Al-Nahrain Journal of Science
ANJS, Vol.27 (1), March, 2024, pp. 108-120

Human Activity Recognition Using Inertial Sensors in a Smartphone: Technical Background (Review)

Hade Khalaf *, Musaab Riyadh

Department of Computer, College of Science, Mustansiriyah University, Baghdad, Iraq.

Article's Information

Received: 16.09.2023
Accepted: 25.02.2024
Published: 15.03.2024

Keywords:
Deep Learning
HAR Approach
Machine Learning
Technical Background

Abstract

Human Activity Recognition (HAR) stands at the intersection of machine learning, deep learning, and sensor technology, primarily focusing on leveraging inertial sensors in smartphones and wearable devices. This paper presents a comprehensive technical overview of HAR, examining the amalgamation of machine learning and deep learning systems while considering the data inputs from mobile and wearable inertial sensors. The review encompasses a broad spectrum of methodologies applied to HAR, ranging from classical machine learning algorithms to cutting-edge deep learning architectures. Emphasis is placed on the nuanced challenges and opportunities posed using inertial sensors in smartphones and wearables. This includes discussions on data preprocessing strategies, feature extraction methods, and model architectures, accounting for the unique characteristics of sensor data, such as noise, variability, and power consumption. The paper explores recent advancements, scrutinizing state-of-the-art approaches, innovative model architectures, and emerging trends in HAR. Through a comparative evaluation of various machine learning and deep learning techniques, the review aims to guide researchers and practitioners in selecting the most appropriate methods for HAR applications across diverse scenarios. In conclusion, this paper serves as an inclusive guide to the technical landscape of HAR, incorporating insights from both mobile and wearable inertial sensors. By synthesizing existing knowledge and addressing future research directions, it aims to propel advancements in developing robust and efficient systems for recognizing human activities, accommodating the evolving landscape of sensor technologies in mobile and wearable devices.

http://doi.org/10.22401/ANJS.27.1.10

* Corresponding author email address: hadi_khalaf@uomustansiriyah.edu.iq

1. Introduction

The goal of the complex problem known as Human Activity Recognition (HAR) is predicting human gestures via computer interaction. Using a wide range of applications, it improves human lives. According to Figure 1., there are two primary approaches for identifying human activity: vision-based and sensor-based [1, 2]. The camera is used to identify human activities in video systems. This method calls for costly infrastructure installations for cameras and presents specific difficulties owing to lighting, background, and scale issues that could make motion detection challenging. With regard to the second method, motion is converted into recognized signals using wearable sensors like accelerometers, barometers, gyro-meters, etc. In addition to providing privacy for its users, it offers an alternate method of motion acquisition free from the same environmental limitations as a video-based technique. Yet, there are inherent limits to activity detection depending on this methodology when it comes to gathering enough data regarding all pose motions in the human body, which could adversely impact performance. Since people wear them to automatically detect and track various behaviors, like jogging, sitting, sleeping, and running, wearable devices are illustrative instances of sensor-based HAR [3]. On the other hand, a sensor is ineffective when a subject is either beyond its detection range or engages in unknown behaviors [4]. The HAR method's foundation is the intricacy of the various data inputs, such as those from wearable sensors, images, object sensors, and audio. On the basis of data from various sensors, it could be described as the capacity for recognizing or detecting present activity [5]. The major contribution of this study is to demonstrate to researchers the datasets most frequently utilized in literature for experimentation processes, to detail the most popular algorithms in their analysis, to identify each data processing method and the outcomes of the algorithms' experiments, and to distinguish the quality metrics related to said applications. This study reviews the research on applying various machine learning (ML) techniques, including unsupervised, supervised, and deep learning (DL). Systems for recognizing actions and gestures depending on video analysis were thoroughly researched [5,6].

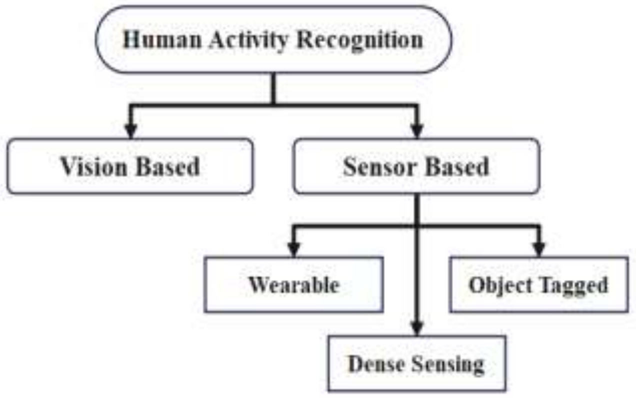

Figure 1. Classification Of Human Activity Recognition Approaches

2. Activity recognition architecture

Human activity recognition (HAR) involves identifying and classifying activities performed by individuals based on sensor data, often collected from various sources such as accelerometers, gyroscopes, and sometimes additional sensors like magnetometers and GPS. As shown in Figure 2. the traditional architecture for human activity recognition typically involves several key components [7]:

- Data Collection: sensors are used to collect data related to human activities. Standard sensors include accelerometers, gyroscopes, and sometimes additional sensors like magnetometers and GPS.
- Data Preprocessing: Clean and preprocess the raw sensor data to remove noise, handle missing values, and standardize the format.
- Segmentation: the prepressed data are segmented using over or non-over-lapped windows.
- Feature Extraction: Extract relevant features from the preprocessed data. Features could include statistical measures, frequency domain features, and other relevant characteristics that capture the essence of the activities.
- Feature Selection: Select a subset of the extracted features to reduce dimensionality and focus on the most informative aspects of the data.
- Training: Train the chosen model using labeled data. Machine learning or deep learning is often employed, where the model learns to map input sensor data to corresponding activity labels.
- Testing and Evaluation: Evaluate the trained model on a separate dataset not seen during training to assess its generalization performance. Common metrics include accuracy, precision, recall, and F1 score.

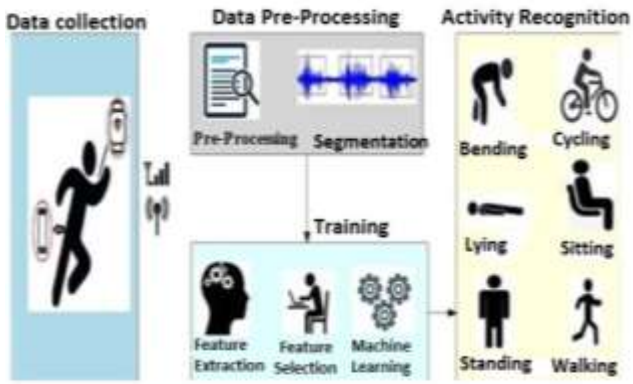

Figure 2. This is a Figure. Example of Activity Recognition Using Sensors [7].

2.1. Data collection

Any HAR system's base is built on data collection, which is the acquisition of sensor signals. The sensor-based HAR method was used in a number of practical applications, particularly those related to smart homes and healthcare. Additionally, a lot of data has been gathered from various sensors, including object sensors, wearable sensors, and environmental sensors, thanks to the rapid growth of wireless sensor networks (WSN) [8].

Wearable sensors are the most common sensor-based HAR, and the recent expansion of the Internet of Things (IoT) and mobile computing has created the ideal setting for their development. Individuals carrying out their daily tasks have sensors installed on their bodies to gather data. Accelerometers, which measure acceleration and indicate the rate of change in an object's motion; gyroscopes, which measure orientation and angular velocity; and other sensors are crucial for detecting human movement. After that, by comparing the signal differences prior to and following an activity, human activity could be identified [9].

2.2. Data Preprocessing

Following the collection of a dataset, it must be pre-processed to reduce noise introduced throughout data collection and the sensors themselves prior to feeding into the ML algorithm for sensor-based HAR. Data pre-processing is considered the most significant process impacting HAR frameworks' overall performance. Denoising and segmentation are both steps in the pre-processing process. Following pre-processing, feature engineering is carried out, which entails feature selection and feature extraction. Feature extraction extracts a set of features from the input while preserving important data. The extracted feature may occasionally need to undergo pre-processing, such as reducing its dimension [10].

2.2.1. Error handling

There were certain problems in the dataset that were corrected, including noise from malfunctions, calibration issues, noisy ambient conditions, positioning issues, and various activities. Some rows are vacant, while others have additional attributes. In order to reduce the noise generated, data pre-processing methods are crucial. Mean filter, low-pass filter, wavelet filter, linear filter, and Kalman filter are examples of common denoising techniques [11].

2.2.2. Segmentation

The inertial sensor signals are organized at the segmentation stage so that the precise time that an activity is taking place may be determined. The inertial sensor signals are divided into time windows, which are sequential subsequences of continuous values. It could be divided into time-driven windows segmentation, event-driven windows segmentation, and action-driven windows segmentation in the context of HAR smartphones. By using non-overlapping or overlapping time windows, segmentation depending on time windows could be controlled in these two different ways [12]. Segmentation with non-overlapping windows is done by dividing a signal or sequence into fixed-size, non-overlapping segments or windows; each segment is then processed independently. This approach is particularly useful when you want to analyze or process a long sequence in a more manageable and computationally efficient way. In contrast to non-overlapping windows, overlapping windows involve partially covering the same portion of the data in multiple adjacent windows. This technique is useful for capturing more context and ensuring that features near the boundaries of windows are considered in multiple segments [13].

2.2.3. Feature extraction

The extraction of features from pre-processed data stands out as a critical phase in developing machine learning and deep learning systems, particularly those designed for human activity recognition. Features serve as foundational elements, given that these systems often do not autonomously extract features, especially in widely utilized datasets like WiSDM, WHARF, OPPORTUNITY, and others. These datasets involve the collection of digital signals from sensors positioned on participants' bodies. In contrast, datasets recorded via video, though less commonly employed, also exist [14]. Similarly, there exists a multitude of techniques for extracting features from digital signals, and the choice of method often depends on the specific signal domain being addressed. Certain methods focus on the time domain of the digital signal, categorized as statistical features. Conversely, others operate in the frequency domain, presenting features of greater complexity compared to statistical features. Notably, a challenge arises in the extraction process due to the prevalence of datasets recorded in the time domain. Converting these datasets to the frequency domain is a prerequisite, followed by analysis using signal analysis methods before feature extraction can take place [15].

Fundamentally, every feature extraction method is intricately linked to the processing of digital signals captured by sensors. Consequently, the precision of outcomes in human activity recognition systems is heavily contingent on the technological advancements in sensors and the accuracy and fidelity of signals they record. This underscores the fact that progress in sensor and accelerator technologies directly influences human activity recognition systems' classification and recognition capabilities [16].

2.3. Feature selection

A sub-set of features which are significant for the classification algorithms to consider is selected by feature selection [17]. An alternate method of feature selection is representation learning, in which models focus on the analysis of data for extracting a useful feature set. This method also decreases high-dimensional spaces as well as temporal complexity through deleting irrelevant features [18]. The three primary techniques used in feature selection are wrapper approaches, embedded techniques, and filter techniques. Utilizing the inherent properties of features and variables, filter approaches choose a subset of the original features by ranking them according to the correlation coefficient [19]. Wrapper approaches, as opposed to filter methods, were shown to yield greater performance since various classifiers are used to assess how well the chosen subsets perform [20]. However, embedded approaches determine the appropriate feature subset through determining the optimal weights regarding a function which has previously shown excellent results. Wrapper and embedded methods are both applicable to regression and multiclass problems [21].

3. Technical Background

Given the abundance of research dedicated to the classification and recognition of human activity, this review will primarily focus on recently published studies, specifically those within the last five years, as illustrated in figure 3. The systems investigated in these studies are categorized into two main methodologies: the machine learning approach and the deep learning approach. Each methodology section is further subdivided into branches based on the specific techniques employed, with a brief discussion of each technique, including references to the studies utilizing them. This classification is intended to comprehensively encompass a wide range of techniques employed in the field of human activity classification and recognition, aiming to draw insights from the experiences of authors in developing novel systems in this domain.

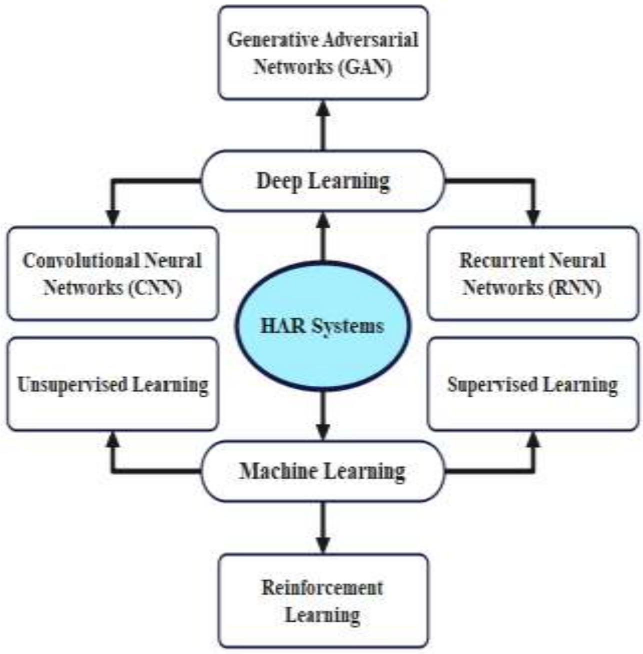

Figure 3. This is a figure. HAR Approach Concepts.

3.1. Machine learning (ML)

Machine learning (ML) can be defined as a subfield of AI that allows computers to "self-learn" from the training data as well as get better over time without being explicitly programmed. In addition, ML algorithms can recognize patterns in data as well as learn from them to make their predictions. ML methods were utilized for a while for addressing the HAR problem, such as K-Nearest Neighbors [22], random forest (RF) [23], Naive Bayes [24, 25], and SVM [26, 27]. ML models and algorithms learn via experience. Conventional ML algorithms have shown outstanding performance in environments with minimal data and strict regulation. Although such algorithms are ineffective and time-consuming, they do need a number of pre-processing stages and well-designed characteristics [28].

3.1.1. Supervised learning

The bulk of biomedical investigations use supervised classification. A set of data that includes certain information regarding the dataset is used by almost all classification algorithms. In other words, the dataset contains the class label data necessary to train the classifier. This type of classification involves supervised learning, where the classifier receives guidance from a supervisor as it builds the classification model. The training set, which is a collection of examples that were manually labeled with the desired output, is assumed to have been provided in order to use the supervised learning technique [29]. The next methods could be emphasized as the ones utilized in ML the most. In their article, Gamal et al. covers multimodal on-body inertial sensor-based ML methods for HAR [30]. Two methods are employed to create a classifier in this study because the goal is to create an accurate one. The first method is represented via model (A), whereas the second is represented through model (B). The two models were created with the use of RF, NB, SVM, and Multilayer Perceptron (MLP) ML classification techniques. Model (A) relies on one classifier to identify 12 activities, whereas Model (B) relies on an ensemble method and is constructed in two steps. In order to lessen the detrimental impacts of imbalanced classes, which may be seen in HAR data, the synthetic Minority Over-sampling Technique (SMOTE) was employed. According to experimental comparison results, model (B) enhanced model (A's) accuracy by 4.7% in MLP, 3.6% in RF, and 23.2% in NB, although model (A's) accuracy in SVM was greater than model (B's). On the other hand, the NB Classifier is a unique type of ML algorithm dealing with the classification problem. The "Bayes theorem" serves as a basis for that. The variables that are utilized for the prediction are considered independent of each other in the NB Classifier method. Put differently, NB could be easily applied to large datasets, and users unfamiliar with classifier technology may quickly understand it. The existence of a set of properties in one data set doesn't necessarily imply a lack of another character in a different data set. Although it might not be the ideal classifier for a given application, it could be counted on to be reliable and perform well.

Randhawa et al. demonstrated the whole structure of detection or recognition from data collection to classification [31]. Also, the study shows where fabric pressure, stretch, and 9 DOF accelerometers should be placed on the subject. It is groundbreaking research that classified violent and normal responses using fabric sensors. Three distinct types of algorithms, like the K-NN, DTs, and SVMs, are utilized in the study to differentiate between normal motions and violent attacks. The results showed that the SVM predicts with a 98.80% accuracy rate compared to other methods. In the case where there is a clear margin of separation between the classes, SVM is more effective in the high-dimensional spaces. It functions well and requires comparatively minimal memory in the case when the number of dimensions is greater than the number of samples. It performs poorly when the dataset has more noise, like when target classes overlap, and is unsuitable for large data sets. The support vector classifier distributes the data points over and below the classifying hyper-plane, so there is no probabilistic basis for the task of classification.

3.1.2. Reinforcement learning

Reinforcement Learning (RL) is a type of machine learning paradigm where an agent learns to make decisions by interacting with an environment. The agent takes actions, and the environment provides feedback in the form of rewards or punishments. The goal of the agent is to learn a strategy or policy that maximizes the cumulative reward over time [32]. An agent is an entity or system that is learning and making decisions. It takes actions in the environment to achieve its goals. The environment is the external system or surroundings with which the agent interacts. The environment responds to the actions of the agent and provides feedback. The state of the agent is a representation of the current situation or configuration of the environment. The state contains the information necessary for the agent to make decisions. Action is a set of possible moves or decisions that the agent can take in a given state. Actions influence the environment and lead to state transitions. A reward is a numerical value that the environment provides to the agent as feedback for a specific action taken in a particular state. The agent's objective is to maximize the cumulative reward over time. Policy is the strategy or mapping from states to actions that the agent follows. The policy defines the behavior of the agent in the given environment. A value function is a function that estimates the expected cumulative future rewards of being in a particular state or taking a specific action. Value functions are crucial for assessing the desirability of different states and actions. Balancing the exploration of new actions to discover potentially better strategies and exploiting known actions to maximize immediate rewards is a fundamental challenge in reinforcement learning [33]. Deep Q Network (DQN) is a reinforcement learning algorithm that combines the principles of Q-learning with deep neural networks, allowing it to handle high-dimensional state spaces commonly found in complex environments [34].

Deep Q Network (DQN) was used by Fan et al. to introduce a hybrid feature selection method that leveraged the integration of Bee Swarm Optimization (BSO) and multi-agent Deep Q-Network (DQN) to enhance human activity recognition [35]. This method was applied for feature selection using the UCI-HAR dataset.

3.1.3. Unsupervised learning

Unsupervised learning is a category of machine learning where the algorithm is given data without explicit instructions on what to do with it. The system tries to learn the patterns and the structure from the data without labeled responses to guide the learning process. The goal of unsupervised learning is to explore the hidden structure or relationships within a set of data [36]. Dealing with small quantities of data in unsupervised learning can be challenging due to the potential limitations in the algorithms' ability to generalize patterns effectively. Unsupervised learning algorithms often benefit from larger datasets to discover more robust patterns and structures [37]. Unsupervised learning encompasses various methods like clustering, association rules, and dimensionality reduction methods.

i. Clustering Methods

Clustering is a technique in unsupervised learning that aims to group similar data points together based on certain features or c haracteristics. Unlike supervised learning, where the algorithm is provided with labeled training data to learn from, unsupervised learning deals with unlabeled data and aims to discover patterns or structures within the data [38]. The primary objective of clustering is to partition a set of data points into groups or clusters such that data points within the same cluster are more similar to each other than to those in other clusters. There are various clustering algorithms, and each has its own strengths, weaknesses, and use cases [39].

Jamel et al. provided a desicription for the data sets from accelerometers which can be utilized for gauging human physical activity were provided and reviewed by the authors [40]. To reduce the dataset's heterogeneities as well as high correlations, the features were divided into two components: PCA and low-pass filtration. Data was clustered using the approximation kernel k-means technique. With regard to the approximation step, employ the KNN classifier and Euclidean distance techniques. For selecting an impartial sample for approximation methods, multistage sampling approaches depending on stratification and SRS were used. The watch and phone datasets generated the best results with f-score accuracy values of 79% and 90%, respectively. Despite being used to solve a clustering problem, the method outperformed classification methods regarding results. The method was applied in a parallel environment, which produced a noticeable enhancement. Sheng et al. presented a method for unsupervised embedding learning that projects activity data into the embedding space using the temporal coherence and locality preservation of human activities [41]. This method is built on an autoencoder framework. The studies' findings demonstrate that the method can cluster related tasks together in the embedding space, which enhances the performance of the ensuing clustering task. Accuracy performance of 92.11%, 71.49%, and 80.73%, respectively, on PAMAP2, REALDISP, and SBHAR datasets.

ii. Association Rules Methods

The foundation on which the association rules build their analysis is provided by "if-then" algorithmic phrases that permit supporting different probabilities included in different data parts. Numerous databases in different types and formats can be found with such sentences. Depending on algorithmic "if-then" statements and established requirements like trust and support, association rules could select the most significant patterns. Also, support criteria give the association rules the capability to identify how often elements in the dataset appear. The confidence criteria could specify how often the Boolean value of the "if-then" statement is true. Fit is another well-known metric that basically compares the predicted confidence with the confidence seen in data [42]. Zhang et al. described a unique approach (HAR-AWDF) which combines association rules with adaptive weighted decision fusion to outperform the traditional AC approach [43]. The most crucial rule sets and item sets are extracted by this approach using measure integration, thresholding, and item weighting. The successful experimental results from comparisons to various measures and AC-based algorithms show the effectiveness of the method. The mean recognition rate of HAR-AWDF is more than 92% in a case where 10% of training samples are chosen randomly from the German Credit Dataset, and the standard deviation is smaller than 0.02.

iii. Dimensionality reduction methods

Dimensionality reduction is a technique used in unsupervised learning to reduce the number of features (dimensions) in a dataset while preserving its essential information. The main goal is to simplify the data representation, making it more manageable and efficient for subsequent analysis or visualization [44]. Since there are various methods for dimensionality reduction, the Principal Component Analysis (PCA) is one of the common approaches; the primary goal of PCA is to transform the original features into a new set of uncorrelated features, known as principal components. These components are ordered by the amount of variance they capture [45]. The Principal Component Analysis (PCA) is widely used for speeding up machine learning algorithms as in. Aljarrah. et al. proposed a BiLSTM RNN model equipped with PCA to classify human activities out of the mHealth dataset [46]. The PCA steps assumed a variation of no less than 95%, resulting in a dimensionality reduction to 17 instead of 21. Results indicated that the PCA-BLSTM model has registered the highest accuracy of 97.64%.

3.2. Deep Learning

DL is a subset of ML that performs significantly better on unstructured data; it outperforms existing ML methods; it became important because it uses DNNs. DL passes the data through multiple layers, with each layer being capable of gradually extracting features before passing it to the next layer. As a sort of ML, this technology aids AI in continually learning by extracting low-level information from the initial layers and combining them in subsequent layers to create a comprehensive representation. Compared to conventional ML algorithms, DL significantly decreases the effort required to select the best features by automatically extracting abstract features through multiple hidden layers. DL structure was shown to be effective with reinforcement learning and unsupervised learning [47]. As a result, HAR frameworks that are based on DL are being developed more frequently. For instance, using a publicly available dataset of 24 ASL static hand motions, Oyedotun O.K et al. [48] trained a CNN depending on stacked autoencoders. The model's 91.33% recognition rate following training has demonstrated the potential of DL in HAR. Even while there were numerous reviews of both HAR [49] and DL [50,51], few of them have addressed both subjects, and the fact that HAR is currently under development means that various new ideas are being introduced. DL depends on the deployment of ANNs. The three most popular types of NNs are.

3.2.1. Convolutional neural networks (CNN)

CNNs, a particular kind of DNN, use a grid-like arrangement to evaluate data. In the mathematical model referred to as CNN, feature extraction is carried out using pooling and convolutional layers, and the final output is accomplished using a fully connected layer (FCL). The convolutional layer, which is the foundation of CNN and is crucial to how it functions, is made up of the activation function and the convolution operation. CNNs are majorly used in computer vision and have also shown promise in text classification tasks related to NLP. By arranging word vectors into a matrix and processing it like an image, CNNs could analyze collections of words using a context window [52]. Xu et al. suggested building a CNN utilizing information gathered from a three-axis accelerometer built into users' smartphones [53]. Jogging, walking, standing, sitting, going upstairs, and going downstairs are among the everyday human activities that have been selected for recognition. Without any elaborate preparation, the 3-D raw accelerometer data is utilized as the input for the CNN training. The accuracy regarding the CNN-based approach for multiple HAR was 91.97%, outperforming the SVM method, which had an accuracy of 82.27% after being trained and evaluated with 6 different feature types from the 3D raw accelerometer data. As a result, the suggested method produced good recognition accuracy at minimal computational cost, and one advantage of CNN is that manual feature extraction is unnecessary. Given that the CNN classifier is fully linked and only accepts raw data as input, the NN contains many parameters. CNN enforces the local connection pattern between neurons and neighboring layers. Because of the 3D nature, the raw accelerometer data has been utilized as input for training CNN without any elaborate preprocessing. SVMs perform poorly in the case when the target classes overlap and the data set contains more noise; hence, using them with large data sets is not advised. In the case when there are more training data samples when compared to features for every one of the data points, SVM will not perform well. As a result, the CNN-based approach beat the SVM for identifying metahuman activity, which was trained and tested using six distinct types of characteristics extracted from 3-D raw accelerometer data.

Bianchi et al. suggested a new IoT system for long-term, personalized monitoring of the individual's actions at home [54]. In order to gather data on various activities and deduce problematic behaviors, the system combines a Wi-Fi wearable sensor with DL methods. The method described here was designed to be expanded to systems needing many wearable sensors that provide information in a personalized way. CNN-based architecture (with one FCL, 4 convolutional layers, as well as a sliding processing window of 2.56s designed for real-time elaboration) was used for classifying activities, and it can do so with a 97% accuracy. Despite having a limited training set, however. This discovery is noteworthy since it shows how simple it is to design and calibrate several HAR systems for different problem classes, like the age groups of individuals. With regard to computing resources, employing a pre-trained network for inference is less expensive in comparison to utilizing a DL method, which, like any ML technique, needs a lot of processing power for training the network. The training phase was developed to remain in the cloud since it commonly demands the most significant processing power, and the network is just updated in the case where a new user must be added, which further minimizes battery usage. Yet, it is acknowledged that there is a concern with energy use. Through buffering data in the sensor's internal memory and transferring compressed data to the cloud, data could be elaborated on the sensor itself.

3.2.2 Recurrent neural networks (RNN)

Recurrent Neural Networks (RNNs) are a class of artificial neural networks designed for sequential data processing and tasks. They are particularly effective for tasks where the input and output data are sequences, such as time series prediction, speech recognition, language modeling, and more. RNNs are designed to work with sequences of data where the order of elements matters. Each sequence component is processed one at a time, and the network maintains a hidden state that captures information about the sequence seen so far [55]. The key feature of RNNs is the presence of recurrent connections, which allow information to be passed from one step of the sequence to the next. This enables the network to maintain a memory of previous inputs and use that information to influence the processing of current inputs. The hidden state in an RNN is a kind of memory that stores information about the past sequence elements. It is updated based on the current input and the previous hidden state at each time step. The hidden state serves as a kind of memory that helps the network capture dependencies and patterns in sequential data [56].

Training RNNs can be challenging due to issues like vanishing gradients or exploding gradients. These problems occur when the gradients become too small or too large during backpropagation through time. Long sequences exacerbate these problems, making it difficult for the network to learn long-term dependencies. Different architectural variations of RNNs have been proposed to address some of the challenges. Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) are two popular types of RNNs designed to mitigate the vanishing gradient problem and capture long-term dependencies more effectively [57]. While RNNs are powerful for sequential data, they have limitations. Training can be computationally expensive, and they may struggle with capturing long-term dependencies in practice. More advanced architectures like transformers have gained popularity for specific tasks due to their ability to capture long-range dependencies more efficiently [58].

Mekruksavanich et al. addressed the HAR problem [59]. The suggested an LSTM-based framework that explores the LSTM network and offers good performance. Four LSTM networks have been used to examine the performance of recognition utilizing various smartphone sensors (tri-axial gyroscope and tri-axial accelerometer). Utilizing the publicly accessible dataset UCI-HAR, such LSTM networks were assessed in terms of prediction accuracy in addition to other performance parameters like recall, precision, AUC, and F1-score. The experimental results demonstrate that, with a high rate of accuracy of 99.39%, the four-layer CNN-LSTM network developed in this paper beats the other basic LSTM networks. Additionally, the suggested LSTM network was contrasted with earlier efforts. The accuracy might be increased by up to 2.24% using the four-layer CNN-LSTM network. CNN layers in this model execute direct mapping in spatial representation regarding raw sensor data for the process of feature extraction, which is an advantage. This study's DL systems are trained and evaluated using the lab data, which is a drawback because prior research has shown that the learning algorithms' performance in a lab environment cannot be accurately compared to performance in the real world. The recommended HAR architecture could also be employed with high performance DL networks for a variety of useful applications in smart homes, including optimum human mobility in sports, safety surveillance and healthcare monitoring for elderly persons, and child and newborn care.

3.2.3. Generative adversarial networks (GAN)
Generative Adversarial Networks (GANs) are a class of artificial intelligence algorithms. They were introduced by Ian Goodfellow and his colleagues in 2014. GANs are designed to generate new data samples that resemble a given dataset. The fundamental idea behind GANs is to have two neural networks, a generator and a discriminator, which are trained simultaneously through adversarial training. The training process involves a continuous feedback loop between the generator and the discriminator. The process continues iteratively, with the generator and discriminator getting better at their tasks over time. Ideally, this results in a generator that can create realistic data samples and a discriminator that is challenged to differentiate between real and generated samples [60]. Hoelzemann.et al. provided an algorithm and contrast of two augmentation techniques, fold-wise, and user-wise, to increase the size of a dataset, which is displayed below on the PAMAP2 dataset using an arbitrary number of synthetic samples [61]. The artificial time-series data is produced using a recurrent Generative Adversarial Network (GAN), where the generator and discriminator are modeled by a collection of LSTM cells. Supervised learning was used to train four DeepConvLSTM models. This network has a validation F1 score of 96% after 200 training epochs.

Table 1. Summary of related works

Authors	Dataset	Method	Feature Extraction	Feature Selection	Accuracy
Gamal et al. [30]	MHEALTH	RF	23 features, which contain motion and physiological data (ECG).	-	96.6%, 99.1%
Randhawa et al. [31]	Real-Time	SVM	Time Domain Feature	-	98.8%
Fan et al. [35]	UCI-HAR	DQN	Time-domain, Spectral Domain Feature	hybrid feature selection methodology, BAROQUE	98.41%
Jamel et al. [40]	UCI	KNN	Time, Frequency, and Statistical Domain features	PCA	90%
Sheng et al. [41]	PAMAP2, REALDISP, SBHAR	k-means	Mean, max, min, Median, Standard deviation, Interquartile range	-	92.11%, 71.49%, 80.73%
Zhang et al. [43]	German Credit	HAR-AWDF	attributes of the credit dataset (A14, A15, A16)	-	92%
Aljarrah [46]	MHEALTH	RNN	auto	PCA	97.64%
Oyedotun et al. [48]	WiSDM	CNN	auto	-	91.33%
Xu al. [53]	WiSDM	CNN	auto	-	91.97%
Bianchi et al. [54]	Collected	CNN	auto	-	97%
Mekruksavanich et al. [59]	UCI-HAR	CNN-LSTM	Auto	-	99.39%
Hoelzemann.et al. [61]	PAMAP2	GAN	Auto	-	F1- score 96%

4. Discussion

Recognizing human activities is a challenging task that is frequently addressed through the application of various machine learning and deep learning approaches. Here, the strengths and weaknesses of different methods in the context of recognizing human activities will be discussed, taking into account both traditional machine learning and deep learning approaches. In human activity recognition, Supervised Learning proves to be a powerful tool, leveraging labeled datasets for accurate model training to recognize specific activities. This approach is particularly advantageous for tasks requiring precise classification, as the model learns from explicit examples during training, enabling it to generalize well to new instances. However, it's crucial to acknowledge the challenges associated with supervised learning in this context. Acquiring large amounts of labeled data for diverse human activities can be an expensive and time-consuming endeavor. Additionally, the model may struggle when faced with activities that are not well represented or adequately captured in the training data, emphasizing the importance of dataset diversity and completeness in achieving robust human activity recognition models. Unsupervised Learning offers a valuable approach to uncovering patterns and relationships in data without the need for labeled examples. This is particularly advantageous for exploring unknown or novel human activities, as the model does not depend on pre-existing labeled data. The absence of a requirement for extensive human annotation is a notable strength, streamlining the training process. However, it's important to acknowledge that working without labeled data presents challenges, especially in accurately classifying activities. Evaluating the performance of unsupervised models can also be complex due to the absence of a clear ground truth, making it essential to employ alternative methods for assessing effectiveness in the absence of labeled examples.

Reinforcement learning stands out due to its aptitude for handling sequential decision-making processes. This characteristic strength allows the model to adapt to dynamic environments and learn from interactions over time. However, it's essential to consider that the efficacy of reinforcement learning in human activity recognition is contingent upon a substantial number of interactions with the environment. In practical, real-world applications, the requirement for a large volume of interactions can be deemed impractical or resource-intensive. Furthermore, achieving a delicate balance between exploration and exploitation in reinforcement learning can pose a challenge, potentially leading to suboptimal behavior if not carefully managed. Convolutional Neural Networks (CNNs) excel in capturing spatial hierarchies, making them highly effective for image-based human activity recognition. Their ability to exhibit invariance to translations in the input adds robustness, enabling them to handle variations in the position of objects within a scene. However, challenges arise when CNNs capture temporal dependencies inherent in sequential human activities. Unlike Recurrent Neural Networks (RNNs), which are specifically designed for such tasks, CNNs may face limitations in modeling dynamic temporal aspects. Additionally, it's important to note that CNNs benefit significantly from a substantial amount of labeled image or video data for effective training in the context of human activity recognition.

Recall that Recurrent Neural Networks (RNNs) are particularly well-suited for tasks involving sequential data, making them an apt choice for human activity recognition, which often exhibits temporal dependencies. Their ability to handle variable-length input sequences renders RNNs flexible, accommodating the diverse durations associated with various human activities. However, it's crucial to acknowledge challenges in training RNNs, such as the potential for vanishing or exploding gradients, which can impact the model's efficacy in capturing long-term dependencies. Additionally, it's worth noting that RNNs can be computationally intensive, particularly when dealing with extended sequences, potentially requiring careful optimization for efficient human activity recognition. Generative Adversarial Networks (GANs) offer a promising avenue for improving human activity recognition through data augmentation. By leveraging GANs, synthetic examples can be generated, augmenting training data and addressing limitations posed by the scarcity of labeled real-world datasets. GANs demonstrate an ability to create realistic synthetic data, a valuable asset when genuine labeled datasets for human activity recognition are restricted. However, it is important to acknowledge the challenges associated with training GANs, including concerns about stability. One potential issue is the risk of limited diversity in the synthetic data, potentially resulting in mode collapse, which could impact the robustness of human activity recognition models trained on such data.

5. Conclusions

In the presented study, many data processing techniques that were employed in this field of expertise were utilized, along with their advantages and disadvantages for sensor-based HAR; due to their potential incorporation in new activity identification applications, such techniques have gained considerable traction in recent years. A discussion from many angles on the previous work, such as feature design and models. We focused on recent advancements in both sensor-based HAR and examined the traits, benefits, and drawbacks of classical ML and DL models employed in HAR. The usefulness of algorithms depending on DTs, RF, NB, and SVM stands out for the particular supervised learning case. The majority of the examined studies used approaches like Dimensionality Reduction, Association Rules Learning, and Clustering for unsupervised learning. Using DL, another contemporary and popular approach, stands out for using the following algorithms: CNN, RNN, and GAN. System design, activity tracking, and speed are a few of the topics that need more research, along with the widespread usage of activity recognition in pattern recognition. This study is expected to spark more research in the area of activity recognition.

Acknowledgments: The authors would like to thank the Computer Science Department, College of Science and Al-Mustansiriyah University for their support.

Funding: This work is self-funded.

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

References

[1] Ahmadi, M.N.; Pfeiffer, K.A.; Trost, S.G.; “Physical Activity Classification in Youth Using Raw Accelerometer Data from the Hip,” Meas. Phys. Edu. Exe. Sci. 24(2): 129–136, 2020.

[2] Wang, Y.; Cang, S.; Yu, H.; “A survey on wearable sensor modality centered human activity recognition in health care,” Exp. Sys. Appl. 137: 167–190, 2019.

[3] Voicu, R.-A.; Dobre, C.; Bajenaru, L.; Ciobanu, R.-I.; “Human Physical Activity Recognition Using Smartphone Sensors,” Sensors 19(3): 458, 2019.

[4] Minh Dang, L.; Min, K.; Wang, H.; Md. Jalil P.; Lee, C.H.; Moon, H.; “Sensor-based and vision-based human activity recognition: A comprehensive survey,” Pat. Recogn. 108: 107561, 2020.

[5] Jung, M.; Chi, S.; “Human activity classification based on sound recognition and residual convolutional neural network,” Auto. Const. 114: 103177, 2020.

[6] Ma, M.; Marturi, N.; Li, Y.; Leonardis, A.; Stolkin, R.; “Region-sequence based six-stream CNN features for general and fine-grained human action recognition in videos,” Patter. Recogn. 76: 506–521, 2018.

[7] Vidya, B.S.P, “Wearable multi-sensor data fusion approach for human activity recognition using machine learning algorithms,” Sens. and Actu. A: Physical 341: 113557, 2022.

[8] Jordao, A.; “Human Activity Recognition Based on Wearable Sensor Data: A Standardization of the State-of-the-Art,” arXiv.org, 2018. doi: 10.48550/arXiv.1806.05226.

[9] Dang L.M.; Piran, M.; Han, D.; Min, K.; Moon H. et al., “A survey on internet of things and cloud computing for healthcare”, Electronics 8 (7): 768, 2019.

[10] Alshwely, M.K.; AlSaad, S.N.; “Image splicing detection based on noise level approach,” Al-Mustansiriyah J. Sci. 31(4): 55–61, 2020.

[11] Maharana, K.; Mondal, S.; Nemade, B.; “A review: Data pre-processing and data augmentation techniques,” Glob. Trans. Proc. 3(1): 91–99, 2022.

[12] Chen, L.; Nugent, C.D.; Chen, L.; Nugent, C.D.; “Time-Window Based Data Segmentation. Human Activity Recognition and Behaviour Analysis”: Cyber-Physical Systems in Smart Environments, pp.103-126, 2019.

[13] Wang, Y.; Li, M.; Shang, Y.; Li, N.; Zhao, Y.; Shang, Z.; "Analysis and Classification of Time Domain Features of Pulse Manifestation Signals between Different Genders". 2019 Chinese Automation Congress (CAC), Hangzhou, China, 2019, pp. 3981-3984, doi: 10.1109/CAC48633.2019.8997260.

[14] Fan, C.; Gao, F.; “Enhanced Human Activity Recognition Using Wearable Sensors via a Hybrid Feature Selection Method,” Sensors 21(19): 6434, 2021.

[15] Gaikwad, N.B.; Tiwari, V.; Keskar, A.; Shivaprakash, N.C.; ”Efficient FPGA implementation of multilayer perceptron for real-time human activity classification”. IEEE Access, 7: 26696–26706, 2019.

[16] D’Arco, L.; Wang, H.; Zheng, H.; ”Assessing impact of sensors and feature selection in smart-insole-based human activity recognition”. Meth. Proto. 5(3): 45, 2022.

[17] Guo, J.; Mu, Y.; Xiong, M.; Liu, Y.; Gu, J.; ”Activity feature solving based on TF-IDF for

activity recognition in smart homes", Complexity: 2019.
[18] Venkatesh, B.; Anuradha, J.; "A Review of Feature Selection and Its Methods," Cyber. Info. Technol. 19(1); 3–26, 2019.
[19] Khaire, U.M.; Dhanalakshmi, R.; "Stability of feature selection algorithm: A review," J. King Saud Uni. Comp. Info. Sci. 34(4): 1060–1073, 2022.
[20] Tawhid, M.A.; Ibrahim, A.M.; "Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm," Int. J. Mach. Lear. Cyber. 11(3): 573–602, 2019.
[21] Effrosynidis, D.; Arampatzis, A.; "An evaluation of feature selection methods for environmental data," Ecol. Info. 61: 101224, 2021.
[22] Saha, A.; Sharma, T.; Batra, H.; Jain, A.; Pal, V.; "Human Action Recognition Using Smartphone Sensors," 2020 International Conference on Computational Performance Evaluation (ComPE), Jul. 2020, Published, doi: 10.1109/compe49325.2020.9200169.
[23] Dewi, C.; Chen, R.-C.; "Human Activity Recognition Based on Evolution of Features Selection and Random Forest," 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 2019, pp. 2496-2501, doi: 10.1109/SMC.2019.8913868.
[24] Nguyen, H.D.; "Wearable Sensor Data Based Human Activity Recognition using Machine Learning: A new approach," arXiv.org, May 09, 2019. https://doi.org/10.48550/arXiv.1905.03809
[25] Asmita, N.; Jayita, S.; Chandreyee, C.; Kundan, P.D.S.; "Detailed Human Activity Recognition using Wearable Sensor and Smartphones," International Conference on Opto-Electronics and Applied Optics, IEEE Conference Publication 2019. https://doi.org/10.1109/OPTRONIX.2019.8862427.
[26] Athavale, V.A.; Gupta, S.C.; Kumar, D.; Savita; "Human Action Recognition Using CNN-SVM Model". Adv. Sci. Technol. 105: 282–290, 2021.
[27] Hossain Shuvo, M.M.; Ahmed, N.; Nouduri, K.; Palaniappan, K.; "A Hybrid Approach for Human Activity Recognition with Support Vector Machine and 1D Convolutional Neural Network," 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington DC, DC, USA, 2020, pp. 1-5, doi: 10.1109/AIPR50011.2020.9425332.
[28] Loris, N.; Matteo, I.; Sheryl, B.; Christian, S.; Sergio, P.; Raffaello, N.; Isabella, C.;, "Comparison of Transfer Learning and Conventional Machine Learning Applied to Structural Brain MRI for the Early Diagnosis and Prognosis of Alzheimer's Disease," Frontiers in Neurology, vol. 11, Nov. 2020.
[29] Burkart, N.; Huber, M.F.; "A Survey on the Explain ability of Supervised Machine Learning," J. Art. Intel. Res. 70: 245–317, 2021.
[30] Gamal, G.; Omar, Y.M.; Maghraby, F.A.; "Machine Learning Approaches for Human Activity Recognition Based on Multimodal Body Sensors". In Intelligent Systems Design and Applications: 20th International Conference on Intelligent Systems Design and Applications (ISDA 2020) held December 12-15, 2020 (pp. 977-987). Cham: Springer International Publishing.
[31] Randhawa, P.; Shanthagiri, V.; Kumar, A.; Yadav, V.; "Human activity detection using machine learning methods from wearable sensors". Sensor Review 40(5): 591-603, 2020.
[32] Levine, S.; "Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems," arXiv.org, 2020. https://doi.org/10.48550/arXiv.2005.01643.
[33] Lorenzo, C.; Gian, C.C.; Luca, D.N.; Rocco, F.; Daniele, G.; Marco, R.; Sergio, S.; "Multi-Agent Reinforcement Learning: A Review of Challenges and Applications". Appl. Sci. 11(11): 4948, 2021.
[34] Sewak, M.; "Deep Q Network (DQN), Double DQN, and Dueling DQN," Deep Reinforcement Learning: 95–108, 2019.
[35] Fan, C.; Gao, F.; "Enhanced Human Activity Recognition Using Wearable Sensors via a Hybrid Feature Selection Method," Sensors 21(19): 6434, 2021.
[36] Kalmár, G.; "Hardware-Software Co-development for Audio and Video Data Acquisition and Analysis" (Doctoral dissertation, Szte).
[37] Berry, M.W.; Mohamed, A.; Yap, B.W.; "Supervised and Unsupervised Learning for Data Science". Springer Nature, 2019.
[38] Manduchi, L.; Hüser, M.; Faltys, M.; Vogt, J.; Rätsch, G.; Fortuin, V.; "T-DPSOM," Proceedings of the Conference on Health, Inference, and Learning, Apr. 2021, Published, doi: 10.1145/3450439.3451872.

[39] Zhao, R.; Ruan, J.; Dong, B.; Meng, L.; Zhang, W.; "Improving unsupervised image clustering with spatial consistency," Knowledge-Based Systems 246: 108673, 2022.

[40] Jamel, A.A.; Akay, B.; "Human Activity Recognition Based on Parallel Approximation Kernel K-Means Algorithm. Comput". Syst. Sci. Eng. 35(6): 441-456, 2020.

[41] Sheng, T.; Huber, M.; "Unsupervised embedding learning for human activity recognition using wearable sensor data". in The Thirty-Third International Flairs Conference, 2020.

[42] Telikani, A.; Gandomi, A.H.; Shahbahrami, A.; "A survey of evolutionary computation for association rule mining," Info. Sci. 524: 318–352, 2020.

[43] Zhang, Y.; Huang, Q.; Zhao, K.; "Hybridizing association rules with adaptive weighted decision fusion for personal credit assessment". Sys. Sci. Contr. Eng. 7(3): 135-142, (2019).

[44] Espadoto, M.; Martins, R.M.; Kerren, A.; Hirata, N.S. T.; Telea, A.C.; "Toward a Quantitative Survey of Dimension Reduction Techniques". IEEE Trans. Vis. Comp. Graph. 27(3): 2153-2173, 2021.

[45] Kherif, F.; Latypova, A.; "Principal component analysis," Mach. Lear.: 209–225, 2020.

[46] Aljarrah, A.A.; Ali, A.H.; "Human Activity Recognition using PCA and BiLSTM Recurrent Neural Networks" in 2019 2nd International Conference on Engineering Technology and its Applications (IICETA), 2019.

[47] Kelleher, J.D.; Deep Learning. MIT Press, 2019.

[48] Oyedotun, O.K; Khashman, A.; "Deep learning in vision-based static hand gesture recognition", Neur. Comp. Appl. 28(12): 3941–3951, 2017.

[49] Chen, C.; Jafari, R.; Kehtarnavaz, N.; "A survey of depth and inertial sensor fusion for human action recognition", Multimed. Tool. Appl. 76(3): 4405–4425, 2017.

[50] Zhang, Q.; Yang, L.T.; Chen, Z.; Li, P.; "A survey on deep learning for big data", Info. Fusion (42): 146–157, 2018.

[51] Mohammadi, M.; Al-Fuqaha, A.; Sorour, S.; Guizani, M.; "Deep learning for IoT big data and streaming analytics: a survey", IEEE Comm. Surv. Tutor. 20(4): 2923–2960, 2018.

[52] Khassaf, N.M.; Shaker, S.H.; "Image Retrieval based Convolutional Neural Network," Al-Mustansiriyah J. Sci. 31(4): 43–54, 2020.

[53] Xu, W.; Pang, Y.; Yang, Y.; Liu, Y.; "Human activity recognition based on convolutional neural network". In 2018 24th international conference on pattern recognition (ICPR) (pp. 165-170) (2018). IEEE.

[54] Bianchi, V.; Bassoli, M.; Lombardo, G.; Fornacciari, P.; Mordonini, M.; De-Munari, I.; "IoT wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment". IEEE Inter. Thin. J. 6(5): 8553-8562, 2019.

[55] Haqi Al-Tai, M.; Nema, B.M.; Al-Sherbaz, v; "Deep Learning for Fake News Detection: Literature Review," Al-Mustansiriyah J. Sci. 34(2): 70–81, 2023.

[56] Sherstinsky, A.; "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," Physica D: Nonlin. Phenomena 404: 132306, 2020.

[57] Schmidt, R.M.; "Recurrent Neural Networks (RNNs): A gentle Introduction and Overview," arXiv.org, 2019. https://doi.org/10.48550/arXiv.1912.05911.

[58] Bouarara, H.A.; "Recurrent Neural Network (RNN) to Analyse Mental Behaviour in Social Media," Int. J. Soft. Sci. Comp. Intel. 13(3): 1–11, 2021.

[59] Mekruksavanich, S.; Jitpattanakul, A.; "Lstm networks using smartphone data for sensor-based human activity recognition in smart homes". Sensors 21(5): 1636, 2021.

[60] Ian, G.; Jean, P.A.; Mehdi, M.; Bing, X.; David, W.F.; Sherjil, O.; Aaron, C.; Yoshua, B.; "Generative adversarial networks," Communications ACM 63(11): 139–144, 2020.

[61] Hoelzemann, A.; Sorathiya, N.; Van-Laerhoven, K.; "Data Augmentation Strategies for Human Activity Data Using Generative Adversarial Neural Networks," in 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops), 2021.