IRAQI Academic Scientific Journals
العراقية المجلات الاكاديمية العلمية
University of Kut Journal
ISSN (E): 2616 - 7808 II ISSN (P): 2414 - 7419
www.kutcollegejournal.alkutcollege.edu.iq
k.u.c.j.sci@alkutcollege.edu.iq
Vol. 10 , Issue 2 , December 2025

Human Activity Recognition Using Deep Learning: A Review of the Last Ten Years

Haider Rasheed Hassan [1] , Dheyab Salman Ibrahim [2] , Ziyad Tariq Mustafa Al-Ta'i [3]

Abstract

This review systematically examines recent advances in Human Activity Recognition (HAR) enabled by deep learning techniques, which play a critical role in improving human–machine interaction across applications such as healthcare, surveillance, and intelligent environments. The study analyzes peer-reviewed publications published between 2014 and 2024 to trace the evolution of HAR from traditional approaches to state-of-the-art deep learning models. It provides a comprehensive overview of commonly used datasets for both vision-based and sensor-based HAR systems, as well as deep learning architectures including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and hybrid models. The findings indicate a clear shift toward sensor-based HAR systems due to the flexibility and non-intrusive nature of wearable and ambient sensors. Moreover, the review highlights the capability of deep learning models to handle complex and unstructured data, resulting in notable improvements in recognition accuracy. Finally, this study identifies key research challenges and future directions, including data privacy concerns, model generalization, and the optimization of deep learning models for deployment on resource-constrained devices. Overall, this review offers a structured analysis of datasets, methodologies, and performance trends, providing valuable insights and a clear roadmap for future research in the HAR field.

Keywords: Deep Learning, Human Activity Recognition, HAR, vision-based, sensor-based

Affiliation of Authors
[1,2,3] College of Science, University of Diyala, Iraq, Diyala, 32000

[1] scicompms222307@uodiyala.edu.iq
[2] dr.dheyab@uodiyala.edu.iq
[3]ziyad1964tariq@uodiyala.edu.iq

[1] **Corresponding Author**

Paper Info.
Published: Dec. 2025

التعرّف على الأنشطة البشرية باستخدام التعلّم العميق: مراجعة لأبحاث السنوات العشر الماضية

حيدر رشيد حسن [1] ، ذياب سلمان ابراهيم [2] ، زياد طارق مصطفى الطائي [3]

المستخلص

تستعرض هذه الدراسة بشكل منهجي أحدث التطورات في مجال التعرف على الأنشطة البشرية (Human Activity Recognition – HAR)باستخدام تقنيات التعلم العميق، والتي تُعد عنصرًا أساسيًا في تحسين التفاعل بين الإنسان والآلة في تطبيقات متعددة مثل الرعاية الصحية، والمراقبة، والبيئات الذكية. تعتمد الدراسة على تحليل منهجي للأبحاث العلمية المحكمة المنشورة خلال الفترة من **2014إلى 2024**، بهدف تتبع تطور تقنيات التعرف على الأنشطة البشرية من الأساليب التقليدية إلى نماذج التعلم العميق المتقدمة. كما تقدم المراجعة عرضًا شاملًا لمجموعات البيانات المستخدمة في أنظمة HAR المعتمدة على الرؤية الحاسوبية والمستشعرات، إلى جانب استعراض معماريات النماذج العميقة مثل الشبكات العصبية الالتفافية(CNN) ، والشبكات العصبية التكرارية (RNN)، وشبكات الذاكرة طويلة وقصيرة الأمد(LSTM) ، والنماذج الهجينة. تشير النتائج إلى وجود توجه متزايد نحو أنظمة HAR المعتمدة على المستشعرات، نظرًا لمرونتها وطبيعتها غير التدخلية عند استخدام المستشعرات القابلة للارتداء أو المحيطية. بالإضافة إلى ذلك، توضح الدراسة قدرة نماذج التعلم العميق على معالجة البيانات المعقدة وغير المهيكلة، مما أدى إلى تحسينات ملحوظة في دقة التعرف على الأنشطة. كما تحدد هذه المراجعة عددًا من التحديات البحثية والاتجاهات المستقبلية، مثل قضايا خصوصية البيانات، وتعزيز قابلية تعميم النماذج، وتحسين كفاءة النماذج لتناسب العمل على الأجهزة ذات الموارد المحدودة. وبشكل عام، تقدم هذه الدراسة تحليلًا شاملًا لمجموعات البيانات والمنهجيات واتجاهات الأداء، مما يوفر رؤى علمية قيمة وخارطة طريق واضحة للأبحاث المستقبلية في مجال التعرف على الأنشطة البشرية.

الكلمات المفتاحية: التعلم العميق، التعرف على النشاط البشري، HAR، القائم على الرؤية، القائم على أجهزة الاستشعار

انتساب الباحثين
[1, 2, 3] كلية العلوم, جامعة ديالى, االعراق ، ديالى، 32000

[1] scicompms222307@uodiyala.edu.iq
[2] dr.dheyab@uodiyala.edu.iq
[3]ziyad1964tariq@uodiyala.edu.iq

[1] **المؤلف المراسل**

معلومات البحث
تأريخ النشر: كانون الاول 2025

1. Introduction

We, as humans, do a lot and our daily activities can talk much about us. Imagine if you know that someone starts his day by running for half an hour and spends 2 hours training each day. The image of that person would be either he is an athlete or at least he is in good shape, following a healthy lifestyle; that's information or assumption you made by only knowing a little of his daily activity style, and that is what raises the importance of human activity recognition as knowing more will undoubtedly provide a better understanding. The development of HAR systems based on the deep learning paradigm represents a revolution in human-machine interaction that currently covers areas such as healthcare [1], surveillance [2], smart homes [3], and more.

HAR methodologies predominantly fall into two categories: vision and sensor approaches. Vision-based HAR focuses on visual data from cameras and is successful in situations where visual cues are apparent. On the other hand, Sensor-based HAR uses data from multiple sensors and provides flexibility and practicality in situations where visual surveillance is not applicable or invasive.

Due to the fast-advancing area of deep learning, within the group of machine learning, HAR systems accuracy and efficiency has latterly taken a step forward. The feature that it possesses to learn from data and to analyze big data gives a way to the better diagnose of human behavior patterns; the deep learning has achieved a groundbreaking growth when it is about the video and photo processing which also lies in the vision-based HAR. CNN algorithms emerged as the competent machine vision tool, because of their superior ability to utilize pixel data from images and videos that mostly occur in real life. The organization and analysis of activity intelligence have led to more precise identification and right timing [4].

Sensor-based HAR relies on deep learning models that can combining data from several different sensors such as accelerometers or gyroscopes and even from heart rate monitors, for more comprehensive perceptions regarding human activities. RNNs and their sub-types such as LSTM, are the key ones in this case. As RNN models works very well in processing time-series data like complex motions sampled by sensors.

Furthermore, deep learning has enabled researchers to design more precise and flexible HAR systems [5].

This review article focuses on research using deep learning approaches in vision-based and sensor-based HAR by methodically going over more than 50 research publications between 2014 -2024. It explores the methods, findings, and innovations mentioned in these works, offering a thorough summary of the state-of-the-art; the aim of this work can be summarized as follows:

1. Exploring classification techniques for both sensor and vision-based HAR.
2. Organizing the frequently used datasets based on their year of evolution, mode of representation, classes, and data type.
3. Chronologically summarizing the related research articles by comparing their underlying architecture, datasets, data type, and accuracy.
4. Identifying potential research gaps and future directions concerning the deep learning HAR systems.

The paper is organized as follows: the research methodology is discussed in Section 1. An overview of deep learning, including its definition and significance and architecture related to HAR, is demonstrated in Section 3. Section 4 introduces

various HAR datasets, their classification, and hierarchical tabular representation with the specification. Section 5 outlines the classification techniques used in HAR. The performance of various significant articles is summarized in Section 6. The challenges and multiple aspects of future directions are briefly expressed in section 7. In section 8, the contributions and practical implications of our work are briefly discussed. Finally, the conclusion in section 9, followed by the references.

2. Methodology

We followed Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols [6] to single out relevant and significant articles related to our research domain. We accomplished this review by adopting three protocols: searching protocol, inclusion and exclusion protocol, and scoping review protocol.

2.1 Searching Protocol

Initially, we established our research framework by selecting appropriate digital platforms, including search sites, libraries, and databases. Most articles referenced in this review were obtained from popular online libraries such as Google Scholar, Web of Science, and IEEE Explore. We employed precise and relevant search terms to identify pertinent literature, individually or in combination. Key phrases included "human activity recognition," "HAR," "video action classification," "sensor-based HAR," "deep learning," "CNN," and various terms related to activity recognition databases. Additionally, we explored queries that integrated multiple keywords to yield more meaningful results. For the first phase of our research, we downloaded more than 250 articles for further examination.

2.2 Inclusion and Exclusion Protocol

We included relevant, English, peer-reviewed vision-based, and sensor-based activity recognition articles that adopt deep learning techniques for model designing. Meanwhile, non-English, non-peer-reviewed, and non-pertinent to our research were excluded. We considered the date (2014-2024), publication type (journal or conference), publishing house, and cite score during preliminary screening. Furthermore, we extended this screening procedure to the abstract composition level, where we validated the themes of the searched articles against our survey theme. Finally, we filtered out the 100 most significant articles for further review.

2.3 Reviewing and Data Extraction

In the final part of our methodology, we carefully considered a wide range of contextual aspects before conducting a systematic review of the chosen papers. We started by organizing the summary according to the abstract section's context, objective, evidence source, eligibility requirement, databases, model algorithms, results, and conclusion. Then, we proceeded into detailed illustration, taking into account the elements above, as well as some other specifics, such as computational complexity, the potential for real-time deployment, constraints, research gaps, and opportunities as shown in Figure (1).

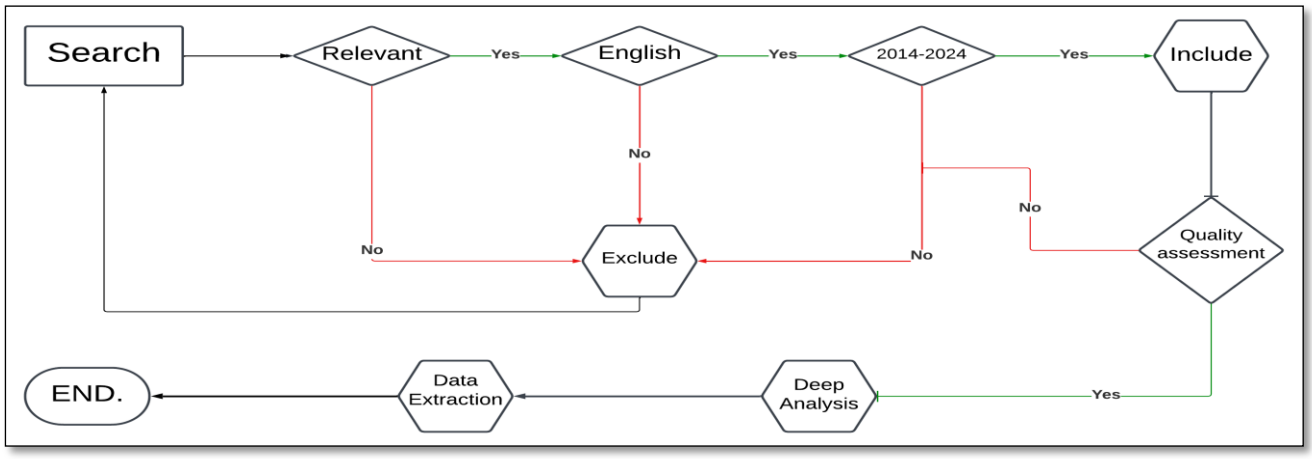

Figure (1): State the flow of our methodology

3. Overview

3.1 What is Deep Learning?

Deep learning is a subcategory of machine learning that is also part of artificial intelligence. It is based on algorithms similar to the structure and function of the brain's neural networks. Deep learning algorithms have multiple layers of processing that help extract and transform features from the data. These layers represent a hierarchical order of progressively more elaborate and abstract features.

The core characteristics of deep learning include:

1. Many layers of nonlinear processing units are used for feature extraction and transformation.
2. Each layer uses the output as input to the next layer.
3. The capacity of learning several levels of representations that represent varying degrees of abstraction.
4. The algorithms can be either supervised or unsupervised.

Deep learning models are beneficial for discovering complex structures in high dimensional data in fields like computer vision, automatic speech recognition, natural language processing, and bioinformatics, for which they have given remarkable results [7]. Since these models automatically find the necessary representations for feature detection or classification from raw data, human intervention is no longer required in feature extraction.

3.2 Why Deep Learning?

Since deep learning techniques have the best success rate in real-world situations, they have gained popularity in the HAR area. Unlike machine learning, which uses handcrafted feature extraction, deep learning provides a foundation for automatic feature selection and learning. Among the most used deep learning approaches are generative adversarial networks (GAN), recurrent neural networks (RNN), auto-encoders, Aslam and Kolekar (2022), restricted Boltzmann machines (RBM), and deep neural networks (DNN and CNN). These methods may be applied to clustering, regression, classification, and detection or identification of human activity. Section 5 provides a detailed discussion of several categorization and detection techniques.

3.3 HAR Framework

3.3.1 Data collection

The first phase of deep learning HAR models is

the data collection which includes collecting data from multiple sensors such as cameras, accelerometers, and gyroscopes in order measure human activities in different scenarios ,which raises the need for producing large amount and high-quality data. Then, the data is organized in a deep learning suitable way accounting for live or past use among real-time and historical usage. Techniques such as data augmentation and sensor fusion are utilized to increase the dataset's variety and depth, thereby improving the model's resilience and accuracy in identifying human activities.

3.3.2 Preprocessing

In this phase, sensors' raw data are subjected to a number of transformations in order to prepare them for analysis which includes normalization to standardize different data types, segmentation in order to split continuous data streams into tolerable parts, and filtering to eliminate noise. On the other hand, manual feature selection can be used to reduce data dimensionality along with some techniques of data augmentation, such as rotation or adding noise that aims at increasing model robustness. Processing missing data through the imputation practice, as well as time-series analysis techniques, is also crucial in terms of temporal data for achieving a quality dataset that is cleared up and structured to allow acceptable learning by the chosen model.

3.3.3 Model architecture

Choosing the right deep learning model architecture is essential for the HAR system; deep learning model architecture typically involves Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). CNNs are used for processing spatial data like images or videos, as CNNs excel in identifying spatial patterns in visual data, making them suitable for vision-based activity recognition. In contrast, RNNs and LSTMs are designed to handle the temporal dynamics and dependencies in sequential data, such as accelerometer readings, making them ideal for recognizing activities based on motion patterns over time (sensor-based activity). The choice between these architectures, or a combination thereof, depends on the nature of the data and the specific requirements of the activity recognition task.in section 4, we illustrate the different deep learning algorithms used with HAR systems for both vision and sensor-based HAR systems, providing a comprehensive and systematic review.

3.3.4 Model Training

Deep learning model training is a challenging task that needs to be planned carefully. Still, with the right approach and appropriate resources, deep learning models are capable of producing excellent results on multiple tasks. At this point, data has been prepared and the architectural framework of the model is selected. For selection of the loss function (MSE for regression and categorical cross-entropy for classification tasks), the loss function will define the robustness of the model in dealing with the task. In second order, one should pick SGD (stochastic gradient descent), Adam (fundamental method), or RMS prop (root mean square propagation) algorithm to fine-tune the rate of convergence and parameter update performance while training the model which improves the training. Also, the over-fitting problems may be solved using regularization techniques like dropout or L2 regularization, which it will not be doubtful to the model to generalize the new data. So, the last issue is a continuous observation of the

model's training and validation, and contingent methods for example but not limited to early stopping to avoid over-fitting and achieve the model with the best checkpoints.

3.3.5 Model Evaluation

Evaluation of deep learning models for HAR involves careful caution when using a different technique and methodology. As a matter of fact, it is aimed at exhaustive accuracy check as well as stability and generality, which help to identify its potential in the health care environment as the reliable tool for human activity recognition. The list of confusions matrix is an irreplaceable instrument which works out the model predictions result down to the finest possible detail. It allows the advanced end-user to depict functioning i.e. modeling of data for various operations, while highlighting domains where the technique gives positive or negative results, with mild confusion between similar activities. Ranking is carried out on a multi-dimensional scale. Accuracy, precision, recall, and F1 score are the set metrics that is mostly used. Accuracy measures the overall correctness of the model, while precision reflects whether optimistic predictions are accurate; recall assesses a model's ability to identify all relevant cases, and F1 is a balance between precision and recall. Generally, there are two types of labels: prediction labels and truth labels. Truth labels are the actual class label (or ground truth) to which that sample belongs, whereas prediction labels are model-predicted tags during validation or testing. These definitions state that a sample is considered true positive (TP) if the ground truth and predictive labels are positive, and a sample is considered true negative (TN) if the ground truth and predictive labels are negative. Prediction labels are classified as false positives (FP) if the ground truth label is negative and the prediction label is positive or false negatives (FN) if the ground truth label is positive and the prediction label is negative. The most used metrics are listed and the equations of the data is given by:

$$\textbf{Accuracy} = \frac{\textbf{TP} + \textbf{TN}}{\textbf{TP} + \textbf{FP} + \textbf{TN} + \textbf{FN}} \textbf{x100\%} \quad (\mathbf{1})$$

$$\textbf{Recall} = \frac{\textbf{TP}}{\textbf{TP+FN}} \quad (\mathbf{2})$$

$$\textbf{Precision} = \frac{\textbf{TP}}{\textbf{TP+FP}} \quad (\mathbf{3})$$

$$\textbf{F1 Score} = \frac{\textbf{2x(Precision x Recall)}}{\textbf{Precision+ Recall}} \quad (\mathbf{4})$$

Other than these, some papers mention the Area Under the Curve (AUC) metric. We consider only accuracy as the performance measure when sorting deep learning papers from 2014 to 2024 as shown in Table (2).

4. Datasets

Human Activity Recognition (HAR) is a cutting-edge and progressive branch of data science and artificial intelligence with great application in medical spheres, sports sciences, ergonomics, and personal fitness. At its very foundations, HAR involves activity recognition and classification of physical activities- performed by individuals through the data obtained primarily via multiple sensors. These activities may range from mundane, lightly strenuous tasks to vigorous physical efforts. HAR algorithms and systems has been greatly improved by sharing of datasets, datasets compilation, and provision of datasets to the research community for the same purpose. Those data which frequently are with the times series data form sensors like accelerometers and gyroscopes are very important and useful to know the various human motion and its behaviors across

different sports. To perform this purpose they pay attention to training machine learning algorithms, that can separate and categorize a wide range of human actions.

Here we design this classification protocol so as to categorize the datasets into five levels according to the activity intensity to give a unified way for studying and exploitation of them. In consequence, the categorization is not only appreciated from the data analyses point of view but also this categorization becomes more important for health monitoring & fitness tracking as well as behavioral studies. The five levels are:

1. **Sedentary Activities:** This category includes datasets that capture minimal physical movement and contain actions such as sitting and standing.
2. **Light Activities:** Data sets in this category can include activities such as gentle lifting, strolling, light housework.
3. **Moderate Activities:** This level is for exercises that take more energy than light activities but are not very strenuous, such as brisk walking or moderate cycling.
4. **Vigorous Activities:** This category consists of collections of activities such as running, fast riding, and more demanding physical workouts, which are crucial for athletic training, high-performance sports analytics, and very intense biological health studies.
5. **Dynamic Activities:** The highest-level category includes datasets reflecting complex dynamics of movements that are typically observed in competitive sports or advanced exercise programs. These are essential for Field Comprehensive Performance Analytics and Specialized Athletics training programs.

In this work, we organize the frequently used datasets **as shown in Table (1)** based on their year of evolution, mode of representation, classes, and data-source type, to help researchers have a broad overview of the existing datasets.

Table (1): State the frequently used dataset.

Dataset	Reference	Activity level	classes	Data type
KTH	[8]	light	6	video
Weizmann	[9]	light	10	video
CMU-MMAC	[10]	Vigorous	55	multimodal
WISDM ACTi tracker	[11]	Sedentary	6	sensor
HMDB51	[12]	Moderate	51	video
CAD-60 Dataset	[13]	Moderate	12	video
PAMAP2	[14]	Sedentary	18	sensor
USC-HAD	[15]	Sedentary	12	sensor
MSR DailyActivity3D	[16]	light	16	sensor
USC-HAD	[15]	light	12	sensor
UCF101	[17]	Moderate	101	video
Opportunity Activity Recognition	[18]	dynamic	58	sensor

UCI HAR Dataset	**[19]**	**Sedentary**	**6**	**sensor**
Berkeley MHAD	**[20]**	**Light**	**11**	**multimodal**
MHEALTH Dataset	**[21]**	**Sedentary**	**12**	**sensor**
Human3.6M	**[22]**	**Light**	**17**	**multimodal**
Sports-1M	**[23]**	**dynamic**	**487**	**video**
Charades Dataset	**[24]**	**Vigorous**	**157**	**video**
SALSA Dataset	**[25]**	**Vigorous**	**18**	**multimodal**
Kinetics	**[7]**	**Moderate**	**600**	**video**
Unimib SHAR	**[26]**	**dynamic**	**17**	**sensor**
NTU RGB+D	**[27]**	**dynamic**	**120**	**multimodal**
w-HAR dataset	**[28]**	**Light**	**7**	**sensor**
KU-HAR	**[29]**	**Moderate**	**18**	**sensor**

We believe that this categorization will help researchers and practitioners choose the suitable dataset and develop more accurate, versatile, and context-sensitive activity recognition systems.

5. Deep Learning Algorithms

Deep learning has evolved into a core technology for HAR, thus providing complex and precise systems that are capable of recognizing, classifying, and predicting human activities from different data types. Various deep learning structures were effective in HAR, using temporal and spatial features of human activities; in this section, we demonstrate how the researchers use each architecture.

5.1 Deep Belief Networks (DBNs)

A class of deep neural networks constituted by several layers that learn sophisticated representations of data. DBNs have thus become popular in the deep learning community because they are also suitable for unsupervised learning making it possible to pre-train the entire network, then applying supervised learning for classification, we can therefore say that these are very powerful tools for exploring interesting relationships in Big Data. The authors applied DBN in a study to track and interpret human activity on the basis of the sensor data picked from the smartphone in the HAR (Human Activity Recognition) process [30]. Their methodology involves two primary phases: meta-learning and fine-tuning. Initially, the network is unsupervisedly conditioned using the Persistent Contrastive Divergence (PCD) for adjusting weights accustomed toward RBM level. Secondly, the network is tuned via supervised learning to achieve optimal performance in activity recognition. By combining pre-training within the DBN framework, this study has successfully demonstrated the appropriate modeling of activities using sensor data with a better performance in recognizing multiple types of human activities as the model has been tested on the UCI dataset, reaching 95.85%. Accuracy.

5.2 Convolutional Neural Network (CNN)

CNNs are a specific type of neural network that is designed to operate with data that has a grid-like organization. The success of CNNs in resolving various tasks of computer vision, such as image

and video recognition, image categorization, medical image analysis, and so on, in the context of HAR systems CNNs, used not only for vision-based data but also for sensor-based data like data from a wearable sensor, and even from non-visual sensor arrays that capture environmental or physiological signals. Repeated convolution procedures are used to investigate data for various applications. Convolutional layers have undergone several stages of transformation in order to extract more important information efficiently; additional supporting layers that reduce noise and increase computational efficiency and feature quality are the activation, pooling, batch normalization, and dropout layers.

In 2015, Dobhal et al. used CNNs to detect human activities through binary motion images. These binary motion images allow us to capture the entire human movements in a video sequence by distinguishing between the background and the foreground, where the foreground indicates the activity [31].

In 2019, Alemayoh et al. used a new structuring way as the multi-channel motion time series data of an accelerometer and gyroscope motion sensors (collected from a smartphone) are structured into a 14x60 virtual image. Similarly, their respective amplitudes of 1 dimensional DFT (Discrete Fourier Transformation) are organized into 14x60 image format, and then they feed these images to a CNN model [32].

In 2019, Bianchi et al. suggested the implementation of a cutting-edge Internet of Things system for long-term human activity recognition. They mentioned the use of four convolutional layers and one fully connected layer [33].

There are three types of CNN based on their dimensionality: 1D-CNN, 2D-CNN, and 3D-CNN. Thus, each type is designed to handle distinct data structures, and this attribute makes CNNs valuable tools both in research and implementation across various domains.

1. 1D-CNN: 1D CNNs are designed to process temporal information by featuring convolution with a unitary dimension range. Such sequences can take advantage of sequence structure features, making them useful for analyzing time series data, where every sequence element can be considered a step into time [34]. Demonstrates the efficacy of 1D-CNNs processing time-series data from wearable sensors, offering a robust solution for human activity recognition, as they mentioned that the accuracy reached by this model was 95.72%.

   The same author published another paper [35] showing that the accuracy raised to 97.49% on the same dataset (UCI dataset) when adding feature fusion with the 1D-CNN; they also tested the model on self-recorded data, reaching an accuracy of 96.27%.
2. 2D CNNs: the most common type of CNNs, it deals with grid-like data (such as images), where the convolution operation slides over two dimensions (x, y).In 2021, Fard Moshiri et al. used a Raspberry Pi 4 to collect data for seven different human activities, but in order to feed these data to a 2D-CNN classifier, the data was converted into virtual images. They aimed to leverage the CSI data's unique patterns caused by human activities to classify these activities accurately. By using This approach and the high accuracy levels achieved, they demonstrated the potential to outperform traditional HAR methods, showcasing the effectiveness of using CSI data combined with deep learning techniques [36].

3. 3D CNNs: 3D CNNs are architected for spatiotemporal data, where convolutions go through three dimensions. Most suitable for Video Analysis and Event Detection, Processing sequences of frames for activities recognition data, which include spatiotemporal dimensions. In 2019, Phyo et al. introduced the three key steps involved in this method include data acquisition from depth sensors capturing skeletal joint movements, creation of Color Skeleton Motion History Image (Color Skl-MHI) and Relative Joint Image (RJI) for feature extraction, and a 3D Deep Convolutional Neural Network (3D-DCNN)used for action recognition. Using this process effectively captures and analyzes motion, translating skeletal data into a format suitable for deep learning models to classify actions like standing, sitting, or bending, showcasing a novel approach to human activity recognition using skeletal information [37].

As the CNN model gets more depth in terms of either increasing the number of hidden layers or the number of nodes in the hidden layer, Deep Convolutional Neural Networks (DCNN) are introduced, which is nothing but a deeper CNN used to extract more features and increase the accuracy of the prediction. In 2015, Jiang and Yin employed Deep Convolutional Neural Networks (DCNN) to process data from accelerometers and gyroscopes, transforming this data into images and feeding them to the DCNN model; they aim to reduce computational costs while improving recognition accuracy[38]. They demonstrate promising results on public datasets. The proposed DCNN was tested on three public datasets. The accuracy result came as 95.18% on the UCI dataset [19], 97.01% on the USC dataset[15], and the best accuracy was on the SHO dataset 99.93% [39].

In 2019, Khelalef et al. proposed a novel deep learning-based human activity recognition system works as follows: Initially, each frame of a video is tracked, and the human body is extracted. Subsequently, they employ human "silhouettes" to generate binary space-time mappings (BSTMs) that encapsulate human actions inside a designated time frame. Lastly, features from BSTMs are extracted, and the actions are categorized using convolutional neural networks (CNNs) [40].The model evaluated on keck gesture ,Weizmann and KTH datasets the accuracy came as 100%,100%, and 92.5% respectively [8] [9] [41].

ResNet (Residual Network) is another type of CNN; more specifically, it aims at developing more robust neural networks by tackling the issue of vanishing gradients via residual blocks; these blocks enable gradients to pass through neural networks in a better manner, thereby allowing for the training of neural networks with many more layers.

DeepResNeXt is built upon the fundamental theories of ResNet by introducing an additional dimension to the architecture referred to as cardinality, in combination with depth (the number of layers) and width (the number of units in a layer) [42].

Mekruksavanich and Jitpattanakul published two papers using ResNeXt for human activity recognition. Their first paper in 2023 proposed an enhanced version of ResNeXt, adding an attention mechanism to the model. The proposed model was tested on the w-HAR dataset with accuracy of 97.68% [43].

In Their second paper published in 2024 presents “DeepUserIden” a deep learning method used to identify smartphone users based on their daily

activity employing the same model on their first publication (DeepResNeXt). This model was designed to serve the purpose of user identification by altering its architecture for processing time series sensor data that are collected from smartphones by using one dimensional convolutional blocks(1D-CB) and causal convolutions to extract temporal patterns from the sensor data [44].

5.3 Long Short-Term Memory (LSTM)

A recurrent neural network (RNN) designed to address the vanishing gradient issue that plagues conventional RNNs is the long short-term memory (LSTM) network. Its benefit over other sequence learning techniques, hidden Markov models, and other RNNs is its relative insensitivity to gap length. Its objective is to give RNN a "long short-term memory" that can endure thousands of time-steps. It can be used for time series data classification, processing, and prediction.

An input gate, an output gate, a forget gate, and a cell make up a typical LSTM unit, the three gates control the information flow into and out of the cell, and the cell retains values for arbitrarily long periods of time. Forget gates use a value between 0 and 1 to indicate which information from a prior state should be discarded in relation to the present input. To retain the information, a (rounded) value of 1 is indicated, and to discard it, a value of 0. Using the same mechanism as forget gates, input gates determine which new pieces of information to store in the existing state. By giving each piece of information, a value between 0 and 1, output gates are able to regulate which bits of information in the current state are output while taking into account the previous and current states. To retain valuable long-term dependencies for prediction-making in both present and future time-steps, the LSTM network selectively outputs pertinent information from the current state [45].

The ability of LSTM to process time series data made it offer a big advantage In the context of recognizing human activities, the author Hayat et al., 2022 created an LSTM model by fully connecting and integrating several layers, doing this, the network is able to utilize the cell correlations completely, leading to the achievement of more complex features prediction . the model consists of three LSTM layers, each layer has 64 memory cells [46].

This model achieved an overall accuracy of 95.05% using 10-fold cross-validation on UCI dataset.

In 2024, Cob-Parro et al. aimed to achieve a real-time HAR system on edge devices while minimizing computing costs and maintaining high accuracy [47].

They proposed a HAR deep learning framework tailored for edge computing, outlining an approach consisting of three primary steps: feature extraction, human activity recognition, and people detection and tracking. A MobileNetV2-SSD model is used in conjunction with Kalman filters to track individuals. The process of feature extraction entails creating a lightweight feature vector using bounding box data. These feature vectors are processed by an LSTM-based neural network for action recognition. They tested the proposed model on several datasets, but the accuracy level wasn't mentioned.

5.4 Autoencoders

Autoencoders are an artificial neural network that is used for learning efficient coding in an unsupervised manner. The goal of an autoencoder is to find a representation (encoding) for a given set of data, usually for the reasons of

dimensionality reduction or learning features. Autoencoders are artificial neural networks that seek to reproduce their input at their output layer as closely as possible; indeed, passing the data through a bottleneck at the hidden layers forces the network to encode its input as compactly as possible.

In 2014, Hasan and Roy-Chowdhury proposed continuous learning framework for human activity models employing a simple, sparse autoencoder (one layer) to automatically learn features from the unsupervised data derived from activity segments in streaming videos. They test the model on four datasets, namely KTH Human Action Dataset, UCF11 Human Action Dataset, VIRAT Dataset, and TRECVID Dataset. The best performance of the model was 96.6% accuracy on the KTH dataset [48].

In 2014, Li et al. proposed an unsupervised feature learning framework that improves the feature representation of accelerometer and gyroscope sensor data for HAR. It compares three unsupervised learning techniques: Sparse Auto-Encoder (SAE), Denoising Auto-Encoder (DAE), and Principal Component Analysis (PCA), proving the superiority of their framework over typical manual features and unsupervised learning method. By the example of channel-wise feature extraction method such as channel-wise convolution layer, mainly using Spatial Auto Encoders (SEA), the study reveals that this method outperforms monolithic approaches and conventional ones (95.24% accuracy on UCI dataset), paving toward a promising direction in HAR design [49].

5.5 Hybrid Model

In deep learning. A hybrid model refers to the combination of different types of neural networks to leverage the strengths of each approach to improve performance, enhance learning capabilities, or tackle specific challenges that might be difficult to address with a single model type. Multiple types of hybrid models can be created based on the task and the combined approaches.

By using both MLP (for classification) and CNN (for feature extraction) Mo et al., in 2016 demonstrated that this combination successfully extracts spatial characteristics of human activity skeleton data from the Microsoft Kinect sensor, categorizing the activity into different types. where CNN provide a robust feature extraction and MLP for subsequent classification demonstrates a hybrid deep learning approach that seeks to enhance the accuracy of human activity recognition [50].

In 2017, Tomas and Biswas integrated a hybrid model consist of Convolutional Neural Networks and Stacked Auto-Encoders (SAE). Where the CNNs used to process Motion History Images (MHIs) that are derived from RGB frames, to obtain motion representations, and the SAEs used to analyze human skeletal joint motions for discriminative feature learning. This two-sided approach uses spatial and temporal data, demonstrating better recognition accuracy on the benchmark dataset (91.3% accuracy on MSR Daily Activity 3D) by efficiently combining the advantages of both deep learning architectures [51].A combination of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) lightweight model design which is suitable for deployment on mobile devices is Proposed by Ankita et al., 2021, to captures both spatial features from sensor data and temporal patterns in activity sequence, tested on the UCI dataset , the model succeeded in enhancing both

accuracy and efficiency in recognizing human activities with an accuracy of 97.89% [52].

In 2021, Kang et al. presented combination of CNN and CBAM for accelerometer data processing and a bidirectional LSTM network coupled with ResNet for image data processing. The combination of these methods allows the model to take advantage of spatial features from images and temporal features from accelerometer with a high accuracy. With little input signal preprocessing, the model exhibits resilience to noise and a propensity to maintain performance. while using a skeleton image and accelerometer data, the overall accuracy was reported as 94.8%; while using a combination of skeleton image coordinates and accelerometer data, it came out as 93.1% [53].

Convolutional neural networks (CNN) and bi-directional long short-term memory (BiLSTM) networks were combined in an effective way by Nafea et al. in 2021, CNN was used to extract spatial features, while BiLSTM networks were utilized to record temporal dynamics for sensor data. By incorporating the spatial and temporal elements in sensor data, this unique strategy enhances the model's potential to differentiate human activities better, displaying a significant boost in performance in activity detection tasks. On the WISDM dataset and the UCI-HAR dataset, the hybrid model's accuracy results were 98.53% and 97.05%, respectively [54].

In 2022, Basly et al. proposed a hybrid model combining deep residual neural networks (ResNet) for spatial feature extraction and Long Short-Term Memory (LSTM) networks for temporal dynamics, this model effectively leveraged the strengths of both architectures to improve the recognition of human activities by processing raw color (RGB) data Achieving 91.65% accuracy on the MSRDailyActivity3D dataset and 91.18% on the CAD-60 dataset [55].

In 2022, Khan et al. developed a hybrid model combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, the model reached an accuracy of 90.89% on dataset that comprised 12 different physical activities performed by 20 participants created by the author [56].

Convolutional Block Attention Modules (CBAM) for refined feature emphasis, Convolutional Block Attention Modules (CBAM) are used in the hybrid model that Akter et al. in 2023, presents to improve feature extraction by utilizing the attention mechanisms. CBAM integrates features from multiple convolutional layers and significantly increases model accuracy. The AM-DLFC model was evaluated using the KU-HAR dataset, and it demonstrated an impressive classification accuracy of 96.86% with a peak accuracy of 97%. The UCI-HAR and WISDM datasets, which yielded classification accuracy results of 93.48% and 93.89% overall and 94.25% and 94.26% at their peaks, respectively, were also used to assess the durability of the AM-DLFC model [57].

6. Performance Comparison

In this thorough review, we have systematically compared and discussed 24 research articles that have used deep learning algorithms to solve the problem of human activity recognition within the period from 2014 to 2024. The comparative analysis is carefully tabulated in Table 2, which forms a core point of reference for our study. This collection of papers illustrates the progression and utilization of deep learning methods within the domain, reflecting a decade of developments and

knowledge. The comparison is structured around several critical dimensions: the datasets used in each study, whether the source data was sensor or vision-derived, the specific deep learning algorithms applied, and the highest accuracies obtained on the relevant datasets. The multifaceted nature of this approach not only illuminates the current state of the art but also reveals the path of technological advancement and methodological preferences in human activity recognition research. This enables us to summarize the way in which deep learning has been utilized to enhance accuracy and efficiency in the detection and categorization of human activities, providing helpful information on the strengths and weaknesses of different approaches within this rapidly developing and still vital field of research

Table (2): State the Comparison of selected papers.

Reference	Dataset	Data Type	Algorithms	Accuracy
[48]	KTH	Vison-based	sparse autoencoder	96.6%
[49]	UCI HAR Dataset	Sensor-based	SAE	95.24%
[31]	MSR Daily Activity 3D	Vison-based	CNN	98.5%
[38]	SHO	Sensor-based	DCNN	99.93%
[58]	CAD-60	Vison-based	CNN-MLP	81.8%
[51]	MSR Daily Activity 3D	Vison-based	CNN-SAE	91.3%
[30]	UCI HAR Dataset	Sensor-based	DBN	95.85%.
[40]	Weizmann	Vison-based	DCNN	100%
[32]	Noval data	Sensor-based	CNN	99.5%
[37]	UTKinect Action-3D	Vison-based	3D-DCNN	97%,
[33]	Noval data	Sensor-based	CNN	97%
[34]	UCI HAR Dataset	Sensor-based	1D-CNN	95.72%
[36]	CSI dataset	Sensor-based	2D-CNN	95%
[52]	UCI HAR Dataset	Sensor-based	CNN-LSTM	97.89%.
[35]	UCI HAR Dataset,	Sensor-based	1D-CNN	97.49%,
[53]	Berkeley MHAD	Multimodal	CNN-CBAM, Bi-LSTM-ResNet	94.8%
[54]	WISDM	Sensor-based	CNN-BiLSTM	98.53%
[55]	MSRDailyActivity3D	Vison-based	CNN- LSTM	91.65%
[46]	UCI HAR Dataset	Sensor-based	LSTM	95.04%
[56]	Noval data	Sensor-based	CNN-LSTM	90.89%
[43]	w-HAR dataset.	Sensor-based	DeepResNeXt	97.68%.
[57]	KU-HAR dataset	Sensor-based	CNNs-attention mechanism	96.86%

[47]	**KTH**	**Vison-based**	**LSTM**	**-**
[44]	**pervasive dataset**	**Sensor-based**	**DeepResNeXt**	**-**

Referring to Table 2, several key statistics can be extracted:

1. Datasets: the most used data set is the UCI dataset, as used by six researchers; MSR Daily Activity 3D Came second as three studies used it, and three studies mentioned collecting their data from different sensor devices (novel data). Notice that we only mentioned the datasets that the model reaches the best accuracy on; many researchers used more than one dataset as shown in Figure (2).

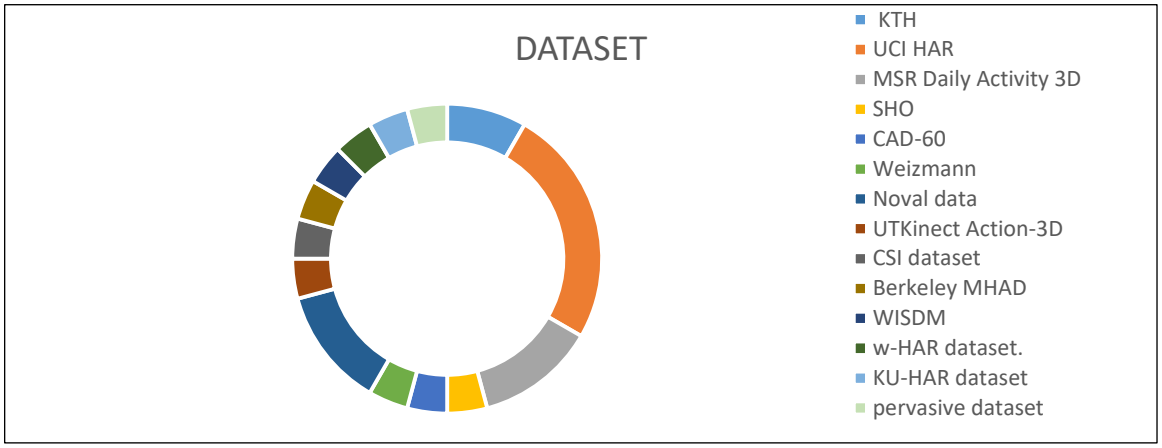

Figure (2): State the dataset distribution for the selected article

2. Data source type: based on the data type, the HAR system can be classified into vision-based, sensor-based, and multimodal. As most datasets are sensor-based, most papers used a sensor-based model; even with convolutional neural network architecture, the sensor data is converted to virtual images and classified using CNN. A virtual representation of the model type as shown in Figure (3).

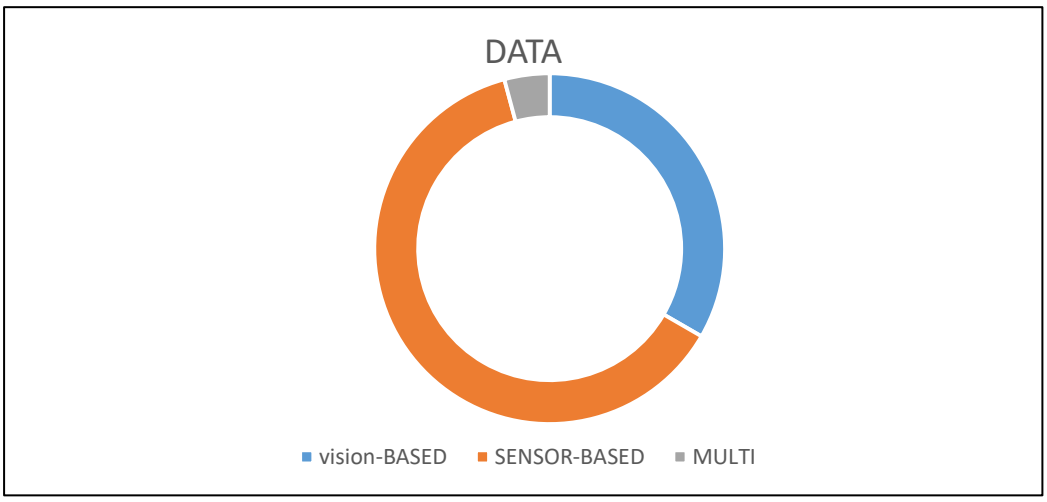

Figure (3): State the source data type models

3. Deep learning algorithm: from the 1D-CNN to DCNN, CNNs are the most used architecture as used by 11 Studies leveraging the strength of this architecture handling human activity recognition problems as it is used for both sensor and vision-based model, the second used architecture is the hybrid model combining more than one on algorithm in single model also some researchers used auto-encoders and DBN, the most use of LSTM are in hybrid model while only to studies used it alone as shown in Figure (4).

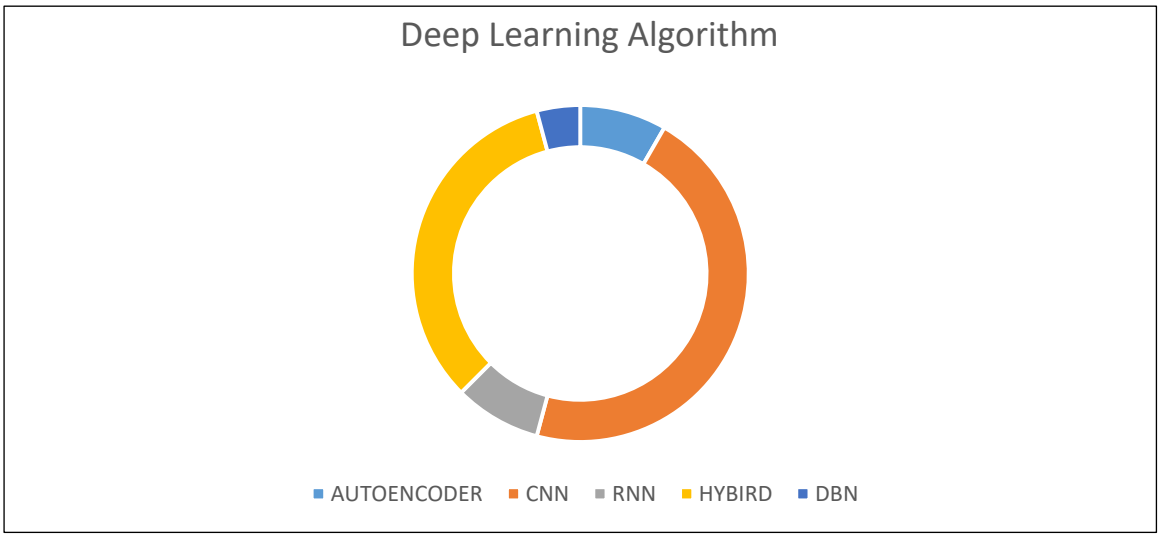

Figure (4): State the deep learning architecture

4. Accuracy level: In 2016, Besides Mo et al. who reached 81.8% on the CAD-60 dataset, all the accuracy levels ranging from 90% to 100% the most are between 95% and 100%, which leads us to the fact that deep learning algorithms are a very powerful tool to handle human activity recognition problems [50]. The best accuracy level was mentioned by Khelalef et al., 2019, as they reported that the DCCN architecture model tested on the Weizmann dataset shows 100% accuracy, while only two papers didn't mention the accuracy level [40]. Accuracy level distribution can be visualized as shown in Figure (5).

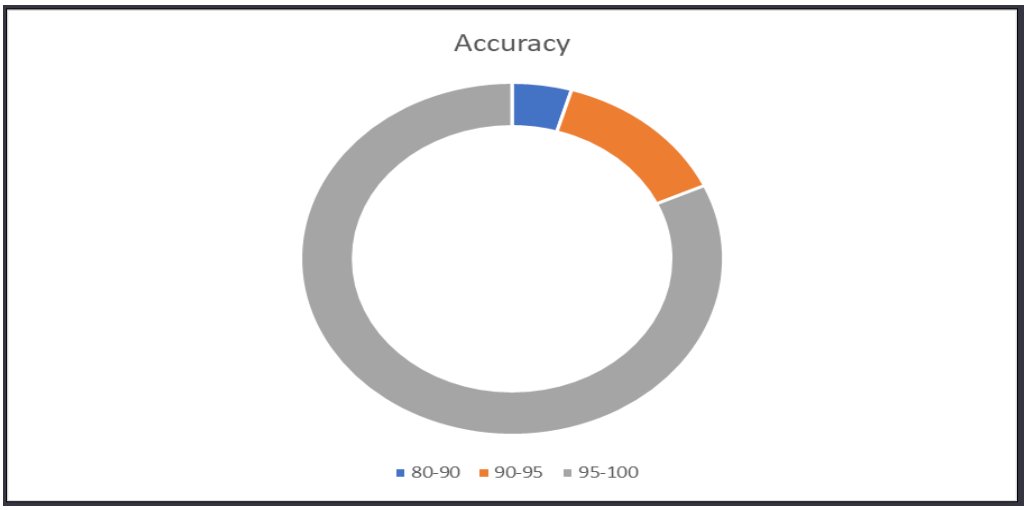

Figure (5): State the Accuracy

7. Challenges and Recommendation

The Challenges associated with the use of deep learning in HAR systems span a wide range of areas, such as data collection, label acquisition, modeling, and model deployment.

7.1 Data Collection

The requirement for large, high-quality, and diverse data sets to build models that generalize well across different scenarios – a task that is usually compromised due to the labor-intensive and costly nature of data collection. The application of data augmentation methods such as synthetic noise injection and Generative Adversarial Networks (GANs) aims to increase the size and variety of the dataset but also indicates the difficulty of addressing data quality and missing information due to real-world collection conditions. The privacy issues arising from the collection of data, however, also bring the need to implement methods that protect user data but allow efficient HAR, demonstrating the numerous challenges of optimizing deep learning for HAR within wearable technologies.

7.2 Label Acquisition

Obtaining wide-scale, heterogeneous datasets for training strong models, and the issues related to properly labeling complex activity data, which often requires external sources for ground truth. The laborious nature of label acquisition, coupled with difficulties in maintaining synchronization across different devices, makes these challenges even more complicated. To overcome these problems, methods including data augmentation, semi-supervised, and active learning techniques are used, as well as using pre-existing labeled datasets to improve model performance. Furthermore, the deployment of these computationally expensive models on low-powered wearable devices, while providing user privacy and real-time processing, provides further challenges that require continued research and development to fine-tune HAR systems for everyday practical use.

7.3 Modeling

Challenges of modeling a variety of activities from different individuals accurately, along with the temporal dependencies due to the sequential nature of movements. The issue of the critical balance is also that the problem of over-fitting and better generalizable models to new and unseen data and activities is an extra challenge. The nature of these challenges suggests that for HAR systems to be more accurate and generalizable, advanced modeling techniques and strategies that can successfully cope with the complexities of human behavior, temporal sequences, and activity variability are required.

7.4 Model Deployment

Improving inference time and minimizing power consumption while deploying deep learning models on mobile systems.

Reducing neural network complexity for deployment on resource-limited platforms and exploring emerging trends like application-specific integrated circuits (ASICs) for efficient HAR processing.

There is a need for research on intelligent computation partitioning across cloud, mobile platforms, and edge devices to improve the practical use of HAR systems.

8. Contribution

1. **Comprehensive Review:** We provide a detailed overview of the current research landscape in HAR using deep learning, focusing on both vision and sensor-based methodologies.
2. **Dataset Categorization and Analysis:** Our review organizes the frequently used datasets in HAR research, offering insights into their evolution, representation, and application in developing deep learning models.
3. **Performance Comparison:** We compare the performance of different deep learning architectures across various studies, highlighting advancements in accuracy and efficiency.
4. **Future Research Directions:** By identifying research gaps, such as challenges in data collection, privacy concerns, and the deployment of models, we propose potential areas for future exploration to advance the field of HAR further.

9. Conclusion

Deep learning has played an essential role in the development of HAR, as it has facilitated superior accuracy and performance, especially in different applications. The review points to a rising tendency to use sensor-based HAR, which is non-invasive and flexible, with deep learning models, especially CNNs and hybrid models, being at the forefront of the developments in the field. Because of the reported high levels of accuracy, despite this, problems persist in data gathering, privacy, model complexity, and the use of resource-limited devices. Future research should concentrate on these areas where more advanced deep learning techniques can be implemented, model generalizability can be improved, and the issues of deployment can be addressed. HAR systems driven by deep learning have high prospects for personalized, real-time monitoring and interaction in healthcare, sports, and daily living settings. However, the review may not cover all relevant literature especially new studies or gray literature that could provide more information. Furthermore, the fast development of deep learning technologies may result in new methodologies and findings emerging shortly after this review's publication that could not be taken into consideration during our analysis.

References

[1] Y. Wang, S. Cang, and H. Yu, "A survey on wearable sensor modality centred human activity recognition in health care," Expert Syst. Appl., vol. 137, pp. 167–190, 2019.

[2] A. Taha, H. H. Zayed, M. E. Khalifa, and E.-S. M. El-Horbaty, "Human activity recognition for surveillance applications," in Proceedings of the 7th International Conference on Information Technology, 2015, pp. 577–586.

[3] S. Mekruksavanich and A. Jitpattanakul, "Lstm networks using smartphone data for sensor-based human activity recognition in smart homes," Sensors, vol. 21, no. 5, p. 1636, 2021.

[4] J. D. Kelleher, Deep learning. MIT press, 2019.

[5] S. Zhang et al., "Deep learning in human activity recognition with wearable sensors: A review on advances," Sensors, vol. 22, no. 4, p. 1476, 2022.

[6] A. C. Tricco et al., "PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation," Ann. Intern. Med., vol. 169, no. 7, pp. 467–473, 2018.

[7] W. Kay et al., "The kinetics human action video dataset," arXiv Prepr. arXiv1705.06950, 2017.

[8] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: a local SVM approach," in Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., 2004, vol. 3, pp. 32–36.

[9] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 12, pp. 2247–2253, 2007.

[10] F. De la Torre et al., "Guide to the carnegie mellon university multimodal activity (cmu-mmac) database," 2009.

[11] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," ACM SigKDD Explor. Newsl., vol. 12, no. 2, pp. 74–82, 2011.

[12] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "HMDB: a large video database for human motion recognition," in 2011 International conference on computer vision, 2011, pp. 2556–2563.

[13] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Human activity detection from RGBD images," in Workshops at the twenty-fifth AAAI conference on artificial intelligence, 2011.

[14] A. Reiss and D. Stricker, "PAMAP2 physical activity monitoring monitoring data set," Dataset from Dep. Augment. Vision, DFKI, Saarbrücken, Ger., 2012.

[15] M. Zhang and A. A. Sawchuk, "USC-HAD: A daily activity dataset for ubiquitous activity recognition using wearable sensors," in Proceedings of the 2012 ACM conference on ubiquitous computing, 2012, pp. 1036–1043.

[16] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in 2012 IEEE conference on computer vision and pattern recognition, 2012, pp. 1290–1297.

[17] K. Soomro, A. R. Zamir, and M. Shah, "UCF101: A dataset of 101 human actions classes from videos in the wild," arXiv Prepr. arXiv1212.0402, 2012.

[18] M. Hardegger, D. Roggen, A. Calatroni, and G. Tröster, "S-SMART: A unified bayesian framework for simultaneous semantic mapping, activity recognition, and tracking," ACM Trans. Intell. Syst. Technol., vol. 7, no. 3, pp. 1–28, 2016.

[19] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones.," in Esann, 2013, vol. 3, p. 3.

[20] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Berkeley mhad: A comprehensive multimodal human action database," in 2013 IEEE workshop on applications of computer vision (WACV), 2013, pp. 53–60.

[21] O. Banos, R. Garcia, and A. Saez, "MHEALTH Dataset. UCI Machine Learning Repository." 2014.

[22] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments," IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 7, pp. 1325–1339, 2013.

[23] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in Proceedings of the IEEE

conference on Computer Vision and Pattern Recognition, 2014, pp. 1725–1732.

[24] G. A. Sigurdsson, G. Varol, X. Wang, A. Farhadi, I. Laptev, and A. Gupta, "Hollywood in homes: Crowdsourcing data collection for activity understanding," in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, 2016, pp. 510–526.

[25] X. Alameda-Pineda et al., "Salsa: A novel dataset for multimodal group behavior analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 8, pp. 1707–1720, 2015.

[26] D. Micucci, M. Mobilio, and P. Napoletano, "Unimib shar: A dataset for human activity recognition using acceleration data from smartphones," Appl. Sci., vol. 7, no. 10, p. 1101, 2017.

[27] J. Liu, A. Shahroudy, M. Perez, G. Wang, L.-Y. Duan, and A. C. Kot, "Ntu rgb+ d 120: A large-scale benchmark for 3d human activity understanding," IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 10, pp. 2684–2701, 2019.

[28] G. Bhat, N. Tran, H. Shill, and U. Y. Ogras, "w-HAR: An activity recognition dataset and framework using low-power wearable devices," Sensors, vol. 20, no. 18, p. 5356, 2020.

[29] N. Sikder and A.-A. Nahid, "KU-HAR: An open dataset for heterogeneous human activity recognition," Pattern Recognit. Lett., vol. 146, pp. 46–54, 2021.

[30] M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," Futur. Gener. Comput. Syst., vol. 81, pp. 307–313, 2018.

[31] T. Dobhal, V. Shitole, G. Thomas, and G. Navada, "Human activity recognition using binary motion image and deep learning," Procedia Comput. Sci., vol. 58, pp. 178–185, 2015.

[32] T. T. Alemayoh, J. H. Lee, and S. Okamoto, "Deep learning based real-time daily human activity recognition and its implementation in a smartphone," in 2019 16th international conference on ubiquitous robots (UR), 2019, pp. 179–182.

[33] V. Bianchi, M. Bassoli, G. Lombardo, P. Fornacciari, M. Mordonini, and I. De Munari, "IoT wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment," IEEE Internet Things J., vol. 6, no. 5, pp. 8553–8562, 2019.

[34] C.-T. Yen, J.-X. Liao, and Y.-K. Huang, "Human daily activity recognition performed using wearable inertial sensors combined with deep learning algorithms," Ieee Access, vol. 8, pp. 174105–174114, 2020.

[35] C.-T. Yen, J.-X. Liao, and Y.-K. Huang, "Feature fusion of a deep-learning algorithm into wearable sensor devices for human activity recognition," Sensors, vol. 21, no. 24, p. 8294, 2021.

[36] P. F. Moshiri, R. Shahbazian, M. Nabati, and S. A. Ghorashi, "A CSI-based human activity recognition using deep learning," Sensors, vol. 21, no. 21, p. 7225, 2021.

[37] C. N. Phyo, T. T. Zin, and P. Tin, "Deep learning for recognizing human activities using motions of skeletal joints," IEEE Trans. Consum. Electron., vol. 65, no. 2, pp. 243–252, 2019, doi: 10.1109/TCE.2019.2908986.

[38] W. Jiang and Z. Yin, "Human activity recognition using wearable sensors by deep

convolutional neural networks," in Proceedings of the 23rd ACM international conference on Multimedia, 2015, pp. 1307–1310.

[39] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, "Fusion of smartphone motion sensors for physical activity recognition," Sensors, vol. 14, no. 6, pp. 10146–10176, 2014.

[40] A. Khelalef, F. Ababsa, and N. Benoudjit, "An efficient human activity recognition technique based on deep learning," Pattern Recognit. Image Anal., vol. 29, pp. 702–715, 2019.

[41] Z. Jiang, Z. Lin, and L. Davis, "Recognizing human actions by learning and matching shape-motion prototype trees," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 3, pp. 533–547, 2012.

[42] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1492–1500.

[43] S. Mekruksavanich and A. Jitpattanakul, "A deep learning network with aggregation residual transformation for human activity recognition using inertial and stretch sensors," Computers, vol. 12, no. 7, p. 141, 2023.

[44] S. Mekruksavanich and A. Jitpattanakul, "Identifying Smartphone Users Based on Activities in Daily Living Using Deep Neural Networks," Information, vol. 15, no. 1, p. 47, 2024.

[45] C.-L. Hung, "Deep learning in biomedical informatics," in Intelligent Nanotechnology, Elsevier, 2023, pp. 307–329.

[46] A. Hayat, F. Morgado-Dias, B. P. Bhuyan, and R. Tomar, "Human activity recognition for elderly people using machine and deep learning approaches," Information, vol. 13, no. 6, p. 275, 2022.

[47] A. C. Cob-Parro, C. Losada-Gutiérrez, M. Marrón-Romera, A. Gardel-Vicente, and I. Bravo-Muñoz, "A new framework for deep learning video based Human Action Recognition on the edge," Expert Syst. Appl., vol. 238, p. 122220, 2024.

[48] M. Hasan and A. K. Roy-Chowdhury, "Continuous learning of human activity models using deep nets," in European conference on computer vision, 2014, pp. 705–720.

[49] Y. Li, D. Shi, B. Ding, and D. Liu, "Unsupervised feature learning for human activity recognition using smartphone sensors," in Mining Intelligence and Knowledge Exploration: Second International Conference, MIKE 2014, Cork, Ireland, December 10-12, 2014. Proceedings, 2014, pp. 99–107.

[50] L. Mo, F. Li, Y. Zhu, and A. Huang, "Human physical activity recognition based on computer vision with deep learning model," in 2016 IEEE international instrumentation and measurement technology conference proceedings, 2016, pp. 1–6.

[51] A. Tomas and K. K. Biswas, "Human activity recognition using combined deep architectures," in 2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP), 2017, pp. 41–45.

[52] Ankita, S. Rani, H. Babbar, S. Coleman, A. Singh, and H. M. Aljahdali, "An efficient and lightweight deep learning model for human activity recognition using smartphones," Sensors, vol. 21, no. 11, p. 3845, 2021.

[53] J. Kang, J. Shin, J. Shin, D. Lee, and A. Choi, “Robust human activity recognition by integrating image and accelerometer sensor data using deep fusion network,” Sensors, vol. 22, no. 1, p. 174, 2021.

[54] O. Nafea, W. Abdul, G. Muhammad, and M. Alsulaiman, “Sensor-based human activity recognition with spatio-temporal deep learning,” Sensors, vol. 21, no. 6, p. 2141, 2021.

[55] H. Basly, W. Ouarda, F. E. Sayadi, B. Ouni, and A. M. Alimi, “DTR-HAR: deep temporal residual representation for human activity recognition,” Vis. Comput., vol. 38, no. 3, pp. 993–1013, 2022, doi: 10.1007/s00371-021-02064-y.

[56] I. U. Khan, S. Afzal, and J. W. Lee, “Human activity recognition via hybrid deep learning based model,” Sensors, vol. 22, no. 1, p. 323, 2022.

[57] M. Akter, S. Ansary, M. A.-M. Khan, and D. Kim, “Human activity recognition using attention-mechanism-based deep learning feature combination,” Sensors, vol. 23, no. 12, p. 5715, 2023.

[58] L. Mo, F. Li, Y. Zhu, and A. Huang, “Human physical activity recognition based on computer vision with deep learning model,” Conf. Rec. - IEEE Instrum. Meas. Technol. Conf., vol. 2016-July, 2016, doi: 10.1109/I2MTC.2016.7520541.