Human Activity Recognition with Smartphone and Wearable Sensors using Deep Learning Techniques: A Review

E Ramanujam, Thinagaran Perumal, *Senior Member, IEEE* and S Padmavathi.

***Abstract*—Human Activity Recognition (HAR) is a field that infers human activities from raw time-series signals acquired through embedded sensors of smartphones and wearable devices. It has gained much attraction in various smart home environments, especially to continuously monitor human behaviors in ambient assisted living to provide elderly care and rehabilitation. The system follows various operation modules such as data acquisition, pre-processing to eliminate noise and distortions, feature extraction, feature selection, and classification. Recently, various state-of-the-art techniques have proposed feature extraction and selection techniques classified using traditional Machine learning classifiers. However, most of the techniques use rustic feature extraction processes that are incapable of recognizing complex activities. With the emergence and advancement of high computational resources, Deep Learning techniques are widely used in various HAR systems to retrieve features and classification efficiently. Thus, this review paper focuses on providing profound concise of deep learning techniques used in smartphone and wearable sensor-based recognition systems. The proposed techniques are categorized into conventional and hybrid deep learning models described with its uniqueness, merits, and limitations. The paper also discusses various benchmark datasets used in existing techniques. Finally, the paper lists certain challenges and issues that require future research and improvements.**

***Index Terms*—Activity recognition, Machine learning, wearable sensors, smart phones, context-aware, deep learning.**

I. INTRODUCTION

In recent years, the market surrounding wearable technology has grown exponentially, with good reason due to the forefront of the Internet of Things (IoT). This shift transforms the healthcare industry from treatment to prevention. There is a range of devices in practice ranging from smartphones to smart wristbands which are being utilized for industry and research. The data from these sensors are collected, stored and analyzed and having key decisions made upon it. The major purpose of this device is to classify the activity of a person at a given time thereby to provide assistance and guidance to a person termed as Human Activity Recognition (HAR). The objective of HAR is to monitor simple, complex, and postural activities of human behaviors in the area of Ambient Assisted Living (AAL), sports injury detection, well-being management, medical diagnosis and especially in elderly care [1]. These recognition activities are mostly used to identify fall and Activities of Daily Living (ADLs) with elders, provide long-term care, well-being, and maintain their independent quality of life who lives alone [2]. The HAR systems utilize Machine learning or deep learning model to recognize the activities by exploiting the signals received from wearable, environmental sensors, and vision systems [3]. Environmental sensors require installation in the home that is quite costly [4]. Vision systems use a camera for recognition and are perceived as intrusive devices [5]. The other solution is wearable devices that attract the researcher mainly due to their wide usage. The wearable devices, especially fitness trackers and smartwatches, are mostly used for recognition as it has various inbuilt/ integrated sensors such as Accelerometer, Gyroscope, and Orientation sensors [3] available at preferable cost. Besides, modern smartphones are used as an alternative to wearable devices for activity recognition as they are cost-effective, unobtrusive, highly equipped with embedded-sensors like wearable devices, and facilitated mostly in real-time applications [6].

Machine learning-based solutions have successfully achieved HAR recognition activities during the last decade, but they face certain challenges in certain cases like [7-9].

i. Limited amount of training data for multiclass activity recognition
ii. Recognizing arbitrary and complex features using handcrafted/ manual features
iii. Recognizing confusing activities such as searching under the bed with falls and walking with climbing upstairs.
iv. Dynamic mode of activities/ motion by a subject in different simulation

Further, the machine learning-based solutions completely depend on pre-processed data from raw signals, valuable and remarkable features that can improve classification algorithms' performance. These challenges can be easily resolved or overcome with deep learning models [7]. The deep learning models have recently shown improved and promising

E.Ramanujam is with the Information Technology Department, Thiagarajar College of Engineering, Madurai, Tamil Nadu 625015 India. (e-mail: erit@tce.edu)

Thinagaran Perumal is with Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia., (e-mail: thinagaran@upm.edu.my).

S.Padmavathi is with the Information Technology Department, Thiagarajar College of Engineering, Madurai, Tamil Nadu 625015 India. (e-mail: spmcse@tce.edu).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3069927, IEEE Sensors Journal

performance on various benchmark datasets utilized by machine learning-based solutions. It can reduce the works on the data pre-processing and feature extraction phase. Further, it can improve the generalization performance and robustness of the deep learning model. Moreover, certain advantages of deep learning over machine learning approaches are

i. Reduces the dependency of expert knowledge for feature engineering process
ii. Accurate recognition on Temporal dynamics of features
iii. A Shortened testing time for image based recognition
iv. High performance even for weakly labeled data

All of these unique features make deep learning techniques superior in HAR based solutions. To overview various deep learning models proposed for HAR in real-time and benchmark datasets, Electronic databases such as PubMed, PsycINFO, Web of Science, Scopus, and Google scholar were used to search and retrieve the published papers on HAR using deep learning techniques. Though there are certain research reviews available for HAR as in [1, 3-7], certain key contributions of the work compared with existing reviews are summarized as follows,

i. Exploration on benchmark datasets with the usage of sensors and devices for data collection to recognize the type of activity and the environment where the subjects performed the activities.
ii. Categorized the HAR methods in the view of design architecture with is characteristics.
iii. Compared the proposed HAR methods with pros and cons, experimental architecture and its feasibility on challenges and issues.
iv. Discusses challenges in this field and listed scope for future enhancements

The review article is structured as follows. Section 2 details the benchmark datasets and their characteristics; Section 3 categorizes the deep learning-based models. Section 4 discusses the challenges that require future research and scope for improvements. Finally, Section 5 concludes the paper.

II. Benchmark Datasets

Researchers have created various benchmark datasets to evaluate HAR activity to deep learning and machine learning-based solutions freely [10-11]. The dataset consists of motion signals collected through embedded sensors of devices from several parts of the participated subjects including head, shin, forearm, chest, upper arm, thigh, waist, and legs. The wearable devices are attached tightly to the above-said positions, smartphones are placed in trouser pockets or cloths, and smartwatches are tied in dominant hands. The devices contain different sensors, which include Accelerometer (A), Gyroscope (G), Orientation (O), Magnetometer (M), Object sensor (Obj), Temperature (T), and Ambient sensors (AM). In each dataset's, collected signals have a different characterization of their participated subjects concerning their age, height, weight, and other physical conditions. During the sensory data collection, the subjects are instructed to perform simple and complex activities. Walking, jumping, lying, running, jogging, climbing upstairs and down, cycling is simple. Cooking, washing clothes, and cleaning the kitchen are considered complex activities. The transition between the two activities is considered postural transitions such as sitting-to-standing and sitting-to-lying. The detailed descriptions of the benchmark datasets, along with its characterization, are shown in Table I.

TABLE I. LIST OF PUBLICLY AVAILABLE BENCHMARK DATASETS

Dataset name	Sensor (s) data	# of Activities		Type of Activity			# of Subjects	Sampling Frequency	Device(s) used	Device Position	Environment
		ADLs	Falls	Simple	Complex	Postural					
PAMAP2 [12]	T, A, G, O	12	✗	✓	✓	✗	9	100Hz	3 IMU units 1 Heart rate monitor	Wrist, chest & dominant side's ankle	Controlled Lab
HHAR [13]	A, G	6	✗	✓	✗	✗	9	Highest	4 Smartwatches & 8 Smartphone	Smartphone - Waist, Pouch	Out of Lab
MHEALTH [14-15]	A, G, M, ECG signals	12	✗	✓	✗	✗	10	50Hz	Shimmer2 wearable sensors	right wrist, left ankle, and chest	Out of Lab
UCI-HAR [16]	A, G	6	✗	✓	✗	✗	30	50Hz	Smartphone	Left belt and No specific position	Controlled
OPPORTUNITY [17]	A, G, M, Obj, AM	6	✗	✓	✓	✗	4	-	Body-worn, Object and Ambient Sensors	upper body, hip,leg, and users shoes	Home Environment (Hall and Kitchen)
WISDM [18]	A, G	18	✗	✓	✓	✗	51	20Hz	Smartphone Smartwatch	Right pant packet + smartwatch on their dominant hand	Controlled / Out of Lab
UniMiB-SHAR [19]	A	9	8	✓	✗	✗	30	50Hz	Smartphone	Left & Right Trouser pockets	Controlled
MobiAct	A, G	12	4	✓	✗	✗	66	100Hz	Smartphone	Trouser	Controlled

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3069927, IEEE Sensors Journal

[20]										pockets	
HAPT [21]	A	12	×	✓	×	✓	30	50Hz	Smartphone	Waist	Controlled
Cooking dataset [22]	Five IMU	16	×	✓	✓	×	7	110Hz	Wearable device	Various positions	Controlled
Dataset created for Chen et al.[23]	A	5	1	✓	×	×	100	100Hz	Smartphone	Trouser pocket & waist	Controlled
Dataset created for Zhu et al., [24]	A, G, M	7	×	✓	×	×	100	50Hz	Smartphone	Handheld, Trouser Pocket, and Backpack	Controlled
Dataset was created for Khan et al., 2018 [25].	A	8	×	✓	×	×	15	50Hz	2 Smartphones 1 smartwatch	Tight pouch around the waist, smartwatch in the dominant hand	Controlled
Position aware activity recognition (HAR) [26]	A	8	×	✓	×	×	15	50Hz	Wearable device	Chest, head, shin, thigh, upper arm & waist	Out of Lab
Sequential Weakly Labeled Multi-Activity Dataset (SWLM) [27]	A	3	×	✓	×	×	10	50Hz	iPhone	Right wrist	Controlled
w-HAR [28]	A, Stretch Sensor	7	×	✓	×	✓	22	250Hz	Wearable	Right ankle	Controlled

G- GYROSCOPE, O- ORIENTATION, A - ACCELEROMETER, , M- MAGNETOMETER, , AM - AMBIENT SENSOR, T- TEMPERATURE, OB - OBJECT SENSOR.

III. DEEP LEARNING MODELS

The advantages of machine learning algorithms have attracted the researcher's interest in applying deep learning models in Human Activity Recognition (HAR). On categorizing the proposed models, certain techniques have utilized the collected raw sensory signals (time series signals), and others have modified the signals to the imaging of signals such as frequency or virtual images for activity recognition, as shown in Figure 1.

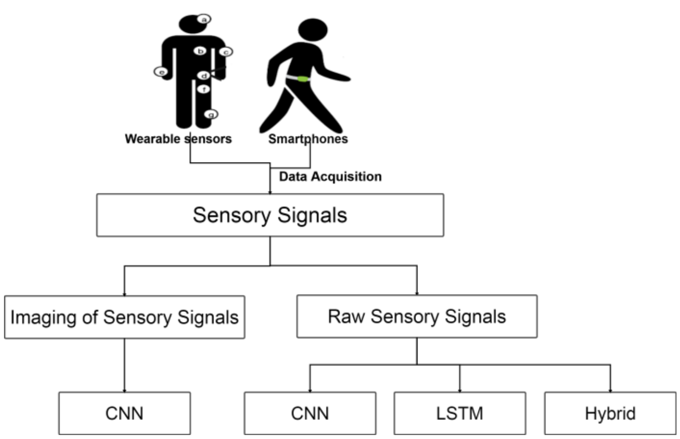

Fig. 1. Categorization of proposed Deep Learning models

A. *Imaging of Sensory Signals*

Convolutional Neural Network (CNN) is highly proficient in extracting efficient features and classifying large scale Images [29]. In addition, the representation of time series signals as visual cues has widespread attention among the researchers and has applied to classify time series signals using CNN. It is very effective. This motivates various HAR systems to encode the raw time-series signals into visual cues such as virtual images and frequency images to classify using CNN, as shown in Figure 2.

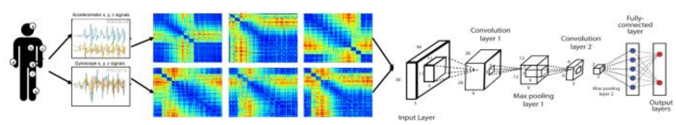

Figure 2. The architecture of classifying virtual images using CNN

With specific to a smartphone-based recognition system, Alemayoh et al. in [30] have encoded the raw sensory signals collected using a specially designed iOS application into a 14X60 virtual image. The resultant image data has been classified using CNN of architecture (1C - 1P - 2C - 2P - 1DL) to classify 8 different simple activities. In recent days, various multi-modal sensor-based activity recognition systems have proposed where multiple sensors are attached to various positions of the subject. To address the multimodal sensor activity recognition system, Lawal et al. [31] fuse various time series (motion) signals generated by multiple wearable sensors into a single frequency image for classification. The frequency images are served as input to two-stream CNN model {1C- 1P - 1D - 2C - 2P - 2D - 3C - 3P - 3D }⊕ {1C- 1P - 1D - 2C - 2P - 2D - 3C - 3P - 3D} - 1F - 1DL - 2DL - 1D - 1D. To deal with the heterogeneous sensor data from multiple sensors, Qin et al. of [32] encoded the time-series signal into a two-channel image integrated with the residual fusion network. In this proposed approach, the layers of Lawal et al., 2020, are modified with residual fusion networks instead of Convolution block 3. Extensive experimentation has been performed using HHAR and MHEALTH dataset for performance evaluation.

B. *Raw Sensory Signals*

With an advantage of local dependency and scaling invariance, the deep learning algorithms are highly capable and efficient in handling time-series signals to extract features and classification. This has recently attracted more researchers in applying deep learning techniques such as CNN, LSTM, and hybrid models towards more human activity recognition applications.

1) Convolutional Neural Network (CNN) Model

The Convolutional Neural Network (CNN) models are competent enough to classify Images and to predict raw time series signals. With an advantage of this, various researchers have integrated CNN models to classify raw inertial sensor signals for human activity recognition. Feature extraction by CNN has certain advantages over traditional shallow learning technique in the field of HAR such as local dependency, scale invariance, and it also captures all possible complex non-linear interactions among the features [33].

Chen and Xue [23] have proposed a CNN model to recognize the simple activities through tri-axial accelerometer signals of a smartphone. The CNN model has been constructed, and the convolution kernels are modified accordingly to process tri-axial accelerometer signals. Ronao and Cho [34] has proposed a novel CNN architecture to extract complex features using an exploited (1x9-1x14) Convolutional layer with a low pooling size of (1x2-1x3). The proposed approach has experimented with the constructed CNN with raw signals and temporal characteristics of Fast Fourier Transformed signals.

Ignatov [35] has proposed a CNN for the local feature extraction along with statistical features such as mean, variance, sum of absolute values and histogram of each input to preserve information about global features of raw signals. This research work also investigates the impact of signal length by varying the sliding window upto 1s on the performance. Avilés-Cruz et al. [36] has developed a deep learning framework that constructs the CNN model to classify and analyze exclusive user-dependent activity recognition. In this approach, three CNN (fine, medium, and coarse) models are constructed parallel for local feature extraction and fused into a single classification model. The CNN models generally require high computing resources for feature extraction and classification as it uses more number of filters and mappings. To reduce the cost of energy consumption and hardware facilities, [37] has designed a CNN model for local feature extraction from the three-axis accelerometer signal of a smartphone. Experimentation has been carried out using the PAMAP2 dataset and HAR dataset to compare with LSTM, BLSTM, MLP, and SVM.

To reduce the memory and computational cost of traditional CNN, [38] has utilized Lego filters instead of Convolutional filters. The proposed lightweight model performs and scales better in experimentation as it doesn't require any special network structure and computational resources. As like [38], Cheng et al [40] have introduced a novel computation efficient CNN using conditionally parameterized convolution for real-time HAR on mobile and wearable devices. Experimentation has been performed to show the efficiency of developed network with large scale capacity with baseline model.

In addition to high computing resources, CNN is more likely to a cold-start problem. To overcome this issue, the research work [41] has evaluated various CNN pre-trained models using real-time complex activities. In addition, all the pre-trained models were assessed on different hyper-parameters to identify the best CNN models for HAR. Then the identified candidate model has been implemented as a feature extractor model to evaluate a large-scale real-world dataset.

To develop a more efficient and real-time HAR system, Nutter et al. [42] presented a deep learning framework that recognizes real-time activities via the smartphone itself. The traditional hand-crafted features of IMU data are dimensionally reduced through Principal Component Analysis (PCA) to 100 features. This minimizes the model size on training and testing features and reduces the computational overhead to run on an embedded processer, thereby preserving a smartphone's battery life. The proposed model introduced a hybrid bi-class SVM instead of a dense layer to improve the classification performance further. The summary of the proposed CNN based HAR systems are shown in Table II.

TABLE II. A SUMMARY OF CONVOLUTIONAL NEURAL NETWORK-BASED HUMAN ACTIVITY RECOGNITION SYSTEM

Authors	Proposed CNN Architecture	Dataset(s) used	Performance Metrics (Avg. Accuracy)	Optimization on Hyper parameters	Issues handled/ Improvements	Scope for Future enhancements
Chen et al., 2015 [23]	C1-P1-C2-P2-C3-P3-FCNN	Self-recorded data by the user	93.8%	Width of Convolution kernel, Epochs	Doesn't require any domain-specific knowledge as like traditional shallow learning techniques	Deal with complex and postural transitions.
Ranao and Chao, 2016 [34]	C_i - P_i - F-FCNN - SL	UCI-HAR	94.79% on raw signals 95.75% on Temporal FFT	Filter size Error rate No of Epochs	Benefited with temporal correlation and low pooling size in construction of CNN layers	To integrate frequency domain convolution with time-domain convolution for better performance.
Ignatov, 2018 [35]	C1 (196 filters) – P1 (1x4) – F(statistical Features) – FCNN – SL	WISDM	90.42%	Time segments of varying size	Investigates the performance on varying time segments/ window size	To emphasize cross dataset experiment on platform-independent architecture not only on user-context, but to devices with different accelermoter calibrations
Avilés-Cruz	{CNN1 { C1 -	UCI-HAR	100%	Filter size,	Evaluated using cross-dataset	To identify associated

et al, 2019 [36]	P1- C2- P2- C3- P3- C4- P4 } ⊕ CNN2 {C1- P1- } ⊕ CNN3 {C1 - P1} } - F - FCNN			Pooling size, Epochs	(Subject Independent Testing).	heath issues by recognizing complex activities of a human subject
Wan et al., 2019 [37]	C1-P1-C2-P2-C3-P3-FCNN	UCI HAR	92.71%	No	Reduces the cost of energy consumption and hardware facilities	The network can be still optimized on the best hyperparameters
		PAMAP2	91.00%			
Tang et al., 2019 [38,39]	Traditional CNN with Lego bricks as lower-dimensional filters	UCI HAR	96.27%	No	Reduces memory and computational cost over traditional CNN on compressed datasets	The architecture can be generalized to process heterogeneous data.
		PAMAP2	91.40%			
		UNIMIB-SHAR	74.46%			
		Opportunity	86.10 %			
		WISDM	97.51%			
Cheng et al, 2020 [40]	Set of linear combinations (experts) + C1 + BN + ReLu	WISDM	99.60%	Experts and Activation functions	Reduces memory and computational cost over traditional CNN on large scale dataset	Still can be optimized using hyperparameters of traditional CNNs.
		PAMAP2	94.01%			
		UNIMIB-SHAR	77.31%			
		OPPORTUNITY	81.18%			
Cruciani et al, 2020 [41]	3 * (1D C1 - P1) -F - FCNN - SL Filters f = 32 kernel k = 2	UCI-HAR	91.98%	Convolutional layers Kernel size filters	Pre-trained models are assessed on various hyperparameters, thus obtaining a best pre-trained candidate model for testing.	Activation functions can be optimized as one of the hyperparameters for best performance. Reusable pre-trained models can be developed
		DCASE (Audio-based HAR)	92.30%			
Nutter et al, 2018 [42]	BN - C1 -AP1 - D1 - F - FCNN - Bi class SVM pipeline	UCI	F1 score 99%	No	Handcrafted 561 features are dimensionally reduced to 100 features through Principal Component Analysis (PCA).	The accuracy rate is less when compared with F1-score. Can be optimized using Hyperparameters of CNN
		UCF	F1 Score 97%			
Zhu et al., 2019 [24]	An ensemble of {{C1-P1-C2-P2-C3-P3-C4-P4-C5-P6-C7-P7}⊕ {C1- P1- C2- P2 } FCNN - SL Using Weighted Voting	Self Recorded data	96.11%	No	Recognizes the confusing activities and people's motion for different individuals (People specific model) with less training data.	To conduct experiments with large datasets under more placements of smartphones (cross domain platform/data)
Khan et al., 2018 [25]	Two models of C1 - C2 - FC- embedded with Kullback-Leibler divergence + SL	Self Recorded data	76% on transfer from Smartphone to Smartwatch 80% on transfer from Smartwatch to Smartphone	No	Proposed a transfer learning mechanism that adapts automatically to a cross domain platform (from smart watch to smart phone and vice-versa) with no or limited labeled data.	To develop a similar automatic adaptation based approach for practical in-the-wild activity recognition
		Heterogeneity Activity Recognition (HHAR)	78.75% on Phone to watch 76.74 Watch to Phone			
		Position aware activity recognition (HAR)	72.24% on Phone to Watch 74.71% on Watch to Phone			
Xiao et al, 2020[43]	SMPL + U*C - S*FCNN - unsupervised penalty function	AMASS dataset	87.46%	No	Proposed an online training and offline testing of real-world complex activities.	Can optimize the execution with C and S connected layers
		DIP dataset	89.08%			
		AMASS & DI	91.15%			
Almaslukh et al, 2018 [44]	C1 - P1- C2- P2- F - FCNN - D - FC - D - FC - D - FC	Position aware activity recognition (HAR)	84 to 88%	Learning rate, filter size, Pooling size, Dropout rate	Proposed time-domain statistical features based position-independent HAR system	To develop a more accurate position-independent model for HAR
Shojaedini et al, 2020 [45]	BN+(C+BN) + (C+BN) + (C+BN) + LSTM+ FC	WISDM	Improves Accuracy to 5% than traditional methods	No	Model saturates the fluctuation in Accuracy using residual network concept	To propose a more accurate model for complex human activities
C - Convolution Layer, P - Pooling Layer, F - Flattened Layer, SL - SoftMax Layer, D - Dropout Layer, FCNN - Fully Connected Neural Network, AP - Average Pooling Layer, FC - Fully Connected, BN-Batch Normalization						

The proposed CNN models are highly competent, efficient, and achieve higher accuracy on simple human activity recognition. However, more complex and confusing activities that any elder may perform in real-time, such as searching under a bed, tie a shoelace with fall. To reduce the accuracy rate fluctuations and improve the sensitivity and specificity rate of confusing HAR [24] presented an ensemble of CNN with varying layers and filters. The proposed model recognizes the confusing activities and people's dynamic motion for activity with less training data.

To investigate the human activity recognition system [25] has proposed Heterogeneous Deep CNN (HDCNN) a transductive transfer learning model. This HDCNN model constructs two-layer CNN with Kullback Leibler divergence to tune the properties of CNN. The proposed model assigns invariant distribution of weights to CNN layers, as long as the process of activities was being monitored. Experimentation has been carried out by training the model with Smartphone signals and tested with smartwatch signals and vice versa for activity recognition. Due to cross-domain adaptability, the proposed model shows optimal performance compared with conventional training and testing. In addition to classifying simple and repetitive daily living activities, [43] has proposed a deep learning model with an unsupervised penalty function to recognize the complex activity. In the proposed system, an AMASS simulated dataset has been used for offline training to enhance the variety and diversity. The dataset contains a rich set of human poses and virtual IMU data. During testing, the proposed system integrates a transfer learning mechanism to improve the performance by fine-tuning the partial neural network.

The proposed smartphone-based human activity recognition systems are mostly position-dependent where the mobiles are kept in a fixed position on the human subject during experimentation. For instance, smartphones fixed in vertical positions on the belt of a human subject, etc.

Few studies use handcrafted features that are less biased to smartphone position to tackle the problem of position-independent HAR system. However, the performances of these studies are relatively low and need further improvement for a reliable real-world HAR implementation. Thus, Alamslukh et al [44] has proposed a robust position-independent HAR system using a deep neural network model. Experimentation has been performed using a real-world HAR system including complex and simple activities to evaluate the performance of the proposed model.

Though the CNN models show better HAR performances, it still faces certain challenges in attaining the saturated accuracy. To reduce the accuracy rate research work [45] engaged a residual network as a substructure inside the CNN that jump over certain layers to reduce the vanishing gradient effects. Therefore, the proposed model achieves higher accuracy in classifying several dynamic activities.

2) Long-short Term Memory (LSTM) Model

The LSTM models are highly competent to predict the raw sequences of time series signals rather than classifying an image like the CNN model. "Spatial correlations" are exploited in CNN models to classify the image whereas the LSTM model process entire sequences of data through feedback connections to classify the time series data. With an advantageous effort of LSTM over the CNN model [34], the researchers have proposed certain techniques on LSTM based HAR models. Table III shows the summary of proposed LSTM based HAR system with its challenges addressed and scope for future enhancements.

Chung et al. [46] have performed a pilot study with 15 participants about the on-body sensor positioning system. This study creates a testbed for various simple activities using eight body-worn Inertial Measurement Units (IMU) sensors. Then, the LSTM model has been deployed to evaluate the performance in both real-world and controlled testbed environment. Further, an ensemble model has been integrated with the LSTM model to demonstrate the class probability of multi-sensor modality. The proposed system has certain challenges as it has created a testbed for simple activity recognition with fewer participants and thus not suitable for large scale applications.

To design a HAR with less computation power and reduced latency, Agarwal et al. [47] have proposed a lightweight deep learning model. The proposed model is suitable to deploy on edge devices as edge computing reduces communication latency traffic and can be applied for real case scenarios. However, this system doesn't evaluate its performance with more complex activities before being taken into real-time implementation on elderly patients. Rashid et al [48] have extended [47] and proposed an energy efficient and memory efficient adaptive CNN for low-power edge devices. Experimentation has been carried out with complex activities to validate the significance of the proposed system in terms of memory and energy efficient using opportunity and w-HAR dataset. Zhao et al [49], have proposed residual bi-directional LSTM architecture with an advantage to concatenate the forward state and backward state i.e. positive and negative time direction. The residual connection utilized between the stacked cells avoids the gradient vanishing problem.

The proposed systems [47-49] have relatively low performance on complex, dynamic, and postural human activity recognition. To solve this problem, Wang and Liu [50] have proposed a novel deep learning structure named Hierarchical deep Long-Short Term memory (H-LSTM). The sensory signals are pre-processed from noise and distortions, and then time-frequency-domain features are extracted by the H-LSTM model to improve the system's performance. To deal with raw time series data, Ashry et al, 2020 [51] have proposed a cascaded LSTM (CHARM) an online and offline model to address the need of large training dataset by utilizing novel hand-crafted features as described in [52-53] along with raw signals.

Mostly, the HAR system has traditionally been solved using local features extracted using heuristic methods. The traditional machine learning algorithms may easily throw into local minimum other than a globally optimal solution. The research work [54] has proposed a deep hybrid framework based on convolution operations of CNN integrated with LSTM recurrent units and classified using ELM to overcome this issue. The proposed framework has been evaluated on opportunity and outperforms non-recurrent neural network models.

The other major issue needs to be addressed by LSTM model in HAR is to analyze the weakly labeled sensor data. , Zhou et al. [55] have designed a semi-supervised learning framework using LSTM with Deep Q-Network an intelligent auto-labeling scheme to improve the performance on weakly labeled sensor data. Deep Q-Network better solves the inadequate labeled samples to improvise the performance of the system. The LSTM model in the proposed system recognizes the fine-grained patterns contextually extracted from sequential motion data.

TABLE III. A SUMMARY OF LONG-SHORT TERM MEMORY (LSTM) BASED HUMAN ACTIVITY RECOGNITION SYSTEM

Authors	Proposed LSTM Architecture	Dataset(s) used	Performance Metrics (Avg. Accuracy)	Optimization on Hyper parameters	Issues handled/ Improvements	Scope for future enhancements
Chung et al, 2018 [46]	Many-to-one LSTM architecture	Multimodal Sensor Test Bed	93.07% (Acc+Gyr + Mag) 93.48% (Hard voting on sensor fusion) 94.47% (soft voting on sensor fusion)	No	Ensemble model has been integrated with LSTM model to demonstrate the class probability of multi-sensor modality	A limited number of subjects involved in creating a test bed
Wang and Liu, 2020 [50]	The hierarchical arrangement of LSTM network - Softmax classifier	UCI - HAR	91.65%	No	Time-frequency domain-based Hierarchical deep LSTM network	Complex, postural transitions and confusing activities can be involved
		HHAR	99.15%			
Agarwal et al., 2020 [47]	RNN-LSTM - SL	WISDM	95.78%	No	Lightweight model suitable to be deployed on edge devices	To support multi-sensor fusion data
Sun et al, 2018 [54]	4 Layers of CNN - LSTM1 - LSTM2 - FCNN - ELM	Opportunity	90.6% F1- score	No	Induce temporal dynamics of features to shorten the model's runtime and solve sequential activity recognition.	Transfer learning can be integrated to improve the adaptive capability
Zhou et al, 2020 [55]	Deep Q Network (DQN)	UniMiB SHAR	F1 score - 79%	No	Designed to recognize weak labeled sensor data	More significant patterns /weak labeled data can be detected to improve the performance
		Position-aware activity recognition with wearable devices	F1-score of 97% on positioning the devices on Chest and Shin			
Rashid et al, 2021 [48]	Adaptive CNN	Opportunity	F1-Score – 91.57%	No	Implemented Real-time on Edge devices	Evaluation can be done with Complex activities
		w-HAR	F1-Score – 97.64%			
Zhao et al, 2018 [49]	Residual Bi-dir LSTM	UCI-HAR	93.6%	Convolutional Layers, dropout, units/layer	Best hyperparameters are chosen by Grid search	Class imbalance makes difference in performance metrics
		Opportunity	90.5%			
RNN - Recurrent Neural Network, ELM - Extreme Learning Machine, SL - Softmax Layer, FCNN - Fully Connected Neural Network						

3) *Hybrid Models*

The hybrid models are specifically proposed to better perform than traditional deep learning models in certain challenges such as limited training, confusing activities, and localization of sensors. The summary of various hybrid models proposed for HAR system with its novelty, issues addressed and scope for future enhancements are shown in Table IV. Xia et al. [56] have proposed a deep hybrid model by integrating LSTM models connected with Convolutional Layers and Global Pooling layers (GAP). The GAP is used instead of a traditional fully connected layer. In addition, the Batch Normalization is introduced after GAP to speed up the convergence of the proposed system.

Most of the HAR systems are proposed for simple activity and has certain challenges in recognition of complex physical activity. To monitor and detect dynamic and complex activities of human subjects, Qi et al. [57] have proposed an adaptive recognition framework. The proposed framework introduced an online learning independent of class constraints and classified 12 classes of activities. Experimentation has been performed to show that the proposed framework achieves the highest accuracy of 95.15% and 92.20% in detecting 5 dynamics, 6 static, and sequence of transitions when the mobile phones are kept in waist and pocket. In addition to complex activities to recognize transitional activities of short and long duration, research work [58] has proposed a deep learning scheme. The proposed scheme constructs a CNN for feature extraction and an LSTM network to involve the dependencies between actions to improve the HAR identification rate. The combined wearable sensor model accurately recognizes activities and their transitions.

The proposed system for HAR such as [57,58] faces certain challenges, particularly recognizing similar and confusing activities such as walking, climbing upstairs, etc with less inter-class and high intra-class scatter. To overcome this challenge, Lv et al.[59] have introduced a margin mechanism to extract discriminative features for classification. Four neural networks have been modified using the proposed margin mechanism, and the performances have been compared with traditional models on various benchmark datasets.

The smartphone-based HAR has certain other challenges such as device location and subject dependency that produces high false alarm rate and less positive alarm rate. This is mainly due to providing information about an acting class and not about the user context, such as location and objects manipulation. To address this challenge, research work [60] has integrated CNN and LSTM models. Experimentation has shown that CNN and LSTM model performed well on location and subject independent recognitions. The research work [61] have also proposed a hybrid activity recognition model to address location and subject dependency that combines a deep neural network model with a symbolic model. Extensive experimentation on the preparation of carrots soup has been performed to analyze the performance of proposed under realistic complex and contextual information.

Mukherjee et al. [62] have proposed EnsemConvNet, an ensemble of CNN-Net, Encoded-Net, and CNN-LSTM models that uses deep learning architectures with traditional shallow learning and statistical features. The proposed uses an automated weighted approach that includes majority voting, score fusion, sum rule and product rule to identify the classifier combinations. Performances are evaluated on WISDM activity prediction, UniMB SHAR, and MobiAct dataset.

Su et al. [63] have integrated the Deep Bidirectional Long Short-Term Memory (Deep DBLSTM) model and CNN model for an automated HAR system. The long sequence of raw data is processed using the DBLSTM model to generate a bidirectional output vector. Then, CNN is used to extract local features from the output vector. Finally, the fully connected network with softmax function classifies human activities.

TABLE IV. A SUMMARY OF HYBRID DEEP LEARNING MODELS FOR HUMAN ACTIVITY RECOGNITION SYSTEM

Authors	Proposed Architecture	Dataset(s) used	Performance Metrics (Avg. Accuracy)	Optimization on Hyperparameters	Issues handled/ Improvements	Scope for future enhancements
Xia et al, 2020 [56]	LSTM 1 - LSTM 2 CNN 1 - Max pooling - CNN 2 - GAP - BN - SL	UCI - HAR	F1 Score - 95.78%	Optimizer Batch size Filters	High robustness and better activity detection using fewer parameters	The more complex activity can be used to evaluate the generalized performance
		WISDM	F1 Score - 95.85%			
		Opportunity	F1 Score - 92.63%			
Qi et al., 2020 [57]	LSTM based Hierarchical Classification	Smartphone-based adaptive HAR dataset	Waist - 92.93 ± 3.32% Pocket - 88.37 ± 4.87%	No	Identifies the human motions in dynamic situations using an unsupervised Ada-HAR learning system	Requires more computation time in classifying the features with integrated clustering and deep learning modules
San Buenaventura et al., 2018 [60]	CNN - LSTM - Bayesian Optimization using GP	The fusion of smartphone motion sensors for physical activity recognition [46]	CNN - Belt 97.01% LSTM Belt 96.51% Overall - 97.17%	Bayesian Optimization using Gaussian Processes - Neurons Hidden layers Learning rate	Proposed Location independent, subject independent and automatic prediction system	Computational complexity analysis has to be considered for large scale data
			CNN - Subject 2 - 93.33% LSTM - Subject 3 - 83.04% Overall - 80.02%			
Mukherjee et al, 2020 [62]	CNN- Net + Encoded - Net + CNN - LSTM - classifier with Adaptive weighted approach	WISDM	97.1%	No	Identifies the sequential activities on data	To develop a customized model using transfer learning model.
		UniMiB SHAR	ADL & Fall -92.3% ADL - 98.7% Fall - 84.8% 2 ADL & 2 Fall - 99.4%			
		MobiAct	95.1%			
Su et al 2019 [63]	BLSTM Layer + .. + n BLSTM Layer - Abstractive representation - CNN Model - P - SL	UCI HAR	97.95%	No	Processes long sequence of raw time series data signals	Poor in the ability to extract features.
Gumaei et al. 2019 [64]	SRU 1 -D1 - SRU 2 + D2 - GRU 1 + D3 + FCNN	MHEALTH (body motion and vital signs)	Accuracy - 99.80% F1- score - 99.6%	No	Feedback connections are used to improve the accuracy	Large dataset with complex activity can be used
Wang et al., 2020a [58]	CNN + LSTM (C1 + P1 + C2 + P2 + C3+ P3 + Feature map + LSTM + BN + SL	HAPT dataset	95.87%	Learning rate Batch normalization Batch size	Improves the recognition between the Postural transition of activities	Complex actions can be added, such as eating and driving and considering different users' behavior differences.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3069927, IEEE Sensors Journal

Lv et al., 2020 [59]	{MLP + CNN + LSTM + Hybrid } + Arcmargin layer + SL	OPPORTUNITY	92.30 LSTM-M	Sliding window Sensor channels Margin value Softmax loss	Enhances the intra-class and inter-class diversity using margin mechanism integrated with softmax loss functions.	Exclusive network has been designed for each dataset
		UniMiB-SHAR	77.88 Hybrid M			
		PAMAP2	93.52 Hybrid M			
Rueda et al. 2019 [61]	Deep learning + symbolic models	The typical meal time routing	61.8 ± 5.5	No	Provides information about an action of events along with user context	To test with new dataset recorded in both laboratory and real-case scenario.
Singh et al, 2020 [65]	CNN + LSTM + SAL + SL	MHEALTH	94.86 ± 7.65	Filters of CNN, LSTM Units, Attention length, Learning rate, Temporal window size	Focuses on specific information on embedding layers of CNN and LSTM	To try out with global and local attention mechanism
		USC_HAD	90.88 ± 1.47			
		UTC-MHAD1	58.02 ± 2.29			
		UTC-MHAD2	89.84 ± 2.97			
		WHARF	82.39 ± 3.93			
		WISDM	90.41 ± 5.59			
Wang et al, 2020 [27]	RAN [CNN+ LSTM + SAL]	SWLM	99.0%	No	Weakly labeled dataset	Collection of annotated ground truth labeled training data
Gao et al, 2020 [66]	Traditional CNN + channel attention + temporal attention submodules	WISDM	98.85%	No	Demonstrates the superiority in improving the comprehensibility of multi-modal and weak labeled dataset	To explore more applications to better understand mixed attention mechanisms
		UNIMIB SHAR	79.03%			
		PAMAP2	93.16%			
		OPPORTUNITY	82.75%			
		SWLM	94.86%			
He et al, 2019 [67]	Context (CNN)+ Glimpse + Controller + classification n/w	UCI-HAR	95.35%	No	Addresses weakly labeled dataset	To improve the model performance with complex weakly labeled dataset
Zhu et al, 2019 [68]	Statistical features + DLSTM+ Dropout	UCI-HAR	97.21%	Layers Probability of Dropout layer	Features are extracted for local dependencies in the recurrent frame	To explore unseen classes recognition problem (cross-dataset)
Wang et al, 2019 [69]	{C1 + C2+ C3 }(AM1)+ {P + C4}(AM2) + {P+C5}(AM3) + P + FC1 + {AM1+AM2+AM3)	UCI-HAR	93.41%	No	Conducted location experiment on weakly labeled data	To deal with multi – sensor weakly labeled data sequences that has multiple types of complex activity
		SWLM	93.83%			
Ordóñez et al, [70]	IL1 + C2+ C3 + C4 + C5 + D1 + D2 + SL fused with LSTM recurrent units	Opportunity	93.0%	Sequence length + number of convolution layers	Focuses on DeepCONVLSTM	Transfer learning approach based models to be developed to recognize large-scale data.
Tang et al, 2021 [71]	Teacher model training + self labeling + student model training + transformation of signals	Fenland (Unlabeled)	95.15%	Covolutional layers, filter size	Training with labeled data and testing with fenland unlabeled dataset.	Downstream performance can be improved on better optimization of the model.
GAP- Global Pooling Layer, BN - Batch Normalization, MLP - Multilayer Perceptron, SL - Softmax Layer, C - Convolutional Layer, P - Pooling Layer, SRU - Simple Recurrent Units, GRU- Gated Recurrent Units, BN - Batch Normalization, SAL – Self Attention Layer						

The recent attraction towards developing robust deep learning techniques in smart health care to make decisions using multimodal sensory data has increased significantly. Gumaei et al. [64] has proposed an effective multi-modal sensor-based activity recognition using a hybrid deep learning model. The model integrates simple recurrent units (SRU) and gated recurrent units (RGUs) of neural networks. The deep SRUs are utilized to process the multimodal input sequences using their internal memory state capability. As well, deep GRUs is used for their feedback mechanism to improve the performance of stability in activity recognition. Ordonez et al [70] have proposed a DeepConvLSTM model based on the success of recurrent neural networks to recognize multimodal activities and to explicitly model the temporal dynamics of feature activations. Xu et al [72], have proposed a combination of Inception neural network and recurrent neural network model to recognize multi-channel sensor inputs. Inception module with various kernel-based convolution layers extracts the multi-dimensional features. Finally the classification has been performed using GRU layer.

As quoted earlier in the advantages of deep neural techniques, various research works [27,67,69] have proposed attention mechanism to perform well even in weakly labeled multi-activity recognition and locality tasks. He et al [67] have proposed a weakly supervised HAR based on recurrent attention learning that uses reinforcement learning to interact with sensor data over time. Initially the context and representative local CNN features are extracted through several iterations by applying a novel rewarding system introduced through recurrent attention learning. As the labeling of dataset collection is more difficult for complex and longer sequence of activities, Zhu et al [68] proposed a semi-supervised deep learning model using temporal ensembling of LSTM memory.

Tang et al [71] has proposed a SelfHAR, a semi-supervised model that effectively learns to leverage unlabeled mobile sensing datasets to complement small labeled datasets. This approach combines self-training along with augmentation with CNN model. Wang et al [28] have introduced recurrent attention networks (RAN) that combines attention mechanism with CNN feature extraction and LSTM decoder. The RAN proposed by [28] often have a weak feature than the traditional CNN. Moreover, the RAN concentrates on hard and soft attention which pays more attention to the target activity for a long sequence and it fails to address the Spatio-temporal dependencies of multimodal sensing signals. To overcome this, Gao et al [66] have proposed DanHAR: novel dual attention (blending channel attention and temporal attention) on CNN to demonstrate the superiority in improving the comprehensibility for multimodal HAR. Wang et al [69] have proposed a novel attention-based HAR which modifies the traditional CNN by an attention mechanism. The proposed model computes compatibility between global and local features extracted at the final fully convoluted layers and convolutional layers respectively.

HAR is primarily important in user-aware and context-aware services that need to be monitor continuously in daily life. In general, human activities have a longer time span than the HAR system, which recognizes for a few seconds. Thus, continuous monitoring of human activities using these proposed systems is computationally inefficient. To address context-aware detection, Jeong & Kim [73] has proposed a low computational cost segment level change detection mechanism to recognize the activities. In the proposed, Fully Convolutional Neural Network (FCNN) with a high recognition rate has been introduced to classify the activities. It outperforms the traditional CNN and other conventional techniques on extensive experimentation. Besides, the proposed system consumes 6.5 times less energy than the conventional models on embedded platforms.

Though CNN, LSTM and hybrid model capture Spatio-temporal features, Singh et al. [65] have proposed a self-attention layer to focus on specific information from the embeddings generated by any deep learning model. The self-attention layer has been used here to learn weights that capture the latent relationships between input time points of raw sensory data to decode human activity in an efficient way. Experimentation has been performed on six different HAR to validate the significance of self-attention layer.

Ascioglu et al [74], have addressed the recognition of activity in outdoor environments. They have generated a new dataset through a novel design of sensor based wireless activity monitoring system and its applications to deep learning neural networks. Experimentation has been carried out through CNN, LSTM, ConvLSTM.

IV. SCOPE FOR FUTURE ENHANCEMENT

Human activity recognition plays a vital role in Ambient Assisted Living, smart home environments, sports, and work injury detections. Various state-of-the-art techniques have recently proposed more traditional, generative, and discriminative deep learning models. However, they are not suitable enough to implement in real-world scenarios, as most of the techniques categorize simple activities rather than complex activities and have more false alarms. Thus, this section deals with certain challenges and problems that require research and future enhancements.

- *Annotation scarcity* – Training and evaluation of deep learning models require large volume of data. It is highly expensive to categorize, collect and annotate sensory activities.
- *Outdoor Activites* – Activities are performed by the subjects mostly in controlled and lab environments. Thus, possibility of real time implementation is a challenge as outdoor activities may differ with controlled environment.
- *Class Imbalance* – Emergent and unexpected activities such as accident, unsupervised fall is other major challenge as it is hard to perform and this may cause more injuries if not performed with safety measures.
- *Complex activities and Postural Transitions* – HAR tasks are mostly simple such as walking, sitting, sleeping etc. However, more complex activities are also carried out by each subject in daily routines such as washing clothes, moping the floor, bathing the pet, etc., In addition, there may be a chance of occurrence of fall between activities termed as postural transitions such as sitting to standing, standing to walking etc., Only UCI-HAPT dataset addresses Postural transitions.
- *Hyperparamter optimizations*- The performances of HAR model have dependency on Hyperparamter optimization as quoted in Table II, Table III and Table IV. Convolutional layers, filters, filter size, neurons, epochs, sliding window are certain parameters that every model needs to select the best value for better performances.
- *Multimodal sensor activities* – The performances of proposed systems perform better for unimodal sensors. However, the major challenge is to address by recognizing activities by fusing multimodal sensor and its activities such as smart watch, stretch sensor, and wearable devices.
- *Cross-dataset* – Various researchers have generated enormous number of benchmark datasets as listed in Table I for performance evaluation of HAR systems. However, the datasets generated are person dependent, time and sensor dependent. These factors lead to

challenge and discrepancies between training and testing of data with person independent data.

- *Analysis on large scale dataset* – The HAR model mostly deals with a maximum of 100K instances of training and testing data. Need a large scale adaptable HAR system to monitor the person throughout the day.
- *Cross-Adaptability data* – Other major challenge that needs to be addressed by training and testing with different platforms of data/ different devices as proposed by [25].
- Onboard implementation on smartphone and smartwatches
- Performance limitations in unsupervised learning

V. Conclusion

The automatic extraction of features for efficient classification of human activities is increasing a lot, and it has a wide scope in various domains. This results from the steady rise in computational resources such as GPU devices and ease availability and collection of sensory signals from smartphones and wearable devices. This paper reviews various deep neural network models that enable automatic extraction of features for human activity recognition with its uniqueness of implementation and design, advantages, and limitations. Several benchmark datasets available freely for performance evaluation with its characterization on creation and collection have also been discussed. Finally, open challenges that are required to be involved in future research enhancements are also discussed.

References

[1]. Nweke, H. F., Teh, Y. W., Mujtaba, G., & Al-Garadi, M. A. (2019). Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. Information Fusion, 46, 147-170.
[2]. Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L., &Amirat, Y. (2015). Physical human activity recognition using wearable sensors. Sensors, 15(12), 31314-31338.
[3]. Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. Pattern Recognition Letters, 119, 3-11.
[4]. Xu, T., Zhou, Y., & Zhu, J. (2018). New advances and challenges of fall detection systems: A survey. Applied Sciences, 8(3), 418. doi:10.3390/app8030418
[5]. Sathyanarayana, S., Satzoda, R. K., Sathyanarayana, S., &Thambipillai, S. (2018). Vision-based patient monitoring: a comprehensive review of algorithms and technologies. Journal of Ambient Intelligence and Humanized Computing, 9(2), 225-251.
[6]. Masoud, M., Jaradat, Y., Manasrah, A., &Jannoud, I. (2019). Sensors of Smart Devices in the Internet of Everything (IoE) Era: Big Opportunities and Massive Doubts. Journal of Sensors, 2019.
[7]. Nweke, H. F., Teh, Y. W., Al-Garadi, M. A., &Alo, U. R. (2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. Expert Systems with Applications, 105, 233-261.
[8]. Lockhart, J. W., & Weiss, G. M. (2014, September). Limitations with activity recognition methodology & data sets. In Proceedings of the 2014 ACM *International Joint Conference on Pervasive and Ubiquitous Computing*: Adjunct Publication (pp. 747-756).
[9]. Ramasamy Ramamurthy, S., & Roy, N. (2018). Recent trends in machine learning for human activity recognition—A survey. Wiley Interdisciplinary Reviews: *Data Mining and Knowledge Discovery*, 8(4), e1254.
[10]. UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/index.php
[11]. Kaggle datasets: https://www.kaggle.com/datasets
[12]. Reiss, A., &Stricker, D. (2012, June). Introducing a new benchmarked dataset for activity monitoring. In 2012 16th International Symposium on Wearable Computers (pp. 108-109). IEEE.
[13]. Stisen, A., Blunck, H., Bhattacharya, S., Prentow, T. S., Kjærgaard, M. B., Dey, A., ...& Jensen, M. M. (2015, November). Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition, in Proceedings of the 13th ACM conference on embedded networked sensor systems (pp. 127-140).
[14]. Banos, O., Garcia, R., Holgado-Terriza, J. A., Damas, M., Pomares, H., Rojas, I., ...&Villalonga, C. (2014, December). mHealthDroid: a novel framework for agile development of mobile health applications. International workshop on ambient assisted living (pp. 91-98). Springer, Cham.
[15]. Banos, O., Villalonga, C., Garcia, R., Saez, A., Damas, M., Holgado-Terriza, J. A., ...& Rojas, I. (2015). Design, implementation, and validation of a novel open framework for agile development of mobile health applications. Biomedical engineering online, 14(S2), S6.
[16]. Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013, April). A public domain dataset for human activity recognition using smartphones. In Esann (Vol. 3, p. 3).
[17]. Chavarriaga, R., Sagha, H., Calatroni, A., Digumarti, S. T., Tröster, G., Millán, J. D. R., &Roggen, D. (2013). The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition. Pattern Recognition Letters, 34(15), 2033-2042.
[18]. Weiss, G. M., Yoneda, K., &Hayajneh, T. (2019). Smartphone and smartwatch-based biometrics using activities of daily living. IEEE Access, 7, 133190-133202.
[19]. Micucci, D., Mobilio, M., &Napoletano, P. (2017). Unimibshar: A dataset for human activity recognition using acceleration data from smartphones. Applied Sciences, 7(10), 1101.
[20]. Chatzaki, C., Pediaditis, M., Vavoulas, G., &Tsiknakis, M. (2016, April). Human daily activity and fall recognition using a smartphone's acceleration sensor. In International Conference on Information and Communication Technologies for Ageing Well and e-Health (pp. 100-118). Springer, Cham.
[21]. Reyes-Ortiz, J. L., Oneto, L., Ghio, A., Samá, A., Anguita, D., & Parra, X. (2014, September). Human activity recognition on smartphones with awareness of basic activities and postural transitions. In International conference on artificial neural networks (pp. 177-184). Springer, Cham.
[22]. Krüger, F., Hein, A., Yordanova, K., &Kirste, T. (2017). Recognising user actions during cooking task (cooking task dataset) imu data. University Library, University of Rostock.
[23]. Chen, Y., &Xue, Y. (2015, October). A deep learning approach to human activity recognition based on a single accelerometer. In 2015 IEEE international conference on systems, man, and cybernetics (pp. 1488-1492). IEEE.
[24]. Zhu, R., Xiao, Z., Li, Y., Yang, M., Tan, Y., Zhou, L., ...& Wen, H. (2019). Efficient human activity recognition solving confusing activities via deep ensemble learning. IEEE Access, 7, 75490-75499.
[25]. Khan, M. A. A. H., Roy, N., &Misra, A. (2018, March). Scaling human activity recognition via deep learning-based domain adaptation. In 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom) (pp. 1-9). IEEE.
[26]. Sztyler, T., Stuckenschmidt, H., & Petrich, W. (2017). Position-aware activity recognition with wearable devices. Pervasive and mobile computing, 38, 281-295.
[27]. Wang, K., He, J., & Zhang, L. (2020). Sequential Weakly Labeled Multi-Activity Localization and Recognition on Wearable Sensors using Recurrent Attention Networks. arXiv e-prints, arXiv-2004.
[28]. Bhat, G., Tran, N., Shill, H., & Ogras, U. Y. (2020). w-HAR: An Activity Recognition Dataset and Framework Using Low-Power Wearable Devices. *Sensors*, *20*(18), 5356.
[29]. Um, T. T., Babakeshizadeh, V., &Kulić, D. (2017, September). Exercise motion classification from large-scale wearable sensor data using convolutional neural networks. In 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 2385-2390). IEEE.
[30]. Alemayoh, T. T., Lee, J. H., & Okamoto, S. (2019, June). Deep Learning-Based Real-time Daily Human Activity Recognition and Its Implementation in a Smartphone. In 2019 16th International Conference on Ubiquitous Robots (UR) (pp. 179-182). IEEE.

[31]. Lawal, I. A., &Bano, S. (2020). Deep Human Activity Recognition WithLocalisation of Wearable Sensors. IEEE Access, 8, 155060-155070.

[32]. Qin, Z., Zhang, Y., Meng, S., Qin, Z., & Choo, K. K. R. (2020). Imaging and fusing time series for wearable sensor-based human activity recognition. Information Fusion, 53, 80-87.

[33]. Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, *119*, 3-11.

[34]. Ronao, C. A., & Cho, S. B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. Expert systems with applications, 59, 235-244.

[35]. Ignatov, A. (2018). Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Applied Soft Computing*, *62*, 915-922.

[36]. Avilés-Cruz, C., Ferreyra-Ramírez, A., Zúñiga-López, A., & Villegas-Cortéz, J. (2019). Coarse-fine convolutional deep-learning strategy for human activity recognition. Sensors, 19(7), 1556.

[37]. Wan, S., Qi, L., Xu, X., Tong, C., &Gu, Z. (2020). Deep learning models for real-time human activity recognition with smartphones. Mobile Networks and Applications, 25(2), 743-755.

[38]. Tang, Y., Teng, Q., Zhang, L., Min, F., & He, J. (2020). Efficient convolutional neural networks with smaller filters for human activity recognition using wearable sensors. arXiv preprint arXiv:2005.03948.

[39]. Tang, Y., Teng, Q., Zhang, L., Min, F., & He, J. (2020). Layer-wise training convolutional neural networks with smaller filters for human activity recognition using wearable sensors. *IEEE Sensors Journal*, *21*(1), 581-592.

[40]. Cheng, X., Zhang, L., Tang, Y., Liu, Y., Wu, H., & He, J. (2020). Real-time Human Activity Recognition Using Conditionally Parametrized Convolutions on Mobile and Wearable Devices. *arXiv preprint arXiv:2006.03259*.

[41]. Cruciani, F., Vafeiadis, A., Nugent, C., Cleland, I., McCullagh, P., Votis, K., ...&Hamzaoui, R. (2020). Feature learning for Human Activity Recognition using Convolutional Neural Networks. CCF Transactions on Pervasive Computing and Interaction, 2(1), 18-32.

[42]. Nutter, M., Crawford, C. H., & Ortiz, J. (2018, July). Design of novel deep learning models for real-time human activity recognition with mobile phones. In 2018 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

[43]. Xiao, F., Pei, L., Chu, L., Zou, D., Yu, W., Zhu, Y., & Li, T. (2020). A Deep Learning Method for Complex Human Activity Recognition Using Virtual Wearable Sensors. arXiv preprint arXiv:2003.01874.

[44]. Almaslukh, B., Artoli, A. M., & Al-Muhtadi, J. (2018). A robust deep learning approach for position-independent smartphone-based human activity recognition. Sensors, 18(11), 3726.

[45]. Shojaedini, S. V., &Beirami, M. J. (2020). Mobile sensor-based human activity recognition: distinguishing of challenging activities by applying long short-term memory deep learning modified by residual network concept. Biomedical Engineering Letters, 10(3), 419-430.

[46]. Chung, S., Lim, J., Noh, K. J., Kim, G., &Jeong, H. (2019). Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning. Sensors, 19(7), 1716.

[47]. Agarwal, P., &Alam, M. (2020). A lightweight deep learning model for human activity recognition on edge devices. Procedia Computer Science, 167, 2364-2373.

[48]. Rashid, N., Demirel, B. U., & Faruque, M. A. A. (2021). AHAR: Adaptive CNN for Energy-efficient Human Activity Recognition in Low-power Edge Devices. *arXiv preprint arXiv:2102.01875*.

[49]. Zhao, Y., Yang, R., Chevalier, G., Xu, X., & Zhang, Z. (2018). Deep residual bidir-LSTM for human activity recognition using wearable sensors. *Mathematical Problems in Engineering*, *2018*.

[50]. Wang, L., & Liu, R. (2020). Human activity recognition based on wearable sensors using hierarchical deep LSTM networks. Circuits, Systems, and Signal Processing, 39(2), 837-856.

[51]. Ashry, S., Ogawa, T., & Gomaa, W. (2020). CHARM-Deep: Continuous Human Activity Recognition Model Based on Deep Neural Network using IMU Sensors of Smartwatch. *IEEE Sensors Journal*, *20*(15), 8757-8770.

[52]. Ashry, S., Elbasiony, R., & Gomaa, W. (2018). An LSTM-based descriptor for human activities recognition using IMU sensors. In *Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics, ICINCO* (Vol. 1, pp. 494-501).

[53]. Gomaa, W., Elbasiony, R., & Ashry, S. (2017, December). Adl classification based on autocorrelation function of inertial signals. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 833-837). IEEE.

[54]. Sun, J., Fu, Y., Li, S., He, J., Xu, C., & Tan, L. (2018). Sequential human activity recognition based on deep convolutional network and extreme learning machine using wearable sensors. Journal of Sensors, 2018.

[55]. Zhou, X., Liang, W., Kevin, I., Wang, K., Wang, H., Yang, L. T., &Jin, Q. (2020). Deep Learning Enhanced Human Activity Recognition for the Internet of Healthcare Things. IEEE Internet of Things Journal.

[56]. Xia, K., Huang, J., & Wang, H. (2020). LSTM-CNN Architecture for Human Activity Recognition. IEEE Access, 8, 56855-56866.

[57]. Qi, W., Su, H., &Aliverti, A. (2020). A Smartphone-Based Adaptive Recognition and Real-Time Monitoring System for Human Activities. IEEE Transactions on Human-Machine Systems.

[58]. Wang, H., Zhao, J., Li, J., Tian, L., Tu, P., Cao, T., ...& Li, S. (2020). Wearable Sensor-Based Human Activity Recognition Using Hybrid Deep Learning Techniques. Security and Communication Networks, 2020.

[59]. Lv, T., Wang, X., Jin, L., Xiao, Y., & Song, M. (2020). Margin-Based Deep Learning Networks for Human Activity Recognition. Sensors, 20(7), 1871.

[60]. San Buenaventura, C. V., Tiglao, N. M. C., & Atienza, R. O. (2018, October). Deep Learning for Smartphone-Based Human Activity Recognition Using Multi-sensor Fusion. International Wireless Internet Conference (pp. 65-75). Springer, Cham.

[61]. Rueda, F. M., Lüdtke, S., Schröder, M., Yordanova, K., Kirste, T., & Fink, G. A. (2019, March). Combining Symbolic Reasoning and Deep Learning for Human Activity Recognition. In 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops) (pp. 22-27). IEEE.

[62]. Mukherjee, D., Mondal, R., Singh, P. K., Sarkar, R., &Bhattacharjee, D. (2020). EnsemConvNet: a deep learning approach for human activity recognition using smartphone sensors for healthcare applications. Multimedia Tools and Applications, 1-28.

[63]. Su, T., Sun, H., Ma, C., Jiang, L., & Xu, T. (2019, July). HDL: Hierarchical Deep Learning Model based Human Activity Recognition using Smartphone Sensors. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

[64]. Gumaei, A., Hassan, M. M., Alelaiwi, A., &Alsalman, H. (2019). A hybrid deep learning model for human activity recognition using multimodal body sensing data. IEEE Access, 7, 99152-99160.

[65]. Singh, S. P., Sharma, M. K., Lay-Ekuakille, A., Gangwar, D., & Gupta, S. (2020). DeepConvLSTM with self-attention for human activity decoding using wearable sensors. IEEE Sensors Journal.

[66]. Gao, W., Zhang, L., Teng, Q., Wu, H., Min, F., & He, J. (2020). DanHAR: dual attention network for multimodal human activity recognition using wearable sensors. *arXiv preprint arXiv:2006.14435*.

[67]. He, J., Zhang, Q., Wang, L., & Pei, L. (2019). Weakly supervised human activity recognition from wearable sensors by recurrent attention learning. *IEEE Sensors Journal*, *19*(6), 2287-2297.

[68]. Zhu, Q., Chen, Z., & Soh, Y. C. (2018). A novel semisupervised deep learning method for human activity recognition. *IEEE Transactions on Industrial Informatics*, *15*(7), 3821-3830.

[69]. Wang, K., He, J., & Zhang, L. (2019). Attention-based convolutional neural network for weakly labeled human activities' recognition with wearable sensors. *IEEE Sensors Journal*, *19*(17), 7598-7604.

[70]. Ordóñez, F. J., & Roggen, D. (2016). Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, *16*(1), 115.

[71]. Tang, C. I., Perez-Pozuelo, I., Spathis, D., Brage, S., Wareham, N., & Mascolo, C. (2021). SelfHAR: Improving Human Activity Recognition through Self-training with Unlabeled Data. *arXiv preprint arXiv:2102.06073*.

[72]. Xu, C., Chai, D., He, J., Zhang, X., & Duan, S. (2019). InnoHAR: A deep neural network for complex human activity recognition. *IEEE Access*, *7*, 9893-9902.

[73]. Jeong, C. Y., & Kim, M. (2019). An energy-efficient method for human activity recognition with segment-level change detection and deep learning. Sensors, 19(17), 3688.

[74]. Ascioglu, G., & Senol, Y. (2020). Design of a Wearable Wireless Multi-Sensor Monitoring System and Application for Activity Recognition Using Deep Learning. *IEEE Access*, *8*, 169183-169195.