detection and monitoring depends on self reporting, it is flawed because it relies on the patient being able to accurately, and without bias, report their activities. However, with HAR systems, these activities can be automatically and accurately detected, and it gives the mental health professional accurate information to work with.

The detection and monitoring of mental health issues using wearable or ambient sensors has been a long-standing research goal [116]. One of the limitations faced in this research area is the availability of large datasets and also the lack of uniformity in the data collected. This is a problem that needs to be tackled for the adoption of HAR systems in the detection and monitoring of mental health issues.

7.3. Personalised Health Recommendations

HAR can be used to provide personalised recommendations on healthy habits. Maintaining a healthy lifestyle is individualistic because it is affected by things like physical habits, job, genetics, and social habits. The daily nutritional requirements for a professional athlete would be different from that of an office worker. Underlying health conditions can also influence the nutritional needs of the user. A personalised nutrition monitoring or recommender system based on HAR would ensure that these peculiarities and differences are considered when it comes to the nutrition needs of the user.

A HAR-based exercise recommender system can be personalised based on the activities of the user. This can be useful for people with Parkinson's and other motor-generative diseases, as personalised exercises can be used to augment medication [117]. For athletes, this system can be used to emphasise weak spots in their training regimen. Personal trainers can also used this for more personalised exercise recommendation. These health recommendations can also change in response to changes in the user's pattern of activities. This allows for dynamic health recommendations that change with changes in the need of the user.

7.4. Early Detection of Diseases

HAR, especially combined with physiological data, can be used for the early detection of diseases. For instance, peripheral neuropathy can be detected early using gait analysis. Diabetic patients who are at risk for peripheral neuropathy can be monitored using the HAR system with smart shoe sensors or other wearable devices.

Biomedical sensors such as ECG and EMG sensors can be used together with inertial sensors to detect and monitor diseases, which can be helpful for caregivers and help patients receive prompt help where necessary. This can also be applied in the detection of strokes, pneumoconiosis and other diseases [118,119]. In the same vein, heart attacks can be detected based on the patient's health vitals and motions detected by a HAR system. Healthcare practitioners can use a HAR-based system to detect and monitor some of these diseases especially for people who are generally at risk due to genetics, age, underlying health conditions, or lifestyle.

Most biomedical sensors like ECG sensors use probes or straps which are uncomfortable to wear, especially for everyday tasks. Although there has been progress in the development of wearable biomedical sensors such as over-the-ear ECG sensors, there is still a concern about the accuracy and adoption of these wearable biomedical sensors. There are also limited datasets with these sensors due to the rarity of their use in HAR. More research needs to be performed in this application area to fully realise the potential of HAR in the early detection of diseases.

7.5. Monitoring of Physical Rehabilitation Performance

Physical therapists can use HAR to monitor the performance of their patients and curate personalised rehabilitation plans for them. HAR systems can help them have a holistic view of the performance of their patients and note areas of weakness with respect to the patient's everyday life. This means that they can curate a rehabilitation plan tailored to each patient and also effectively monitor the progress of each patient. Physical therapy usually involves constantly re-assessing the biomechanical abilities of the patient, and this takes up a considerable amount of time in the recovery process [120]. With a HAR-based rehabilitation system, the time for recovery would be reduced, with the time spent on constant biomechanical assessment being eliminated or greatly reduced.

Patients carrying out rehabilitation in a familiar environment such as their home have shown to improve their performance after being discharged [121]. However, physical therapists may not be available to monitor their progress at home, and family members lack the knowledge to properly monitor and guide them [122]. A HAR-based monitoring system would allow the patients to carry out their rehabilitation at home and still receive the needed guidance from their physical therapists.

It is important to note that due to the fatal nature of mistakes in healthcare, these suggested application areas are to complement the healthcare professionals and not to replace their expertise.

8. Challenges and Future Directions

8.1. Challenges

HAR has a lot of potential, but it also faces some challenges which need to be addressed for optimal performance and a wider adoption. Some of the challenges of HAR systems are as follows.

8.1.1. Data Privacy and Security

HAR relies on the collection of personal data on people's activities, which raises concerns about the privacy of the data in HAR systems, user consent, and the potential misuse of data. Data privacy is a crucial issue for any system that uses sensitive information from the user, and HAR systems are no different. Due to the intrusive and sensitive nature of data collected by HAR systems such as location, daily routines and biometric information, it can be used for surveillance and potentially infringe on the user's privacy rights. It is also important that these data be handled securely because data breaches can have severe consequences for users such as identity theft or financial loss.

These data concerns also make it difficult for user adoption. With users becoming more aware and conscious of how their data are being used, without clear and user-friendly privacy policies, it would be difficult to promote user adoption of HAR systems.

8.1.2. Data Collection and Labelling

HAR systems require a large amount of labelled data, and this can be time-consuming. HAR models can only recognise activities contained in the dataset which they have been trained on. Usually such datasets are labelled manually in real-time, or once the activity is done [123]. This means that for a robust HAR model, a varied number of activities would be needed in the dataset, and the labelling of such a large dataset would be time-consuming.

There is also the complexity of real-world human activities where activities sometimes overlap or can happen simultaneously. Also, due to the diversity of human behaviour, a single activity can be performed different ways by different people. This is not often captured in datasets collected in a laboratory setting. A model trained on general data may perform poorly for certain individuals due to this [124]. This is a challenge that needs to be tackled in order to create robust HAR systems that can handle real-world complexities.

8.1.3. Accuracy and Reliability

While HAR models have achieved an overall high accuracy, there are still issues with differentiating activities with subtle differences. For the real-life deployment of these models, it is important to have highly accurate models, especially with health applications. There is also low accuracy in identifying complex activities in comparison to simple activities. When there are limited training data, there is low recognition accuracy.

When there is data drift from what the model was trained on, there is reduced accuracy, which means that the model would need to be re-trained on these new data. This makes it difficult to implement reliable HAR systems that have to deal with real-world data. It is also computationally expensive to have to re-train a new model every time there is data drift.

8.1.4. Ethical Considerations

HAR models trained on biased data will give discriminatory results. When the training data do not represent the diverse population they are intended to serve, the resulting HAR system becomes inherently biased against certain demographics, such as race, gender or age. This could further reinforce existing healthcare disparities [125] and systematic discrimination. HAR can also be used as a discriminatory tool against individuals based on their activities or behaviours such as in granting or denying access to services like insurance.

Most recent HAR systems use AI models. These AI models are black boxes, and so it is difficult to ascertain how they arrive at the results, and this makes it difficult to determine their clinical validity. When users can understand how these models work, they are more likely to adopt these HAR systems. This also raises the ethical question of accountability. It is not always clear who is to be held accountable in cases where AI-based systems are deployed.

There are also the ethical concerns of reduced human autonomy, where HAR systems could dictate what the user has to do, thus taking away the user's autonomy [126]. It is important that users can continue to maintain their autonomy and can choose to use or not use these HAR systems.

8.1.5. Standardisation

There is a lack of a widely adopted standard for HAR data collection, communication and model representation. There are also a diverse number of devices and sensors used for HAR which use different communication protocols. This poses a challenge for integrating HAR systems with the existing infrastructure.

Due to the absence of a standard, the HAR landscape is fragmented, and there are inconsistent practices. There is no standard to how activities can be defined, and this can lead to inconsistencies in data labelling and annotation. There is also no set protocol for data collection. This lack of standardisation needs to be addressed for interoperability and adoption of HAR systems.

8.2. Future Directions

There are many possibilities for HAR, but first research has to be carried out to address its challenges. This section highlights some areas where future research work in HAR can be focused on.

8.2.1. Explainable AI

Explainable AI techniques can be used to make AI models more transparent and trustworthy. They provide valuable insights into the decision-making process of these models. This will help users understand the reasoning behind the decisions of these AI systems and promote the adoption of HAR technologies. It would also address certain ethical concerns such as the possible of bias and discrimination. It would also allow researchers to identify and rectify points of errors in HAR systems. Some of the techniques for explainable AI include rule extraction, visualisation, attention mechanisms and feature importance. Applying explainable AI techniques for HAR systems would help in its adoption especially in healthcare.

8.2.2. Massive and Diverse Datasets

In order to develop more rounded models, there is a need to create massive and well-annotated datasets, which would cover a wide range of activities and also involve a wide demography of people. A diverse dataset would ensure that models trained on them are less likely to exhibit bias. Ensuring proper representation of the data of people from diverse demography ensures that the resultant HAR systems are unbiased. A massive dataset ensures that the model can handle variability especially in HAR where individuals perform activities differently. Annotations of such massive datasets can be expensive; however, with LLMs and other non-supervised methods, the effort needed is greatly reduced.

8.2.3. Multimodality

The combination of data from different sensors (movement data, physiological data, and context data) can create a more robust picture of human activity and highlight the fine-grained differences between similar activities. This is also important when recognizing complex activities which may have similarities in terms of body movement. Combining different modalities provides complementary information for a more robust system. This

can be used in healthcare applications of HAR, where an addition of physiological data can help differentiate activities like a fall from the patient lying down. Also, context data such as underlying medical conditions can also provide a richer understanding of certain activities.

Future research should focus on addressing the challenges involved in building robust multimodal HAR systems. Some of these challenges include the increased computational cost of developing multimodal systems, the availability of multimodal data, and the synchronisation of multimodal datasets.

8.2.4. Personalised HAR

The development of personalised HAR will account for individual user preferences and profiles. It will improve the accuracy of the HAR system by considering individual bio-mechanics, activity preferences or environmental factors. Personalised HAR systems can adapt to changes in user behaviour or environmental conditions. LLMs can be used in this instance to adapt existing models to the patterns of individual users or develop adaptive HAR models. Future research can be conducted on personalised and adaptive HAR systems using LLMs.

8.2.5. Standardisation

For interoperability and reproducibility of HAR research, it is important to have industry standards. This is especially important with the use of AI in HAR.These standards should cover data collection and representation, communication, and model representation. It should also provide a guideline that guides the ethical use of HAR systems.

Future research should be on developing metadata standards, clear annotation guidelines, API standards, data privacy guidelines, and ethical guidelines to improve the quality of HAR research and promote the adoption of HAR technologies. Creating a unified standard for HAR requires joint effort form government bodies, research institutes, companies, and industry experts. Sharing ideas and resources openly, through projects like open-source initiatives and industry consortium, can provide the needed momentum for developing an industry standard for HAR.

8.2.6. Privacy Policies and Security

It is important to implement detailed privacy policies such as a strong user consent mechanism, data anonymisation where possible, and a clear data usage and storage policy. Users should be clearly informed about what data will be collected, how they will be used, and how they will be stored. This will ensure the protection of user data. There should also be room for users to opt out of data collection for specific applications. Robust security measures must be in place to protect the user data collected and ensure there is no data breach.

9. Conclusions

HAR has gained popularity in different domains including sports, healthcare, robotics, human–computer interaction, security, surveillance, and entertainment. This review paper provides a comprehensive overview of the latest advancements in HAR with a focus on multimodality, DRL, and LLMs. It explores the diverse range of human activities and sensors used in Human Activity Recognition. Then, it surveys recent research on vision and non-vision based HAR. Different multimodal techniques have also been explored in this survey. Ablation studies performed in some of the papers reviewed showed that multimodal approach enhances the precision of HAR systems. Multimodal HAR systems can distinguish between similar activities with subtle differences and adapt to environmental conditions by combining multiple sensor data and even contextual information. They also need multimodal datasets, and these datasets with physiological data are also reviewed in this survey. The integration of DRL in HAR tasks represents a shift from traditional supervised learning methods. It combines the strength of deep learning and reinforcement learning and enables the HAR system to learn optimal policies through trial and error. This

technique has been applied to identify key features in complex datasets, determine optimal weights for the fusion of multimodal systems, and also improve the system's adaptability to new activities without extensively labelled data. LLMS have been introduced to the task of HAR for their ability to understand and predict activities based on textual and sensor data and also personalise their output based on individual preferences. They can be used in the annotation of large datasets and refining models based on contextual information and user preferences.

Despite the advances in the field of HAR, there still remain some challenges such as data privacy, data collection and annotation, accuracy and reliability, ethical considerations, and standardisation. In order to overcome some of these challenges, future research should focus on explainable AI, collecting and annotating massive and diverse datasets, multimodality, standardisation, and personalised HAR systems. This review highlights the potential of multimodal techniques, DRL, and LLMs in advancing HAR. Future research should continue to innovate and refine these techniques, paving the way for more user-centred, accurate and reliable HAR systems that will thrive in real-world scenarios.

Author Contributions: Conceptualisation, U.O., R.O. and A.S.A.; methodology, U.O. and R.O.; resources, R.O.; writing—original draft preparation, U.O.; writing—review and editing, R.O., A.S.A. and U.O.; supervision, R.O. and A.S.A.; project administration, R.O.; funding acquisition, R.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. Suthar, B.; Gadhia, B. Human Activity Recognition Using Deep Learning: A Survey. In *Proceedings of the Data Science and Intelligent Applications*; Kotecha, K., Piuri, V., Shah, H.N., Patel, R., Eds.; Springer: Singapore, 2021; pp. 217–223.
2. Diraco, G.; Rescio, G.; Siciliano, P.; Leone, A. Review on Human Action Recognition in Smart Living: Sensing Technology, Multimodality, Real-Time Processing, Interoperability, and Resource-Constrained Processing. *Sensors* **2023**, *23*, 5281. [CrossRef] [PubMed]
3. Hussain, Z.; Sheng, Q.Z.; Zhang, W.E. A review and categorization of techniques on device-free human activity recognition. *J. Netw. Comput. Appl.* **2020**, *167*, 102738. [CrossRef]
4. Nikpour, B.; Sinodinos, D.; Armanfard, N. Deep Reinforcement Learning in Human Activity Recognition: A Survey and Outlook. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *early access*.
5. Yilmaz, T.; Foster, R.; Hao, Y. Detecting vital signs with wearable wireless sensors. *Sensors* **2010**, *10*, 10837–10862. [CrossRef] [PubMed]
6. Wang, Y.; Cang, S.; Yu, H. A survey on wearable sensor modality centred human activity recognition in health care. *Expert Syst. Appl.* **2019**, *137*, 167–190. [CrossRef]
7. Manoj, T.; Thyagaraju, G. Ambient assisted living: A research on human activity recognition and vital health sign monitoring using deep learning approaches. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *8*, 531–540.
8. 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 Comput. Surv.* **2021**, *54*, 77. [CrossRef]
9. Li, S.; Yu, P.; Xu, Y.; Zhang, J. A Review of Research on Human Behavior Recognition Methods Based on Deep Learning. In Proceedings of the 2022 4th International Conference on Robotics and Computer Vision (ICRCV), Wuhan, China, 25–27 September 2022; pp. 108–112. [CrossRef]
10. Nweke, H.F.; Teh, Y.W.; Mujtaba, G.; Al-garadi, M.A. Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. *Inf. Fusion* **2019**, *46*, 147–170. [CrossRef]
11. Chen, L.; Liu, X.; Peng, L.; Wu, M. Deep learning based multimodal complex human activity recognition using wearable devices. *Appl. Intell.* **2021**, *51*, 4029–4042. [CrossRef]
12. Kumar, N.S.; Deepika, G.; Goutham, V.; Buvaneswari, B.; Reddy, R.V.K.; Angadi, S.; Dhanamjayulu, C.; Chinthaginjala, R.; Mohammad, F.; Khan, B. HARNet in deep learning approach—A systematic survey. *Sci. Rep.* **2024**, *14*, 8363. [CrossRef]
13. World Health Organization. Physical Activity. 2022. Available online: https://www.who.int/news-room/fact-sheets/detail/physical-activity (accessed on 25 June 2024).
14. Spacey, J. 110 Examples of Social Activities-Simplicable. 2023. Available online: https://simplicable.com/life/social-activities (accessed on 25 June 2024).
15. Hamidi Rad, M.; Aminian, K.; Gremeaux, V.; Massé, F.; Dadashi, F. Swimming phase-based performance evaluation using a single IMU in main swimming techniques. *Front. Bioeng. Biotechnol.* **2021**, *9*, 793302. [CrossRef]

16. Adarsh, A.; Kumar, B. Wireless medical sensor networks for smart e-healthcare. In *Intelligent Data Security Solutions for e-Health Applications*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 275–292.
17. Chung, S.; Jeong, C.Y.; Lim, J.M.; Lim, J.; Noh, K.J.; Kim, G.; Jeong, H. Real-world multimodal lifelog dataset for human behavior study. *ETRI J.* **2022**, *44*, 426–437. [CrossRef]
18. González, S.; Hsieh, W.T.; Chen, T.P.C. A Benchmark for Machine-Learning Based Non-Invasive Blood Pressure Estimation Using Photoplethysmogram. *Sci. Data* **2023**, *10*, 149. [CrossRef] [PubMed]
19. Hu, D.; Henry, C.; Bagchi, S. The Effect of Motion on PPG Heart Rate Sensors. In Proceedings of the 2020 50th Annual IEEE-IFIP International Conference on Dependable Systems and Networks-Supplemental Volume (DSN-S), Valencia, Spain, 29 June–2 July 2020; pp. 59–60. [CrossRef]
20. Wu, J.Y.; Ching, C.; Wang, H.M.D.; Liao, L.D. Emerging Wearable Biosensor Technologies for Stress Monitoring and Their Real-World Applications. *Biosensors* **2022**, *12*, 1097. [CrossRef] [PubMed]
21. Mekruksavanich, S.; Jitpattanakul, A. Efficient Recognition of Complex Human Activities Based on Smartwatch Sensors Using Deep Pyramidal Residual Network. In Proceedings of the 2023 15th International Conference on Information Technology and Electrical Engineering (ICITEE), Chiang Mai, Thailand, 26–27 October 2023; pp. 229–233. [CrossRef]
22. Hu, Z.; Lv, C. *Vision-Based Human Activity Recognition*; Springer: Berlin/Heidelberg, Germany, 2022.
23. Basavaiah, J.; Mohan Patil, C. Human Activity Detection and Action Recognition in Videos Using Convolutional Neural Networks. *J. Inf. Commun. Technol.* **2020**, *19*, 157–183. [CrossRef]
24. Andrade-Ambriz, Y.A.; Ledesma, S.; Ibarra-Manzano, M.A.; Oros-Flores, M.I.; Almanza-Ojeda, D.L. Human activity recognition using temporal convolutional neural network architecture. *Expert Syst. Appl.* **2022**, *191*, 116287. [CrossRef]
25. Parida, L.; Parida, B.R.; Mishra, M.R.; Jayasingh, S.K.; Samal, T.; Ray, S. A Novel Approach for Human Activity Recognition Using Vision Based Method. In Proceedings of the 2023 1st International Conference on Circuits, Power and Intelligent Systems (CCPIS), Bhubaneswar, India, 1–3 September 2023; pp. 1–5. [CrossRef]
26. Cheng, K.; Zhang, Y.; Cao, C.; Shi, L.; Cheng, J.; Lu, H. Decoupling GCN with DropGraph Module for Skeleton-Based Action Recognition. In *Proceedings of the Computer Vision–ECCV 2020*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 536–553.
27. Yan, S.; Xiong, Y.; Lin, D. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32. [CrossRef]
28. Liu, R.; Xu, C.; Zhang, T.; Zhao, W.; Cui, Z.; Yang, J. Si-GCN: Structure-induced Graph Convolution Network for Skeleton-based Action Recognition. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8. [CrossRef]
29. Jiang, M.; Dong, J.; Ma, D.; Sun, J.; He, J.; Lang, L. Inception Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. In Proceedings of the 2022 International Symposium on Control Engineering and Robotics (ISCER), Changsha, China, 18–20 February 2022; pp. 208–213. [CrossRef]
30. Lovanshi, M.; Tiwari, V.; Jain, S. 3D Skeleton-Based Human Motion Prediction Using Dynamic Multi-Scale Spatiotemporal Graph Recurrent Neural Networks. *IEEE Trans. Emerg. Top. Comput. Intell.* **2024**, *8*, 164–174. [CrossRef]
31. Minh Dang, L.; Min, K.; Wang, H.; Jalil Piran, M.; Hee Lee, C.; Moon, H. Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognit.* **2020**, *108*, 107561. [CrossRef]
32. Nikpour, B.; Armanfard, N. Joint Selection using Deep Reinforcement Learning for Skeleton-based Activity Recognition. In Proceedings of the 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Melbourne, VIC, Australia, 17–20 October 2021; pp. 1056–1061. [CrossRef]
33. Li, L.; Wang, M.; Ni, B.; Wang, H.; Yang, J.; Zhang, W. 3d human action representation learning via cross-view consistency pursuit. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 4741–4750.
34. Venugopal Rao, A.; Vishwakarma, S.K.; Kundu, S.; Tiwari, V. Hybrid HAR-CNN Model: A Hybrid Convolutional Neural Network Model for Predicting and Recognizing the Human Activity Recognition. *J. Mach. Comput.* **2024**, *4*, 419–430. [CrossRef]
35. Mehmood, K.; Imran, H.A.; Latif, U. HARDenseNet: A 1D DenseNet Inspired Convolutional Neural Network for Human Activity Recognition with Inertial Sensors. In Proceedings of the 2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 5–7 November 2020; pp. 1–6. [CrossRef]
36. Long, K.; Rao, C.; Zhang, X.; Ye, W.; Lou, X. FPGA Accelerator for Human Activity Recognition Based on Radar. *IEEE Trans. Circuits Syst. II Express Briefs* **2024**, *71*, 1441–1445. [CrossRef]
37. Deepan, P.; Santhosh Kumar, R.; Rajalingam, B.; Kumar Patra, P.S.; Ponnuthurai, S. An Intelligent Robust One Dimensional HAR-CNN Model for Human Activity Recognition using Wearable Sensor Data. In Proceedings of the 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 16–17 December 2022; pp. 1132–1138. [CrossRef]
38. Khan, Y.A.; Imaduddin, S.; Prabhat, R.; Wajid, M. Classification of Human Motion Activities using Mobile Phone Sensors and Deep Learning Model. In Proceedings of the 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 25–26 March 2022; Volume 1, pp. 1381–1386. [CrossRef]

39. Hernández, F.; Suárez, L.F.; Villamizar, J.; Altuve, M. Human Activity Recognition on Smartphones Using a Bidirectional LSTM Network. In Proceedings of the 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), Bucaramanga, Colombia, 24–26 April 2019; pp. 1–5. [CrossRef]
40. Mekruksavanich, S.; Jitpattanakul, A. Smartwatch-based Human Activity Recognition Using Hybrid LSTM Network. In Proceedings of the 2020 IEEE SENSORS, Virtual, 25–28 October 2020; pp. 1–4. [CrossRef]
41. Choudhury, N.A.; Soni, B. An Efficient and Lightweight Deep Learning Model for Human Activity Recognition on Raw Sensor Data in Uncontrolled Environment. *IEEE Sens. J.* **2023**, *23*, 25579–25586. [CrossRef]
42. Abdul, A.; Bhaskar Semwal, V.; Soni, V. Compressed Deep Learning Model For Human Activity Recognition. In Proceedings of the 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 24–25 February 2024; pp. 1–5. [CrossRef]
43. El-Adawi, E.; Essa, E.; Handosa, M.; Elmougy, S. Wireless body area sensor networks based human activity recognition using deep learning. *Sci. Rep.* **2024**, *14*, 2702. [CrossRef] [PubMed]
44. Choudhury, N.A.; Soni, B. Enhanced Complex Human Activity Recognition System: A Proficient Deep Learning Framework Exploiting Physiological Sensors and Feature Learning. *IEEE Sens. Lett.* **2023**, *7*, 6008104. [CrossRef]
45. Theodoridis, T. EMG Physical Action Data Set. UCI Machine Learning Repository. 2011. Available online: http://dx.doi.org/10.24432/C53W49 (accessed on 25 June 2024). [CrossRef]
46. Natani, A.; Sharma, A.; Peruma, T.; Sukhavasi, S. Deep Learning for Multi-Resident Activity Recognition in Ambient Sensing Smart Homes. In Proceedings of the 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 15–18 October 2019; pp. 340–341. [CrossRef]
47. Niu, H.; Nguyen, D.; Yonekawa, K.; Kurokawa, M.; Wada, S.; Yoshihara, K. Multi-source Transfer Learning for Human Activity Recognition in Smart Homes. In Proceedings of the 2020 IEEE International Conference on Smart Computing (SMARTCOMP), Bologna, Italy, 14–17 September 2020; pp. 274–277. [CrossRef]
48. Diallo, A.; Diallo, C. Human Activity Recognition in Smart Home using Deep Learning Models. In Proceedings of the 2021 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 15–17 December 2021; pp. 1511–1515. [CrossRef]
49. Jethanandani, M.; Perumal, T.; Chang, J.R.; Sharma, A.; Bao, Y. Multi-Resident Activity Recognition using Multi-Label Classification in Ambient Sensing Smart Homes. In Proceedings of the 2019 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), Yilan, Taiwan, 20–22 May 2019; pp. 1–2. [CrossRef]
50. Foerster, F.; Smeja, M.; Fahrenberg, J. Detection of posture and motion by accelerometry: A validation study in ambulatory monitoring. *Comput. Hum. Behav.* **1999**, *15*, 571–583. [CrossRef]
51. Agrawal, R. Fast Algorithms for Mining Association Rules. In Proceedings of the 20th VLDB Conference, Santiago de Chile, Chile, 12–15 September 1994.
52. Kulsoom, F.; Narejo, S.; Mehmood, Z.; Chaudhry, H.; Butt, A.; Bashir, A. A Review of Machine Learning-based Human Activity Recognition for Diverse Applications. *Neural Comput. Appl.* **2022**, *34*, 18289–18324. [CrossRef]
53. Kumar, P.; Suresh, S. Deep Learning Models for Recognizing the Simple Human Activities Using Smartphone Accelerometer Sensor. *IETE J. Res.* **2023**, *69*, 5148–5158. [CrossRef]
54. Ali, G.Q.; Al-Libawy, H. Time-Series Deep-Learning Classifier for Human Activity Recognition Based on Smartphone Built-in Sensors. *J. Phys. Conf. Ser.* **2021**, *1973*, 012127. [CrossRef]
55. Verma, U.; Tyagi, P.; Aneja, M.K. Multi-head CNN-based activity recognition and its application on chest-mounted sensor-belt. *Eng. Res. Express* **2024**, *6*, 025210. [CrossRef]
56. Rashid, N.; Nemati, E.; Ahmed, M.Y.; Kuang, J.; Gao, J.A. MM-HAR: Multi-Modal Human Activity Recognition Using Consumer Smartwatch and Earbuds. In Proceedings of the 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Orlando, FL, USA, 15–19 July 2024; pp. 1–4. [CrossRef]
57. Lin, F.; Wang, Z.; Zhao, H.; Qiu, S.; Shi, X.; Wu, L.; Gravina, R.; Fortino, G. Adaptive Multi-Modal Fusion Framework for Activity Monitoring of People With Mobility Disability. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 4314–4324. [CrossRef]
58. Bharti, P.; De, D.; Chellappan, S.; Das, S.K. HuMAn: Complex Activity Recognition with Multi-Modal Multi-Positional Body Sensing. *IEEE Trans. Mob. Comput.* **2019**, *18*, 857–870. [CrossRef]
59. Simonyan, K.; Zisserman, A. Two-Stream Convolutional Networks for Action Recognition in Videos. In *Proceedings of the Advances in Neural Information Processing Systems*; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2014; Volume 27.
60. Shi, L.; Zhang, Y.; Cheng, J.; Lu, H. Two-Stream Adaptive Graph Convolutional Networks for Skeleton-Based Action Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
61. Kumrai, T.; Korpela, J.; Maekawa, T.; Yu, Y.; Kanai, R. Human Activity Recognition with Deep Reinforcement Learning using the Camera of a Mobile Robot. In Proceedings of the 2020 IEEE International Conference on Pervasive Computing and Communications (PerCom), Austin, TX, USA, 23–27 March 2020; pp. 1–10. [CrossRef]
62. Shi, H.; Hou, Z.; Liang, J.; Lin, E.; Zhong, Z. DSFNet: A Distributed Sensors Fusion Network for Action Recognition. *IEEE Sens. J.* **2023**, *23*, 839–848. [CrossRef]

63. Mekruksavanich, S.; Promsakon, C.; Jitpattanakul, A. Location-based Daily Human Activity Recognition using Hybrid Deep Learning Network. In Proceedings of the 2021 18th International Joint Conference on Computer Science and Software Engineering (JCSSE), Virtual, 30 June–3 July 2021; pp. 1–5. [CrossRef]
64. Hnoohom, N.; Maitrichit, N.; Mekruksavanich, S.; Jitpattanakul, A. Deep Learning Approaches for Unobtrusive Human Activity Recognition using Insole-based and Smartwatch Sensors. In Proceedings of the 2022 3rd International Conference on Big Data Analytics and Practices (IBDAP), Bangkok, Thailand, 1–2 September 2022; pp. 1–5. [CrossRef]
65. Pham, C.; Nguyen-Thai, S.; Tran-Quang, H.; Tran, S.; Vu, H.; Tran, T.H.; Le, T.L. SensCapsNet: Deep Neural Network for Non-Obtrusive Sensing Based Human Activity Recognition. *IEEE Access* **2020**, *8*, 86934–86946. [CrossRef]
66. Zhang, L.; Yu, J.; Gao, Z.; Ni, Q. A multi-channel hybrid deep learning framework for multi-sensor fusion enabled human activity recognition. *Alex. Eng. J.* **2024**, *91*, 472–485. [CrossRef]
67. Das, A.; Sil, P.; Singh, P.K.; Bhateja, V.; Sarkar, R. MMHAR-EnsemNet: A Multi-Modal Human Activity Recognition Model. *IEEE Sens. J.* **2021**, *21*, 11569–11576. [CrossRef]
68. Zehra, N.; Azeem, S.H.; Farhan, M. Human Activity Recognition Through Ensemble Learning of Multiple Convolutional Neural Networks. In Proceedings of the 2021 55th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 24–26 March 2021; pp. 1–5. [CrossRef]
69. Guo, J.; Liu, Q.; Chen, E. A Deep Reinforcement Learning Method For Multimodal Data Fusion in Action Recognition. *IEEE Signal Process. Lett.* **2022**, *29*, 120–124. [CrossRef]
70. Muhoza, A.C.; Bergeret, E.; Brdys, C.; Gary, F. Multi-Position Human Activity Recognition using a Multi-Modal Deep Convolutional Neural Network. In Proceedings of the 2023 8th International Conference on Smart and Sustainable Technologies (SpliTech), Bol, Croatia, 20–23 June 2023; pp. 1–5. [CrossRef]
71. Shoaib, M.; Bosch, S.; Incel, O.D.; Scholten, H.; Havinga, P.J. Fusion of smartphone motion sensors for physical activity recognition. *Sensors* **2014**, *14*, 10146–10176. [CrossRef]
72. Banos, O.; Garcia, R.; Holgado-Terriza, J.A.; Damas, M.; Pomares, H.; Rojas, I.; Saez, A.; Villalonga, C. mHealthDroid: A novel framework for agile development of mobile health applications. In Proceedings of the Ambient Assisted Living and Daily Activities: 6th International Work-Conference, IWAAL 2014, Belfast, UK, 2–5 December 2014; Proceedings 6; Springer: Berlin/Heidelberg, Germany, 2014; pp. 91–98.
73. Chao, X.; Hou, Z.; Mo, Y. CZU-MHAD: A Multimodal Dataset for Human Action Recognition Utilizing a Depth Camera and 10 Wearable Inertial Sensors. *IEEE Sens. J.* **2022**, *22*, 7034–7042. [CrossRef]
74. Peng, L.; Chen, L.; Wu, X.; Guo, H.; Chen, G. Hierarchical Complex Activity Representation and Recognition Using Topic Model and Classifier Level Fusion. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 1369–1379. [CrossRef]
75. Óscar D. Lara.; Pérez, A.J.; Labrador, M.A.; Posada, J.D. Centinela: A human activity recognition system based on acceleration and vital sign data. *Pervasive Mob. Comput.* **2012**, *8*, 717–729. [CrossRef]
76. Yao, S.; Zhao, Y.; Shao, H.; Liu, D.; Liu, S.; Hao, Y.; Piao, A.; Hu, S.; Lu, S.; Abdelzaher, T.F. Sadeepsense: Self-attention deep learning framework for heterogeneous on-device sensors in internet of things applications. In Proceedings of the IEEE INFOCOM 2019-IEEE Conference on Computer Communications, Paris, France, 29 April–2 May 2019; pp. 1243–1251.
77. Yao, S.; Hu, S.; Zhao, Y.; Zhang, A.; Abdelzaher, T.F. DeepSense: A Unified Deep Learning Framework for Time-Series Mobile Sensing Data Processing. *CoRR* **2016**. Available online: http://arxiv.org/abs/1611.01942 (accessed on 28 July 2024).
78. Münzner, S.; Schmidt, P.; Reiss, A.; Hanselmann, M.; Stiefelhagen, R.; Dürichen, R. CNN-based sensor fusion techniques for multimodal human activity recognition. In Proceedings of the 2017 ACM International Symposium on Wearable Computers, New York, NY, USA, 11–15 September 2017; ISWC '17, pp. 158–165. [CrossRef]
79. Ordóñez, F.J.; Roggen, D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* **2016**, *16*, 115. [CrossRef]
80. Mahmud, T.; Akash, S.S.; Fattah, S.A.; Zhu, W.P.; Ahmad, M.O. Human Activity Recognition From Multi-modal Wearable Sensor Data Using Deep Multi-stage LSTM Architecture Based on Temporal Feature Aggregation. In Proceedings of the 2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS), Springfield, MA, USA, 9–12 August 2020; pp. 249–252. [CrossRef]
81. Jarchi, D.; Casson, A.J. Description of a Database Containing Wrist PPG Signals Recorded during Physical Exercise with Both Accelerometer and Gyroscope Measures of Motion. *Data* **2017**, *2*, 1. [CrossRef]
82. Dong, W.; Zhang, Z.; Tan, T. Attention-Aware Sampling via Deep Reinforcement Learning for Action Recognition. *AAAI* **2019**, *33*, 8247–8254. [CrossRef]
83. Wu, W.; He, D.; Tan, X.; Chen, S.; Wen, S. Multi-Agent Reinforcement Learning Based Frame Sampling for Effective Untrimmed Video Recognition. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6221–6230. [CrossRef]
84. Zhang, T.; Ma, C.; Sun, H.; Liang, Y.; Wang, B.; Fang, Y. Behavior recognition research based on reinforcement learning for dynamic key feature selection. In Proceedings of the 2022 International Symposium on Advances in Informatics, Electronics and Education (ISAIEE), Frankfurt, Germany, 17–19 December 2022; pp. 230–233. [CrossRef]
85. Zhang, W.; Li, W. A Deep Reinforcement Learning Based Human Behavior Prediction Approach in Smart Home Environments. In Proceedings of the 2019 International Conference on Robots & Intelligent System (ICRIS), Haikou, China, 15–16 June 2019; pp. 59–62. [CrossRef]

86. Raggioli, L.; Rossi, S. A Reinforcement-Learning Approach for Adaptive and Comfortable Assistive Robot Monitoring Behavior. In Proceedings of the 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), New Delhi, India, 14–18 October 2019; pp. 1–6. [CrossRef]
87. Ghadirzadeh, A.; Chen, X.; Yin, W.; Yi, Z.; Björkman, M.; Kragic, D. Human-Centered Collaborative Robots With Deep Reinforcement Learning. *IEEE Robot. Autom. Lett.* **2021**, *6*, 566–571. [CrossRef]
88. Sarker, I.H. LLM potentiality and awareness: A position paper from the perspective of trustworthy and responsible AI modeling. *Discov. Artif. Intell.* **2024**, *4*, 40. [CrossRef]
89. Gao, J.; Zhang, Y.; Chen, Y.; Zhang, T.; Tang, B.; Wang, X. Unsupervised Human Activity Recognition Via Large Language Models and Iterative Evolution. In Proceedings of the ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Republic of Korea, 14–19 April 2024; pp. 91–95. [CrossRef]
90. Kim, Y.; Xu, X.; McDuff, D.; Breazeal, C.; Park, H.W. Health-llm: Large language models for health prediction via wearable sensor data. *arXiv* **2024**, arXiv:2401.06866.
91. Ji, S.; Zheng, X.; Wu, C. HARGPT: Are LLMs Zero-Shot Human Activity Recognizers? *arXiv* **2024**, arXiv:2403.02727. Available online: http://arxiv.org/abs/2403.02727 (accessed on 20 July 2024).
92. Xu, H.; Zhou, P.; Tan, R.; Li, M.; Shen, G. LIMU-BERT: Unleashing the Potential of Unlabeled Data for IMU Sensing Applications. In Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, New York, NY, USA, 6–9 July 2021; SenSys '21, pp. 220–233. [CrossRef]
93. Imran, S.A.; Khan, M.N.H.; Biswas, S.; Islam, B. LLaSA: Large Multimodal Agent for Human Activity Analysis Through Wearable Sensors. *arXiv* **2024**, arXiv:2406.14498. Available online: http://arxiv.org/abs/2406.14498 (accessed on 5 August 2024).
94. Fang, C.M.; Danry, V.; Whitmore, N.; Bao, A.; Hutchison, A.; Pierce, C.; Maes, P. PhysioLLM: Supporting Personalized Health Insights with Wearables and Large Language Models. *arXiv* **2024**, arXiv:2406.19283. Available online: http://arxiv.org/abs/2406.19283 (accessed on 5 August 2024).
95. Gorelick, L.; Blank, M.; Shechtman, E.; Irani, M.; Basri, R. Actions as Space-Time Shapes. *Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 2247–2253. [CrossRef]
96. Schuldt, C.; Laptev, I.; Caputo, B. Recognizing human actions: A local SVM approach. In Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK, 23–26 August 2004; ICPR 2004; Volume 3, pp. 32–36. [CrossRef]
97. Gaglio, S.; Re, G.L.; Morana, M. Human Activity Recognition Process Using 3-D Posture Data. *IEEE Trans. Hum.-Mach. Syst.* **2015**, *45*, 586–597. [CrossRef]
98. Koppula, H.S.; Gupta, R.; Saxena, A. Learning human activities and object affordances from rgb-d videos. *Int. J. Robot. Res.* **2013**, *32*, 951–970. [CrossRef]
99. Wang, J.; Liu, Z.; Wu, Y.; Yuan, J. Mining actionlet ensemble for action recognition with depth cameras. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1290–1297. [CrossRef]
100. Reddy, K.K.; Shah, M. Recognizing 50 human action categories of web videos. *Mach. Vis. Appl.* **2013**, *24*, 971–981. [CrossRef]
101. Müller, M.; Röder, T.; Clausen, M.; Eberhardt, B.; Krüger, B.; Weber, A. *Documentation Mocap Database HDM05*; Technical Report CG-2007-2; Universität Bonn: Bonn, Germany, 2007.
102. Shahroudy, A.; Liu, J.; Ng, T.T.; Wang, G. Ntu rgb+ d: A large scale dataset for 3d human activity analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1010–1019.
103. 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 Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 2684–2701. [CrossRef] [PubMed]
104. Ionescu, C.; Papava, D.; Olaru, V.; Sminchisescu, C. Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1325–1339. [CrossRef] [PubMed]
105. Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. Activity recognition using cell phone accelerometers. *ACM SigKDD Explor. Newsl.* **2011**, *12*, 74–82. [CrossRef]
106. Alemdar, H.; Ertan, H.; Incel, O.D.; Ersoy, C. ARAS human activity datasets in multiple homes with multiple residents. In Proceedings of the 2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, Venice, Italy, 5–8 May 2013; pp. 232–235.
107. Roggen, D.; Calatroni, A.; Nguyen-Dinh, L.V.; Chavarriaga, R.; Sagha, H. OPPORTUNITY Activity Recognition. UCI Machine Learning Repository. 2012. Available online: http://dx.doi.org/10.24432/C5M027 (accessed on 24 June 2024). [CrossRef]
108. Reyes-Ortiz, J.; Anguita, D.; Ghio, A.; Oneto, L.; Parra, X. Human Activity Recognition Using Smartphones. UCI Machine Learning Repository. 2012. Available online: http://dx.doi.org/10.24432/C54S4K (accessed on 24 June 2024). [CrossRef]
109. Reiss, A.; Stricker, D. Introducing a new benchmarked dataset for activity monitoring. In Proceedings of the 2012 16th International Symposium on Wearable Computers, Newcastle, UK, 18–22 June 2012; pp. 108–109.
110. Banos, O.; Villalonga, C.; Garcia, R.; Saez, A.; Damas, M.; Holgado-Terriza, J.A.; Lee, S.; Pomares, H.; Rojas, I. Design, implementation and validation of a novel open framework for agile development of mobile health applications. *BioMed. Eng. OnLine* **2015**, *14*, S6. [CrossRef]
111. Reiss, A.; Indlekofer, I.; Schmidt, P. PPG-DaLiA. UCI Machine Learning Repository. 2019. Available online: http://dx.doi.org/10.24432/C53890 (accessed on 25 July 2024). [CrossRef]

112. Javeed, M.; Jalal, A. Deep Activity Recognition based on Patterns Discovery for Healthcare Monitoring. In Proceedings of the 2023 4th International Conference on Advancements in Computational Sciences (ICACS), Lahore, Pakistan, 20–22 February 2023; pp. 1–6. [CrossRef]
113. Elkahlout, M.; Abu-Saqer, M.M.; Aldaour, A.F.; Issa, A.; Debeljak, M. IoT-Based Healthcare and Monitoring Systems for the Elderly: A Literature Survey Study. In Proceedings of the 2020 International Conference on Assistive and Rehabilitation Technologies (iCareTech), Gaza, Palestine, 28–29 August 2020; pp. 92–96. [CrossRef]
114. Kalita, S.; Karmakar, A.; Hazarika, S.M. Human Fall Detection during Activities of Daily Living using Extended CORE9. In Proceedings of the 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), Gangtok, India, 25–28 February 2019; pp. 1–6. [CrossRef]
115. Thaduangta, B.; Choomjit, P.; Mongkolveswith, S.; Supasitthimethee, U.; Funilkul, S.; Triyason, T. Smart Healthcare: Basic health check-up and monitoring system for elderly. In Proceedings of the 2016 International Computer Science and Engineering Conference (ICSEC), Chiang Mai, Thailand, 14–17 December 2016; pp. 1–6. [CrossRef]
116. Pinge, A.; Jaisinghani, D.; Ghosh, S.; Challa, A.; Sen, S. mTanaaw: A System for Assessment and Analysis of Mental Health with Wearables. In Proceedings of the 2024 16th International Conference on COMmunication Systems & NETworkS (COMSNETS), Bengaluru, India, 3–7 January 2024; pp. 105–110. [CrossRef]
117. Aswar, S.; Yerrabandi, V.; Moncy, M.M.; Boda, S.R.; Jones, J.; Purkayastha, S. Generalizability of Human Activity Recognition Machine Learning Models from non-Parkinson's to Parkinson's Disease Patients. In Proceedings of the 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 24–27 July 2023; pp. 1–4. [CrossRef]
118. Mekruksavanich, S.; Jantawong, P.; Jitpattanakul, A. Enhancing Clinical Activity Recognition with Bidirectional RNNs and Accelerometer-ECG Fusion. In Proceedings of the 2024 21st International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Khon Kaen, Thailand, 27–30 May 2024; pp. 1–4. [CrossRef]
119. Verma, H.; Paul, D.; Bathula, S.R.; Sinha, S.; Kumar, S. Human Activity Recognition with Wearable Biomedical Sensors in Cyber Physical Systems. In Proceedings of the 2018 15th IEEE India Council International Conference (INDICON), Coimbatore, India, 16–18 December 2018; pp. 1–6. [CrossRef]
120. Hamido, M.; Mosallam, K.; Diab, O.; Amin, D.; Atia, A. A Framework for Human Activity Recognition Application for Therapeutic Purposes. In Proceedings of the 2023 Intelligent Methods, Systems, and Applications (IMSA), Giza, Egypt, 15–16 July 2023; pp. 130–135. [CrossRef]
121. Jin, F.; Zou, M.; Peng, X.; Lei, H.; Ren, Y. Deep Learning-Enhanced Internet of Things for Activity Recognition in Post-Stroke Rehabilitation. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 3851–3859. [CrossRef]
122. Yan, H.; Hu, B.; Chen, G.; Zhengyuan, E. Real-Time Continuous Human Rehabilitation Action Recognition using OpenPose and FCN. In Proceedings of the 2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), Shenzhen, China, 6–8 March 2020; pp. 239–242. [CrossRef]
123. Mohamed, A.; Lejarza, F.; Cahail, S.; Claudel, C.; Thomaz, E. HAR-GCNN: Deep graph CNNs for human activity recognition from highly unlabeled mobile sensor data. In Proceedings of the 2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops), Pisa, Italy, 21–25 March 2022; pp. 335–340.
124. Bursa, S.O.; Incel, O.D.; Isiklar Alptekin, G. Personalized and motion-based human activity recognition with transfer learning and compressed deep learning models. *Comput. Electr. Eng.* **2023**, *109*, 108777. [CrossRef]
125. Umer, L.; Khan, M.H.; Ayaz, Y. Transforming Healthcare with Artificial Intelligence in Pakistan: A Comprehensive Overview. *Pak. Armed Forces Med. J.* **2023**, *73*, 955–963. [CrossRef]
126. Emdad, F.B.; Ho, S.M.; Ravuri, B.; Hussain, S. Towards a unified utilitarian ethics framework for healthcare artificial intelligence. *arXiv* **2023** arXiv:2309.14617.