Computer Modeling in Engineering & Sciences

Tech Science Press

Doi:10.32604/cmes.2025.071858

REVIEW

Deep Learning and Federated Learning in Human Activity Recognition with Sensor Data: A Comprehensive Review

Farhad Mortezapour Shiri*, Thinagaran Perumal, Norwati Mustapha and Raihani Mohamed

Faculty of Computer Science and Information Technology, University Putra Malaysia (UPM), Serdang, 43400, Malaysia

*Corresponding Author: Farhad Mortezapour Shiri. Email: gs63904@student.upm.edu.my

Received: 13 August 2025; Accepted: 06 October 2025; Published: 26 November 2025

ABSTRACT: Human Activity Recognition (HAR) represents a rapidly advancing research domain, propelled by continuous developments in sensor technologies and the Internet of Things (IoT). Deep learning has become the dominant paradigm in sensor-based HAR systems, offering significant advantages over traditional machine learning methods by eliminating manual feature extraction, enhancing recognition accuracy for complex activities, and enabling the exploitation of unlabeled data through generative models. This paper provides a comprehensive review of recent advancements and emerging trends in deep learning models developed for sensor-based human activity recognition (HAR) systems. We begin with an overview of fundamental HAR concepts in sensor-driven contexts, followed by a systematic categorization and summary of existing research. Our survey encompasses a wide range of deep learning approaches, including Multi-Layer Perceptrons (MLP), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory networks (LSTM), Gated Recurrent Units (GRU), Transformers, Deep Belief Networks (DBN), and hybrid architectures. A comparative evaluation of these models is provided, highlighting their performance, architectural complexity, and contributions to the field. Beyond Centralized deep learning models, we examine the role of Federated Learning (FL) in HAR, highlighting current applications and research directions. Finally, we discuss the growing importance of Explainable Artificial Intelligence (XAI) in sensor-based HAR, reviewing recent studies that integrate interpretability methods to enhance transparency and trustworthiness in deep learning-based HAR systems.

KEYWORDS: Human activity recognition (HAR); machine learning; deep learning; sensors; Internet of Things; federated learning (FL); explainable AI (XAI)

1 Introduction

The recent advancements on the Internet of Things (IoT) have led to the proliferation of embedded sensors across a vast array of applications, enabling the collection of massive volumes of real-world data. These developments facilitate the monitoring and control of various IoT-enabled devices, enhancing the interactivity between physical objects and digital data platforms. [1] The IoT encompasses a network of interconnected physical entities ranging from vehicles and buildings to everyday objects, equipped with sensors, electronics, and network connectivity, which collectively gather and exchange data [2]. A significant application of IoT technology is in Human Activity Recognition (HAR), which involves the identification and classification of human actions using data sourced from multiple sensors and devices such as Wi-Fi signals, cameras, radar, and wearable sensors. HAR aims to detect a wide range of human motions, including, but not limited to, running, walking, stair climbing, falling, sitting, and standing [3]. The primary goal of HAR

Copyright © 2025 The Authors. Published by Tech Science Press.
This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

is to analyze and understand human interactions with their surroundings, focusing on detailed movements of the whole body and individual limbs. By interpreting these activities, it is possible to predict outcomes, infer intentions, and assess the psychological state of individuals involved [4].

Researchers in the field of Human Activity Recognition (HAR) are developing methods to observe and analyze the actions of individuals to identify the types of activities being performed. HAR systems are broadly classified into two categories: sensor-based and video-based systems [5]. Video-based HAR systems utilize one or more cameras to record videos of human activities, capturing multiple perspectives to detect movements. However, these systems face significant challenges related to privacy concerns, as individuals may be reluctant to be continuously recorded during their daily activities. Additionally, processing video data for HAR can be computationally intensive, posing another significant barrier to the widespread adoption of video-based systems [6]. In contrast, sensor-based HAR has gained popularity among both users and researchers due to its numerous advantages over video-based systems. Sensor-based systems involve the automatic recognition of human activities using data collected from various sensing devices, including wearable and ambient sensors [7]. These sensors are advantageous as they are less susceptible to environmental disturbances, capturing continuous and precise motion signals. This robustness enhances the reliability and applicability of sensor-based HAR across diverse environments, significantly improving the efficiency and accuracy of activity recognition [3]. Overall, the choice between sensor-based and video-based HAR systems depends on the specific requirements of the application, balancing factors such as accuracy, privacy, and computational demands.

Machine learning (ML) approaches are commonly employed to address human activity recognition in smart environments, categorized into conventional machine learning (CML) models and deep learning (DL) models. While CML models have been historically utilized for HAR, they exhibit several limitations that have spurred the shift towards the adoption of deep learning methods. CML techniques require the manual extraction of features, which depend heavily on domain-specific expertise or human experience. This necessity for heuristic feature design not only restricts these models to environments where expert knowledge is available but also limits their applicability to more generalized scenarios and diverse tasks [8]. Furthermore, CML models are typically constrained to learn shallow features that align with human expertise, which predominantly facilitate the recognition of basic activities such as running or walking. They often struggle to recognize high-level or context-aware activities due to their inability to interpret complex data patterns without explicit feature guidance.

In contrast, DL models overcome these limitations by eliminating the need for manual feature extraction, thus allowing for more scalable and robust activity recognition systems. Deep learning (DL) is the process of learning hierarchical data representations by using architectures with several hidden layers that make up the depth of a neural network. In DL algorithms, data flows through these layers in a cascading fashion, with each layer gradually extracting complex features and passing crucial information to its successor. Low-level features are captured by the first layers, and these fundamental features are then combined and improved upon by later layers to provide a thorough and complex data representation [9]. In fact, deep learning architectures inherently learn to identify intricate patterns and features directly from raw data, enabling them to discern more complex, high-level activities. This capability significantly enhances the accuracy and adaptiveness of HAR systems. Moreover, while CML methods generally require extensive labeled datasets for training, deep generative networks can effectively learn from unlabeled data, offering substantial benefits for developing efficient HAR systems. DL models also demonstrate superior performance in handling variability in data due to different individuals, device models, and device poses, making them more versatile and effective in practical applications [10].

In this study, we offer concise, high-level overviews of pivotal deep learning techniques that have significantly influenced sensor-based Human Activity Recognition (HAR) systems. For detailed insights into specific methods or fundamental deep learning procedures, we encourage our readers to consult specialized research papers, comprehensive surveys, textbooks, and tutorials. Below, we outline the contributions of our study:

1. Related Review Works on Sensor-Based HAR: We begin by examining several review papers that focus on deep learning applications in sensor-based human activity recognition. This provides a foundational understanding of the current research landscape and identifies key advancements and methodologies.
2. Background of Human Activity Recognition: We present an overview of the human activity recognition field, discussing its established and emerging applications, challenges, and the sensors utilized. Additionally, we highlight popular publicly available datasets that are instrumental in the development and benchmarking of HAR systems.
3. Evaluation of Deep Learning Models in HAR: Following a brief introduction to prevalent deep learning models, we delve into recent studies that have employed these models for recognizing human activities from sensor data. We discuss various architectures, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Deep Belief Networks (DBN), and Transformer. Lastly, we categorize recent articles in our domain based on the deployment of different models and assess these models based on their accuracy on publicly available datasets, deployment characteristics, architectural details, and overall achievements in the field.
4. Federated Learning (FL) in HAR: introducing the fundamental concepts of Federated Learning (FL), highlighting its key design considerations, architectures, and optimization algorithms. Subsequently, we review the applications of FL in Human Activity Recognition (HAR) and provide a summary of recent research efforts that have employed FL in HAR systems.
5. Use of Explainable AI (XAI) in HAR: Discuss the necessity of Explainable Artificial Intelligence (XAI) in sensor-based HAR systems that rely on deep learning and give an overview of current studies that have used XAI methods in HAR.

This structured approach not only encapsulates the state of the art in sensor-based HAR using deep learning but also sets the stage for future research directions in this rapidly evolving field. The article is organized as follows: The research method including research questions and research scope is explained in Section 2. The related existing review works in the field of HAR using deep learning are discussed in Section 3. Section 4 provides an in-depth overview of HAR concepts, including applications, challenges, sensor data, and datasets, aligning these elements with the posed research questions. Section 5 introduces significant deep learning contributions to HAR, detailing specific models and their roles in advancing the field. Section 6 provides an overview of the applications of Federated Learning (FL) in HAR systems. The application of XAI models in DL-based HAR systems is discussed in Section 7. Research directions and future aspects are covered in Section 8. The paper concludes with Section 9.

2 Method

2.1 Research Questions

In this study, we address several critical research questions that explore various facets of Human Activity Recognition (HAR) systems.

- Q1: What are the real-world applications of HAR systems?

This question seeks to identify and explain the diverse practical implementations of HAR technologies across different sectors.

- Q2: What challenges are we facing in this field, and what potential solutions may exist?

Here, we explore the current obstacles impeding HAR development and effectiveness, along with innovative strategies that might overcome these challenges.

- Q3: What are the mainstream sensors and major public datasets in this field?

This inquiry focuses on detailing the sensors predominantly used in HAR systems and highlighting the key datasets that facilitate research and development in this area.

- Q4: What deep learning approaches are employed in the field of HAR, and what are the pros and cons of each?

We aim to review the various deep learning models applied to HAR, assessing their strengths and limitations in context.

- Q5: What are the applications of Federated Learning (FL) in Human Activity Recognition (HAR)?

Investigating the benefits of FL in addressing key challenges of real-world HAR systems, including privacy, communication costs, scalability, and latency.

- Q6: What is the necessity of Explainable Artificial Intelligence (XAI) for HAR systems?

Examining the Explainable Artificial Intelligence (XAI) models that can be used in sensor-based HAR systems that rely on deep learning.

This article is organized to systematically address the research questions and provide a comprehensive review of the state-of-the-art in HAR using deep learning.

2.2 Research Scope

In this paper, we conduct a comprehensive systematic review of the field of human activity recognition (HAR) utilizing deep learning methodologies. We conducted a targeted search on the Google Scholar database. Our search strategy employed a carefully selected set of keywords pertinent to HAR, including "human activity recognition", "HAR", "action detection", "multi-resident activity recognition", and "fall detection". These were combined with terms related to deep learning terms including "deep learning", "CNN", "convolutional neural network", "RNN", " recurrent neural network", "LSTM", " long short-term memory", "GRU", "gated recurrent unit", "MLP", "multi-layer perceptron", "DBN", and "deep belief network". Initially, 920 papers were identified through our keyword-based search. Subsequent screening processes involved the exclusion of duplicates, papers using visual data for activity recognition, non-English papers, and non-technical studies. The final selection focused on 125 recent high-quality papers that are most relevant to advancements in HAR. The methodological flow of article selection is illustrated in Fig. 1.

The author co-citation network and recurrent keywords were analyzed and shown using VOS Viewer, a free and open-source visualization software. Fig. 2 illustrates the distribution of scholarly articles focused on human activity recognition (HAR) utilizing deep learning models over recent years. The chart depicts a significant uptick in research within this area, particularly highlighting the robust growth throughout this decade. This trend clearly demonstrates the increasing reliance on deep learning methodologies in developing HAR systems, reflecting both a growing academic interest and advancements in technological applications.

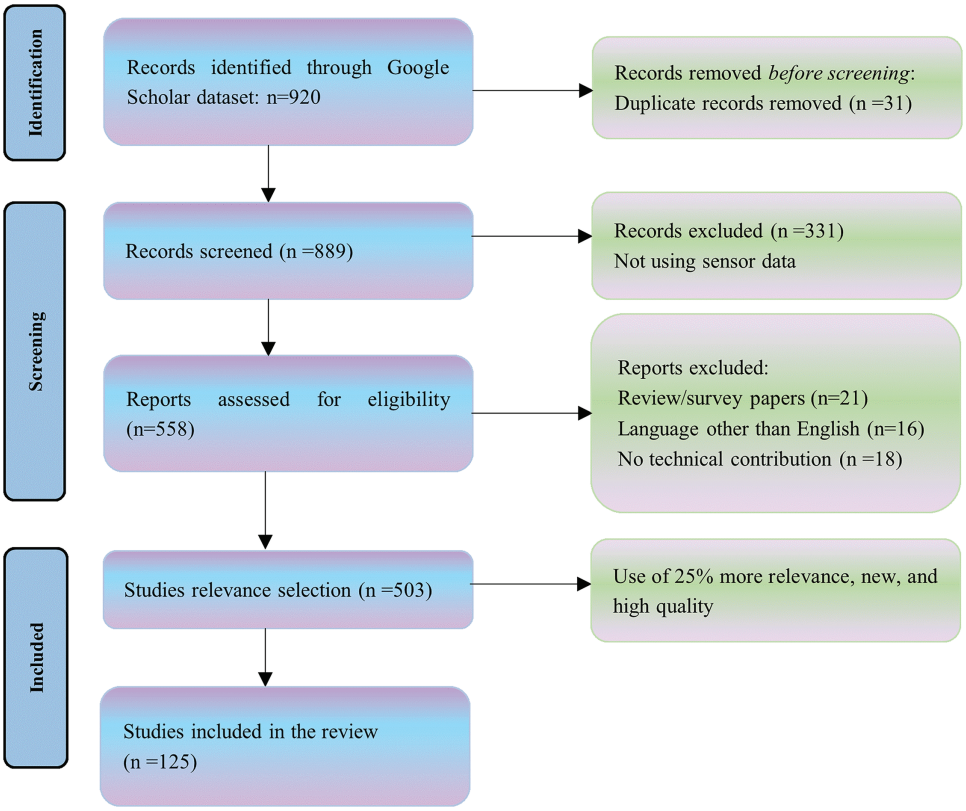

Figure 1: Steps performed for selection of articles

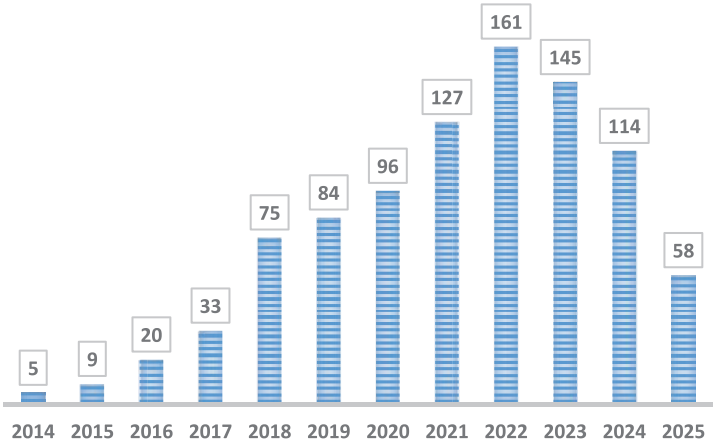

Figure 2: The number of articles in the field of HAR using deep learning during recent decades

A thorough breakdown of the publishing output distribution per nation is shown in Fig. 3. The analysis of publication output by nation revealed that the countries with the largest publication output in this discipline were China, India, South Korea, and the United States.

Finding research trends in a particular field of study is made easier with the help of keyword analysis. The density of keyword occurrences in the chosen documents was visualized in this experiment using VOS Viewer as the tool. The most often used keywords are deep learning, activity recognition, sensors,

classification, mobile, and accelerometer. To view the keyword density graphically, please refer to Fig. 4. The density of the identified keywords is shown visually in the figure.

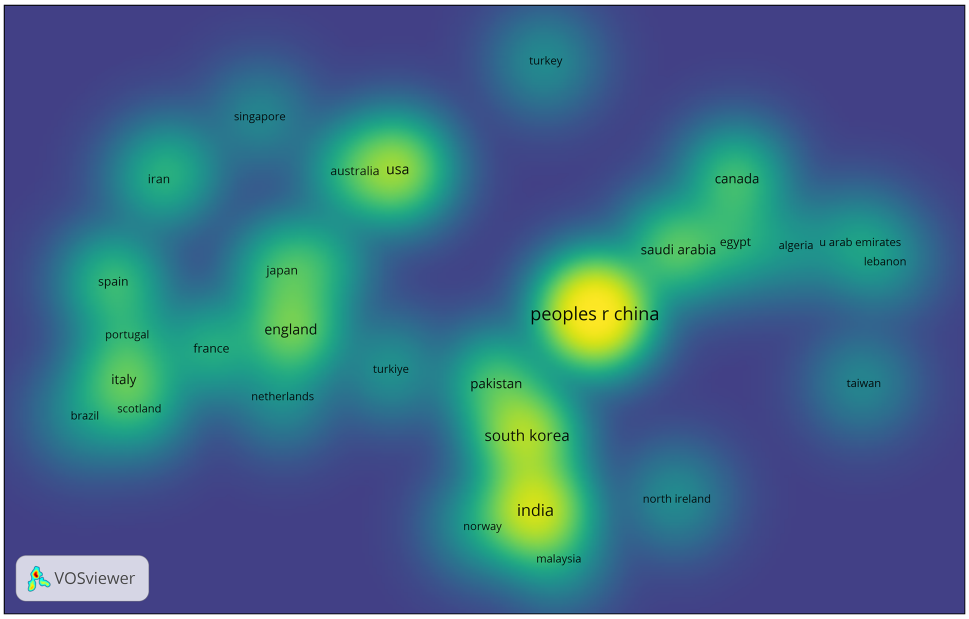

Figure 3: Distributions of publication output by countries in the field of sensor-based human activity recognition

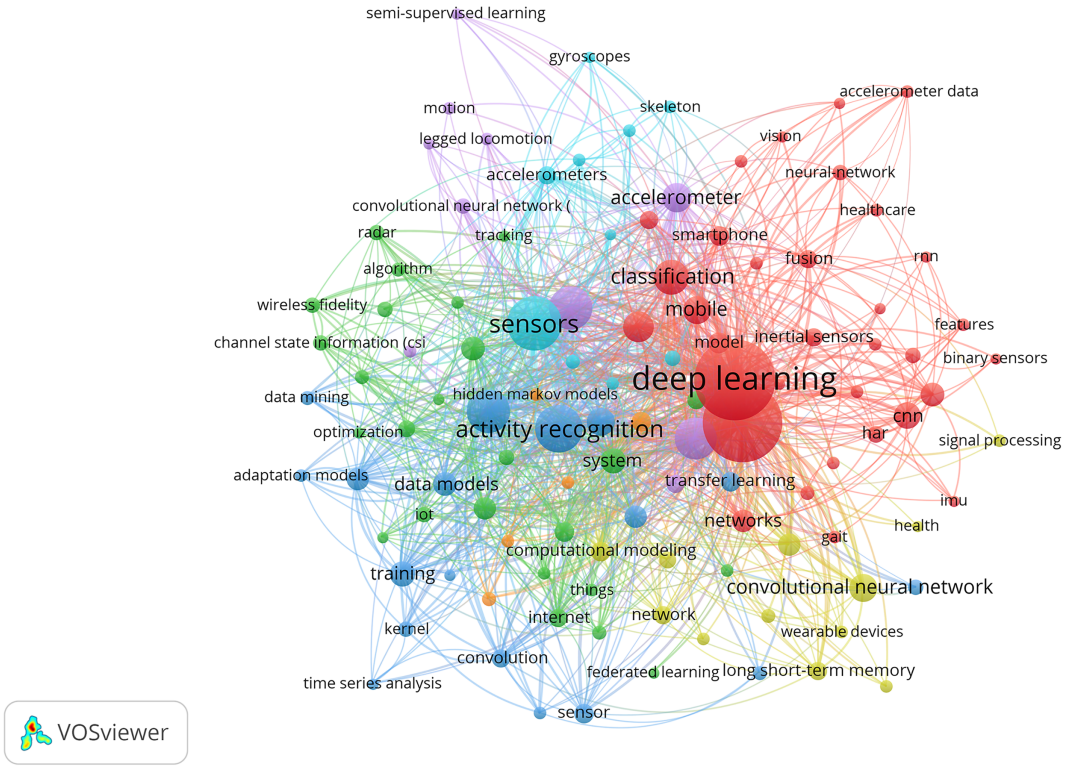

Figure 4: The density visualization of the identified keywords

3 Related Works

Numerous studies have been conducted on human activity recognition (HAR), yet the bulk of these have centered on delineating the taxonomy of HAR and evaluating the most sophisticated systems that utilize traditional machine learning techniques [11–14]. While there have been reviews on the use of deep learning models for HAR, these studies often focus on a narrow selection of deep learning architectures and their variants, providing a somewhat limited perspective on the field.

A significant advancement in this domain is provided by [15], which detailed advanced deep learning approaches for sensor-based HAR. This review illuminated the multimodal nature of sensory data and discussed publicly available datasets that facilitate the assessment of HAR systems under various challenges. The authors proposed a new taxonomy to categorize deep learning approaches based on the specific challenges they address, providing a structured overview of the state of research by summarizing and analyzing the challenges and corresponding deep learning approaches.

Ramanujam et al. [16] focused on deep learning methods used in wearable and smartphone sensor-based systems. They distinguished between traditional and hybrid deep learning models, discussing each in terms of their advantages, limitations, and unique characteristics. The review also covered benchmark datasets commonly used in the field, concluding with a list of unresolved issues and challenges that warrant further investigation.

Gu et al. [10] delivered a comprehensive review on recent advancements and challenges in deep learning for HAR. They categorized deep learning models into generative, discriminative, and hybrid types. Notably, they discussed popular generative models like Restricted Boltzmann Machines (RBMs), autoencoders [17], and Generative Adversarial Networks (GANs) [18], as well as discriminative models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and their variants. Their analysis of hybrid models highlighted how outputs from generative models could be utilized as inputs for discriminative models to enhance classification or regression tasks.

Zhang et al. [19] provided a thorough examination of the latest developments, emerging trends, and significant challenges in wearable-based HAR. Their review started with an overview of common sensors, real-world applications, and accessible HAR datasets. Following an assessment of each deep learning technique's strengths and weaknesses, they discussed the advancements in wearable HAR and provided guidance on selecting optimal deep learning strategies. The review concluded with a discussion on current challenges from data, label, and model perspectives, each offering potential research opportunities.

Despite the growing body of work on human activity recognition (HAR) using deep learning, there remains a critical need for a comprehensive and in-depth exploration of the latest deep learning techniques applied to HAR. Our study aims to fill this gap by focusing on state-of-the-art deep learning methods tailored for sensor-based HAR, which distinguishes our work from other publications currently available. We extend our analysis beyond single-sensor systems to include a variety of sensor types, such as ambient and wearable sensors, thereby broadening the scope of our review.

Furthermore, our research delves into the latest advancements aimed at overcoming existing barriers and challenges within the field. We provide insights into potential areas for future research, emphasizing the complexity of HAR systems, especially in environments with multiple residents. This aspect is particularly challenging, yet crucial for the advancement of HAR systems. We include a review of significant studies that have addressed the issue of activity recognition in multi-resident scenarios, highlighting their methodologies and findings. Our aim is to provide a comprehensive overview that not only summarizes the current state of HAR technologies but also sets the stage for future innovations and applications in this dynamic field.

4 Overview of HAR Concepts

4.1 HAR Application

Human Activity Recognition (HAR) utilizing sensor data has been extensively deployed across various research domains, significantly enhancing capabilities in ambient assisted living (AAL), healthcare systems, behavior analysis, and security frameworks. Furthermore, HAR technologies are integral in facilitating human-robot interactions and in the accurate recognition of sport-specific movements.

4.1.1 Ambient Assisted Living

Ambient Assisted Living (AAL) is an emerging communication technology designed to support the elderly in maintaining independence and activity in their daily lives [20]. HAR systems integrated into AAL environments play a crucial role in ensuring the safety and well-being of elderly or disabled individuals by monitoring daily activities and alerting caregivers to potential issues [21].

The demographic shift towards an older population has become more pronounced, with a notable increase in the average lifespan leading to a larger proportion of seniors living with various disabilities. Statistics reveal that the global population of individuals aged 65 and older has surged by over 360 million, now representing more than 8.5% of the worldwide population. This demographic change has significantly impacted community needs, escalating demand for home assistance, rehabilitation, and physical support, thereby driving up healthcare costs [22].

HAR is an integral component of smart home technologies that enable seniors to live autonomously, thereby enhancing their quality of life and standard of care. The primary function of these smart home environments, also known as AAL systems, is to remotely monitor and evaluate the health and safety of elderly individuals, those with dementia, and others with relevant disabilities [23]. Moreover, the integration of HAR in smart homes facilitates a transparent representation of the surrounding context, allowing for the implementation of various health technology applications. These applications range from monitoring disease progression and recovery to detecting anomalies, such as falls, highlighting the versatility and critical importance of HAR in modern healthcare technology [24].

4.1.2 Healthcare Systems

Human Activity Recognition (HAR) systems are increasingly utilized within healthcare settings to monitor and manage patients, particularly the elderly and disabled. These e-health systems encompass a wide range of applications, including remote patient care, respiratory biofeedback, comprehensive activity monitoring, mental stress evaluation, both physical and mental rehabilitation, weight training, as well as real-time assessment of posture, vision, and movement [13]. The capability of HAR systems to accurately recognize human activities is pivotal in identifying various health disorders such as cardiovascular issues and Alzheimer's disease. Early detection through HAR can facilitate timely medical interventions, significantly enhancing patient outcomes [25].

Furthermore, HAR plays a critical role in maintaining the overall physical and mental health of the population. For chronic diseases such as obesity, diabetes, and cardiovascular disorders, physicians can leverage HAR to continuously monitor patients' daily activities. This continuous monitoring allows for the strict management of dietary and exercise regimes essential for disease control [26]. For example, individuals affected by these conditions are often required to maintain a balanced diet and engage in regular physical activity [27]. HAR systems provide a way to record daily activities, supplying clinicians with up-to-date information and offering patients real-time feedback on their progress. Additionally, for those suffering from

mental health issues or cognitive decline, HAR systems are crucial for the continuous observation necessary to promptly identify any abnormal behavior, thereby preventing potential adverse outcomes [11].

The implementation of activity recognition systems empowers patients to take charge of their health and enables healthcare providers to monitor their patients' conditions more effectively and tailor their recommendations. By facilitating continuous monitoring, HAR systems can reduce the length of hospital stays, enhance diagnostic accuracy, and ultimately improve patients' quality of life [28].

4.1.3 Fitness and Sports

Human Activity Recognition (HAR) systems are revolutionizing the fitness industry by enabling individuals to monitor their physical activities, such as walking, running, cycling, and swimming, through advanced wearable technologies like smartwatches and fitness bands [29]. These systems provide valuable data on the duration, intensity, and caloric expenditure of physical activities. Recognized as a crucial paradigm, physical activity (PA) recognition is linked to significant benefits for both physical and mental health and is integral to various fitness and rehabilitation programs [30].

Physical activity is vital for reducing the risk of numerous chronic and non-communicable diseases including diabetes, hypertension, depression, obesity, as well as cardiovascular and cerebrovascular disorders [31]. Furthermore, regular physical activity physiologically improves mood, supports active lifestyles in older adults, enhances self-esteem, and helps in managing blood pressure, anxiety, stress levels, and weight. It also reduces the risk of cognitive disorders like Alzheimer's disease in the elderly [32].

Traditionally, dietitians and medical professionals have relied on self-completed questionnaire methods, asking participants to log their daily activities. These questionnaires are then analyzed to assess the individual's physical activity level and provide tailored feedback. However, the analysis of self-reported data is time-consuming and labor-intensive, especially with large populations. To overcome these challenges, recent advances have seen the adoption of sensor-based technologies that provide a more effective means of capturing and compiling daily contextual data and life logs [33,34].

Moreover, in the realm of elite sports, the objective evaluation of an athlete's performance is crucial. Automated sport-specific movement detection and recognition systems, facilitated by sensor-based HAR technologies, provide significant improvements over manual performance analysis methods [35]. These systems are employed during training sessions and competitions across various sports [36], including ski jumping [37,38], tennis [39–41], running [42–44], Boxing [45,46], Golf [47–49], volleyball [50–52], swimming [53–55], cricket [56–58], skateboard [59–61], and other. The adoption of HAR in these contexts eliminates the inaccuracies associated with traditional performance analysis, enabling detailed and precise assessments that are essential for optimizing athlete training and competitive performance.

4.1.4 Security and Surveillance

In the realm of security and surveillance, the urgency for immediate action is pronounced in situations involving suspect behaviors, such as extended periods of loitering, sudden running, theft of mobile devices, confrontational arguments, suspicious activities towards others, or potential threats of suicide bombing. These scenarios require a robust intelligent surveillance system capable of not only detecting but also swiftly responding to potential threats through timely alerts. In the current global context, where numerous attacks are attributed to terrorism, the ability to predict and preemptively address such threats can significantly mitigate or even prevent loss of life [62].

HAR systems have become integral to surveillance strategies that utilize both sensor and visual data. These systems are adept at continuous monitoring, detecting unauthorized entries, and identifying abnormal

activities. The effective deployment of HAR in these contexts is crucial for enhancing situational awareness and enabling rapid response to potential security incidents, thereby playing a critical role in safeguarding public safety [13].

4.2 HAR Systems Challenges

Human Activity Recognition (HAR) has emerged as a crucial field with diverse applications ranging from healthcare to security. Despite significant research advancements, HAR systems continue to face substantial challenges that hinder their full potential.

4.2.1 Security and Surveillance

A fundamental aspect of HAR research is data collection, which often faces issues such as unlabeled datasets, absence of temporal context, uncertain class labels, and restrictive data conditions. These challenges need to be addressed to improve the accuracy of activity anticipation and recognition [63,64].

4.2.2 Hardware Limitations

Hardware Limitations: Hardware forms the backbone of HAR endeavors, especially in scenarios dealing with large volumes of data. Commonly used hardware includes smartphones, smartwatches, and various sensors. However, limitations related to hardware capabilities, computational costs, and algorithmic constraints remain a concern and can hinder the scalability and effectiveness of HAR systems [12].

4.2.3 Complex Human Activities (CHA) Recognition

Recognizing complex human activities (CHAs) poses significant challenges in the field of Human Activity Recognition (HAR). CHAs often encompass multiple concurrent or overlapping actions sustained over extended periods. Examples include cooking, which may simultaneously involve multiple tasks such as chopping vegetables, monitoring cooking progress, and cleaning dishes, or writing, which can involve organizing thoughts, typing, and reviewing text concurrently. These are contrasted with simple human activities (SHAs), which are defined as singular, brief actions such as sitting or standing. The difficulty lies in accurately identifying and distinguishing CHAs from SHAs, given the complexity and duration of CHAs compared to the more transient nature of SHAs [65]. Moreover, CHAs can occur in both concurrent and interleaved modes. For instance, an individual might engage in cooking and cleaning dishes in an alternating sequence, or cook while listening to music, and also manage these tasks independently but simultaneously. The capability to model these dynamic and overlapping actions accurately remains a formidable challenge in HAR, necessitating advanced algorithms capable of capturing the nuanced intricacies of human behavior [5]. The literature on Human Activity Recognition (HAR) identifies three primary methodologies for recognizing complex human activities (CHAs), each with distinct approaches and limitations.

SHA-Based Recognition: This method involves applying simple human activity (SHA) identification techniques to CHAs, essentially treating complex activities as if they were simple ones. The primary limitation here is that the features extractable from SHAs, which are inherently less complex, do not adequately capture the nuances of CHAs. Consequently, this approach may not accurately represent the intricacies of complex activities, leading to potential inaccuracies in activity recognition [66].

Composite SHA Representation: In this approach, CHAs are modeled as combinations of multiple SHAs that have been meticulously labeled and predefined. While this method benefits from leveraging detailed SHA insights, it heavily depends on domain expertise to define and label the SHAs accurately. Additionally,

this method is constrained by the nature of SHAs themselves, which may not fully encompass the non-semantic and unlabeled components frequently present in CHAs, thereby limiting the recognition capability of truly complex activities [67,68].

Latent Semantic Analysis: The third approach utilizes topic models to identify latent semantics in sensor data, which are presumed to reflect the characteristics of CHAs. This method can potentially uncover underlying patterns and structures in activity data that are not explicitly labeled. However, its major drawback is that topic models generally focus on the distribution of data and often neglect the sequential aspects of activities. This oversight can lead to a significant gap in recognizing the temporal dynamics that are critical for accurately modeling and understanding CHAs [69–71].

Each of these methods provides a framework for addressing the challenges posed by CHA recognition but also illustrates the complexity and the need for further refinement to enhance accuracy and applicability in real-world scenarios.

4.2.4 Multi-Resident Activity Recognition

Recognizing activities within systems that house multiple residents introduces unique challenges. Unlike single-resident systems, where activities can be directly attributed to the sole inhabitant, multi-resident systems must discern which individual is responsible for which activity. This complexity is heightened by the residents' ability to engage in both solitary and collaborative activities, often influenced by the dynamics of social interactions. A particularly critical issue in multi-resident environments is data association, which is the process of accurately linking each environmental sensor-detected event, such as the opening of a refrigerator door, to the initiating individual [72]. The ambiguity of sensor readings, compounded by events triggered in close temporal and spatial proximity, significantly complicates this task. For example, if two sensors in different rooms are activated simultaneously, it becomes challenging to accurately determine which individual is in which room [73]. To address these challenges, research suggests three main approaches including wearable-based approaches, data-driven approaches, and single-model approaches [74].

Wearable-Based Approaches: These methods require residents to wear personal sensors, like smart jewelry or wristbands, which can provide precise data to associate activities with specific individuals [72,75].

Data-Driven Approaches: These techniques treat data association as a distinct learning task that precedes activity classification, allowing for more a nuanced understanding and categorization of activities [76,77].

Single-Model Approaches: Leveraging raw ambient sensor data, these models implicitly learn data associations during the training process, aiming to integrate and correlate data without explicit labeling [78,79].

4.2.5 Group Activity Recognition (GAR)

Recognizing group activities within HAR systems poses another significant challenge. GAR often begins with the recognition of individual activities, which are then synthesized to infer group behavior. However, the structural differences between individual and group actions, along with the variability in behaviors exhibited by individuals within the same group setting, complicate the recognition process. Direct modeling of group dynamics is essential to understanding the roles and relationships among participants accurately. Currently, GAR systems struggle with precision and are limited by a lack of labeled data, which is crucial for training effective recognition algorithms [80].

4.3 Sensor Data

In smart environments, the effective collection of raw data is crucial for the functionality and intelligence of Human Activity Recognition (HAR) systems. This data is gathered through a diverse array of sensors, actuators, and smart devices such as smartwatches, smart glasses, smartphones, and others. These components are essential for capturing a wide range of activities and interactions within these environments.

Fig. 5 provides an overview of the various sensors employed for data collection in Human Activity Recognition (HAR) systems. Sensors are generally classified into two categories including radio-based sensors and binary sensors. The most common radio-based sensing systems are Bluetooth, ZigBee, Z-waves, RFID, 6LoWPAN, and Wi-Fi. Binary sensors are divided into two groups: Ambient (Environmental) sensors and wearable sensors [73].

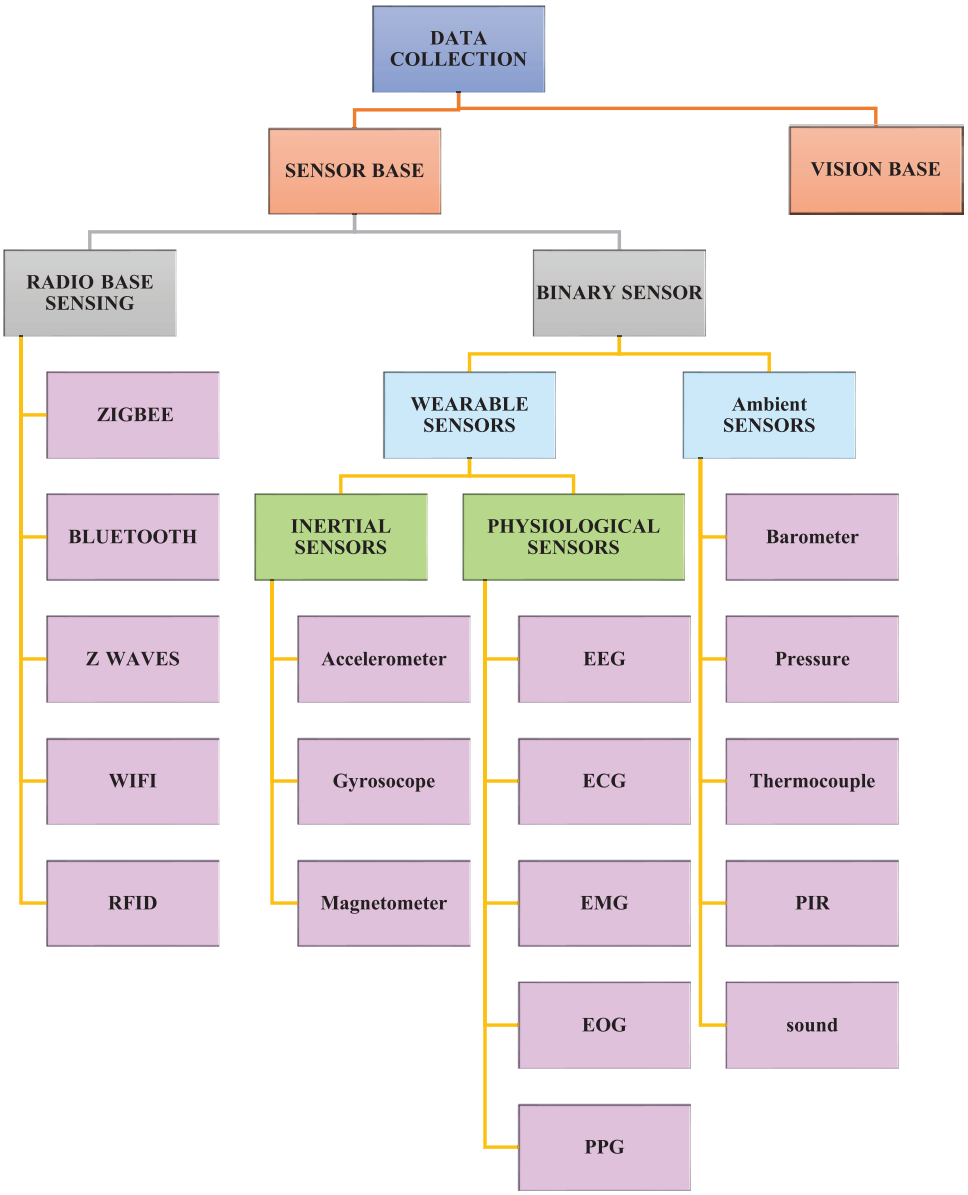

Figure 5: Various sensors for collecting data in HAR systems

4.3.1 Wearable Sensors

Wearable sensors play a pivotal role in Human Activity Recognition (HAR) systems, providing critical data that allow for continuous monitoring of both physical activities and physiological states. These sensors are seamlessly integrated into everyday portable electronics such as glasses, helmets, smartwatches, smart bands, and smartphones, enabling unobtrusive and constant data acquisition. Inertial and physiological sensors are the two categories of wearable sensors. Inertial sensors provide data on movement and orientation, crucial for detecting physical activities like walking, running, or cycling.

The most common inertial sensors are magnetometers, gyroscopes, and accelerometers [81]. Meanwhile, physiological sensors offer insights into an individual's internal state, such as stress levels, brain activity, and cardiac health, enabling a deeper understanding of the user's overall health and activity patterns. The most often used physiological signals are electromyogram (EMG) [82], electroencephalogram (EEG) [83,84], electrocardiogram (ECG) [85], electrooculogram (EOG) [86], and photoplethysmography (PPG) [87].

Accelerometers: An accelerometer is a crucial motion sensor used extensively in Human Activity Recognition (HAR) systems to detect changes in the velocity of a moving object. This sensor measures acceleration in units of gravity (g) or meters per second squared (m/s^2), providing insights into the intensity and direction of motion. Accelerometers typically operate at sampling frequencies ranging from tens to hundreds of Hz, enabling them to capture a wide range of human movements. They are equipped with three axes (X, Y, and Z), allowing them to generate a three-variate time series that offers a comprehensive view of the wearer's movements in three-dimensional space. Accelerometers can be attached to multiple body parts, such as the arm, wrist, ankle, and waist, making them versatile tools for monitoring daily activities [88].

Gyroscopes: Gyroscopes complement accelerometers by measuring an object's orientation and angular velocity rather than linear acceleration. Unlike accelerometers, gyroscopes are specifically designed to track the rotational movements around an axis [89]. A gyroscope consists of a rotating wheel fixed within a frame. The principle behind its operation is the conservation of angular momentum: the spinning wheel tends to maintain its orientation, remaining unaffected by external forces. When the direction of the gyroscope's axis changes, a torque proportional to the rate of change in orientation is generated. This torque is crucial for calculating angular velocity, making gyroscopes essential for accurately determining orientation changes, such as tilts and turns [90].

Magnetometer: Magnetometers measure the strength and direction of magnetic fields, primarily those generated by the Earth. These sensors are vital for determining orientation by detecting the planet's magnetic poles, complementing the data provided by accelerometers and gyroscopes. They operate by measuring the induction caused by moving charges or electrons, typically within the frequency range of tens to hundreds of Hz. Magnetometers also feature a triaxial setup, providing three-dimensional data on magnetic field orientation. In HAR, magnetometers are frequently used for tasks requiring precise [89].

Electromyogram (EMG): Electromyography (EMG) is a biomedical signal that reflects neuromuscular activity within the body. EMG signals are typically captured using specialized sensors known as electromyogram sensors. These signals serve various purposes, including tracking health irregularities, measuring muscle activation, and analyzing the biomechanics of movement [91].

Electroencephalogram (EEG): The Electroencephalogram (EEG) is a vital physiological sensor that detects electrical signals generated by the brain. EEG sensors record the activity of large populations of neurons near the brain's surface over time. EEG signals provide insights into brain activity and are essential for understanding cognitive processes and neurological conditions [92,93].

Electrocardiogram (ECG): An ECG sensor records the electrical activity of the heart using electrodes placed on the skin. It provides valuable information about cardiac function and rhythm [94]. Heart rate

(HR), derived from ECG signals, is a crucial indicator of physical and mental stress, making ECG a valuable tool for assessing human activities [95].

Electrooculogram (EOG): Electrooculography (EOG) measures the corneo-retinal standing potential, which reflects the electrical potential difference between the retina and the front of the eye [96]. EOG signals, captured using electrodes placed around the eyes, can identify sleep states [97], track eye movements [98], and detect human activity [99].

photoplethysmography (PPG): PPG offers an alternative method for monitoring heart rate and cardiovascular rhythm by detecting variations in vascular tissue's light absorption during the cardiac cycle. PPG signals are commonly obtained using pulse oximeter sensors integrated into wearable devices like smartwatches [100,101]. PPG signals complement other sensor data, such as inertial measurement unit (IMU) or ECG signals, to enhance the accuracy of Human Activity Recognition (HAR) systems [102,103].

4.3.2 Ambient Sensors

In smart environments, wearable devices offer a direct method for recognizing human activities. However, this method has notable limitations. Many users, especially in smart homes, may find these devices uncomfortable and inconvenient. Additionally, there's a risk of forgetfulness in usage, particularly among elderly residents. Thus, ambient sensors present a preferable alternative for non-intrusive data collection in such settings. Ambient sensors are integrated into the environment, capturing interactions between humans and their surroundings without causing any disturbance [15]. These sensors include contact sensors, temperature sensors, pressure sensors, passive infrared (PIR) sensors, light sensors, sound sensors, and more. Although ambient sensors efficiently collect data, interpreting this data and resolving associated challenges requires advanced techniques.

Passive Infrared (PIR) Sensor: A passive infrared (PIR) sensor consists of two primary components: a Fresnel lens or mirror, which directs infrared signals toward the sensor, and a pyroelectric sensor, which measures the intensity of the infrared radiation. PIR sensors are typically classified into two categories: binary-based and signal-based. The binary type, commonly used in various applications, such as controlling lighting systems and triggering alarms, provides a simple binary output, indicating '1' for detected motion and '0' for no motion. This straightforward functionality makes PIR sensors a popular choice for motion detection in smart environments, enabling efficient monitoring and documentation of the presence of occupants in specific locations like bedrooms, kitchens, and bathrooms [104].

Pressure Sensors: These sensors measure the force exerted over an area to detect physical interactions like touch. They help monitor whether a person is sitting or lying down on furniture such as chairs or beds, thus discreetly logging their presence [4].

Contact Switch Sensors: These sensors can be attached to a variety of objects, including bedroom doors, living room furniture, kitchen cabinets, or refrigerators, to determine how residents interact with their surroundings. This setup enables a detailed understanding of user behavior within the environment [7].

Temperature and Humidity Sensors (TH): Monitoring changes in temperature and humidity is essential for recognizing activities within a smart home. Sensors that track these environmental factors are strategically positioned in locations like bathrooms and kitchens, where they can detect specific activities, such as cooking or showering. These activities usually cause increases in humidity and temperature, making such sensors valuable for providing insights into the home's dynamics [105].

Light Intensity Sensors (L): These sensors measure the brightness of light within a specific area, playing a crucial role in understanding user behavior, such as determining if a user is asleep. While presence sensors can detect movement, they may not indicate whether someone is actually sleeping, especially during daylight hours. In contrast, a significant drop in light levels often suggests that the individual is likely asleep. Additionally, the use of light intensity sensors is beneficial in other areas like bathrooms, where changes in lighting can further inform about the activities and habits of residents. [105].

Sound (Acoustic) Sensors: These sensors employ sound waves to detect a variety of noises, such as footsteps, voices, or the sound of breaking glass. This capability allows them to play a crucial role in monitoring and security within areas where they are installed [4].

Floor Sensor: Floor sensing is pivotal in creating environments that are both sensitive and non-invasive. Floor sensors are integrated seamlessly beneath the surface, maintaining the appearance of a conventional floor while functioning to monitor activity. These sensors can be deployed across both private and public settings. In smart buildings, for example, floor sensors detect human presence, automating control of lighting and heating systems to enhance efficiency and comfort. In eldercare settings, they provide vital functionality by detecting falls and other emergencies, ensuring timely assistance. Additionally, these sensors are useful in public spaces for counting people and monitoring crowd dynamics during events, contributing to safety and operational management [106,107].

4.4 HAR Dataset

When selecting a dataset for smart living applications, it is crucial to consider the type of human activities represented. These activities should be relevant to the specific contexts of smart living, such as homes, offices, or urban settings, and should reflect the typical duties and routines of individuals. Activities in these datasets can generally be categorized into two types: simple activities, which include actions like walking, lying down, running, jogging, and climbing stairs; and complex activities, such as cooking, cleaning the kitchen, and washing clothes. Other vital factors to consider in choosing a Human Activity Recognition (HAR) dataset include the quality of the data, the types of sensors used, the number of sensors deployed, sensor placement, the duration of recorded activities, and the number of participants. These criteria are essential for ensuring the selected dataset is well-suited to the application, as they significantly influence the performance and reliability of HAR models. Table 1 lists the major publicly available datasets designed specifically for human activity recognition, providing a comprehensive overview to aid in the selection process.

Table 1: A summary of publicly available datasets for human activity recognition

Dataset name	Year	Environment	Type and number of sensors	Number of activities	Duration	Multi-resident	Citations/ (Cite per Year)	Access link
Kasteren Ubicomp [108]	2008	Smart home	Ambient sensors	8	28 days	No	1135/(70)	Link
Skoda [109]	2008	Car Manufacturing	Inertial measurement unit (IMU), Force-sensitive resistor (FSR)	11	–	No	456/(31)	Link
CASAS: Kyoto ADL [110]	2009	4 Smart home	Ambient sensors (motion, item, cabinet, water, temperature)	6	15 days	No	478/(31)	Link

(Continued)

Table 1 (continued)

Dataset name	Year	Environment	Type and number of sensors	Number of activities	Duration	Multi-resident	Citations/ (Cite per Year)	Access link
CASAS: Milan [110]	2009	Smart home	Ambient sensors (motion, temperature, door closure) (33)	15	72 days	Yes	478/(31)	Link
CASAS: Cairo [110]	2009	Smart home	Ambient sensors	13	–	Yes	478/(31)	Link
OPPORTUNITY [111]	2010	Lab	Wearable + Ambient sensors (72)	18	25 h	No	835/(60)	Link
CASAS: Kyoto Multi-resident [112]	2010	Smart home	Ambient sensors (motion, item, cabinet, water, temperature) (37)	15	2 months	Yes	330/(23)	Link
HASC [113]	2011	Lab	Inertial sensors embedded in smartphone	6	–	No	169/(13)	Link
VanKasteren [114]	2011	3 Smart home	Ambient and wireless sensors	10-13-16	25-14-19 days	Yes	278/(22)	Link
PAMAP2 [115]	2012	Lab	Inertial (accelerometers, Gyroscope, Magnetometer), heart rate sensor	18	10 h	No	1228/(102)	Link
CASAS: Tulum [116]	2012	2 Smart home	Ambient sensors	16	–	Yes	583/(38)	Link
CASAS: Aruba [116]	2012	Smart home	Ambient sensors (34)	11	220 days	No	583/(38)	Link
USC-HAD [117]	2012	Lab	Accelerometer, Gyroscope, Magnetometer	12	6 h	No	542/(45)	Link
RealDISP [118]	2012	Lab	Wearable sensors (Accelerometers, Gyroscopes, Magnetometer) (9)	33	9 h	No	195/(15)	Link
UCI-HAR [119]	2013	Lab	Accelerometers and Gyroscopes embedded in smartphone	6	–	No	2506/(227)	Link
ARAS [120]	2013	2 Smart home	Ambient sensors (20)	27	2 months	Yes	213/(20)	Link
WHARF [121]	2013	Lab	Accelerometer	12	–	No	176/(15)	Link
DSADS [122]	2014	Lab	Magnetometers, Inertial sensors (5)	19	–	no	390/(36)	Link
MHEALTH [123]	2015	Controlled	Wearable sensors (Accelerometers, Gyroscopes, MAGNETOMETER, ECG)	12	–	No	350/(38)	Link
HHAR [124]	2015	Controlled	Accelerometers and Gyroscopes embedded in smartphone and smartwatch	6	–	No	758/(84)	Link
UTD-MHAD1 [125]	2015	Lab	Accelerometer, Gyroscope	27	–	No	914/(90)	Link

(Continued)

Table 1 (continued)

Dataset name	Year	Environment	Type and number of sensors	Number of activities	Duration	Multi-resident	Citations/ (Cite per Year)	Access link
UTwente [126]	2016	Lab	Accelerometer, Gyroscope, and Linear Acceleration embedded in two mobile phones	13	–	No	466/(58)	Link
REALWORLD [127]	2016	Controlled	Wearable sensors (Accelerometers, Gyroscopes, Magnetometer)	8	–	No	315/(40)	Link
AreM [128]	2016	Lab	Inertial sensors embedded in the smartphone, wearable sensors	7	–	No	172/(20)	Link
MobiAct 1 [129]	2016	Lab	Accelerometer, Gyroscope, Orientation Sensors embedded in smartphone	9	–	No	277/(34)	Link
MobiAct 2 [130]	2017	Lab	Inertial sensors (Accelerometers, Gyroscope)	12	–	No	105/(15)	Link
UniMiB SHAR [131]	2017	Lab	Accelerometers embedded in smartphone	9 + 8	–	No	518/(74)	Link
ExtraSensory [132]	2017	Lab	Accelerometers, Gyroscope, GPS embedded in smartphone	10	–	No	301/(43)	Link
ContextAct@A4H [133]	2017	Smart home	Ambient sensors + Actuators (219)	7	1 month	No	62/(9)	Link
SIMADL [134]	2018	Smart home	Ambient sensors (29)	5	63 days	No	36/(6)	Link
DU-MD [135]	2018	Lab	Wearable (accelerometer)	7	–	No	42/(7)	Link
HuGaDB [136]	2018	Lab	Inertial (Accelerometers, Gyroscope)	12	10 h	No	106/(17)	Link
MotionSense [137]	2018	Lab	Accelerometers and Gyroscopes embedded in smartphone	6	–	No	170/(28)	Link
WISDM [138]	2019	Controlled	Accelerometers and Gyroscopes embedded in smartphone and smartwatch	18	60 h	No	308/(61)	Link
Ulster [139]	2019	Lab	Accelerometer, Gyroscope, GPS, light sensors, embedded in smartphone and smartwatch	6	6 weeks	No	26/(5)	Link
W-HAR [140]	2020	Controlled	Inertial (Accelerometers, Gyroscope) and Stretch sensors	7	–	No	69/(17)	Link
MARBLE [141]	2021	Lab	Wearable + Ambient Sensors (8)	13	16 h	No	12/(4)	Link

(Continued)

Table 1 (continued)

Dataset name	Year	Environment	Type and number of sensors	Number of activities	Duration	Multi-resident	Citations/ (Cite per Year)	Access link
CSL-SHARE [142]	2021	Lab	Wearable sensors (Accelerometers, Gyroscopes, EMG, Electrogoniometer, Airborne microphone) (10)	22	2 h	No	31/(10)	Link
SaMO-UJA [143]	2021	Smart lab	Ambient sensors + BLE beacons (85)	25	9 days	Yes	5/(1)	Link
HARTH [144]	2021	Lab	Wearable (Accelerometer) (2)	10	2 h	No	88 (20)	Link
KU-HAR [145]	2021	Lab	Accelerometers and Gyroscopes embedded in smartphone	18	–	No	133 (33)	Link
SDHAR-HOME [105]	2022	1 Smart home	Wearable + Ambient sensors (35) + Positioning (7)	18	2 Months	Yes	28/(8)	Link
HAR70+ [146]	2023	Lab	AX3 accelerometers	6	–	NO	51/(20)	Link

5 Deep Learning Models in HAR

The design and implementation of Human Activity Recognition (HAR) systems involve a systematic process encompassing data acquisition, preprocessing, model training, and deployment. This process begins with the collection of data through diverse sensors including ambient, radio-based, physiological, and inertial types. The data then undergo preprocessing steps such as cleaning, normalization, and feature extraction. Subsequently, an appropriate machine learning model is selected, trained, and utilized to predict activities based on processed data [147].

In recent years, deep learning (DL) approaches have surpassed traditional machine learning methods in various HAR tasks due to several key advancements. The widespread availability of large datasets has facilitated the development of models that effectively learn complex patterns and relationships, significantly enhancing performance. Moreover, advances in hardware acceleration technologies like Graphics Processing Units (GPUs) and Field-Programmable Gate Arrays (FPGAs) [148] have dramatically shortened model training times. These technologies enable more rapid computations and parallel processing, which accelerates the overall training process. Additionally, improvements in algorithmic techniques for optimization and training have boosted the speed and efficiency of deep learning models, enabling faster convergence and improved generalization capabilities [19].

The upcoming discussion explores numerous studies that employ deep learning techniques to recognize human activities. These studies illustrate the versatility and effectiveness of deep learning models in overcoming the challenges of human activity recognition, highlighting significant contributions to the field, and pointing towards future research directions.

5.1 Multi-Layer Perceptron (MLP)

The Multi-Layer Perceptron (MLP) is a fundamental architecture underpinning deep learning and deep neural networks (DNNs). Classified within the sphere of feedforward artificial neural networks (ANNs), the MLP consists of three principal layer types: the input layer, one or more hidden layers, and the output layer. This configuration is pivotal for its role in supervised learning. In the MLP architecture, each neuron from

one layer is fully interconnected with every neuron in the subsequent layer, ensuring a densely connected structure [149]. This architecture is critical for the effective propagation of information during the learning process, characterized by two main phases: forward and backward propagation. In forward propagation, input data are sequentially processed from the input layer through the hidden layers to the output layer, with each neuron's output computed using a combination of incoming inputs and trainable parameters. In forward propagation, input data are sequentially processed from the input layer through the hidden layers to the output layer, with each neuron's output computed using a combination of incoming inputs and trainable parameters [150]. The output at each neuron is given by the equation [151]:

$$y = \varphi\left(xw + b\right) \tag{1}$$

where x represents the input vector, w denotes the weighting vector, b is the bias term, φ is a nonlinear activation function that introduces non-linearity into the model, enabling it to learn complex patterns, and y represents the output value. Fig. 6a exemplifies the operation of a single-neuron perceptron model, highlighting the computational mechanics involving the input vector, weighting factors, bias, and the resultant output through the activation function.

Upon computing the final outputs, the network evaluates the prediction error, and backward propagation commences. During backward propagation, error gradients are transmitted back through the network from the output layer to the input layer. This process adjusts the weights and biases to minimize the error, refining the model's predictions. The backward pass relies on gradient-based optimization techniques, such as stochastic gradient descent, to update the trainable parameters effectively.

The structural composition of the MLP is elucidated in Fig. 6b, which illustrates the interconnected layers and the directional flow of data within the network.

Activation functions, or transfer functions, are essential in shaping the output of an MLP (Multi-Layer Perceptron) network. Essential activation functions include Rectified Linear Unit (ReLU) [152], hyperbolic tangent (Tanh), Sigmoid, and SoftMax [153,154]. These functions impart non-linearity to the network, enabling it to decipher complex patterns and relationships within the data. Throughout an MLP's training phase, a variety of optimization methods are used to fine-tune network parameters and reduce the loss function. Notable among these methods are Adaptive Moment Estimation (Adam) [155], Stochastic Gradient Descent (SGD) [156], adaptive gradient algorithm (ADAGRAD) [157], and Nesterov-accelerated Adaptive Moment Estimation (Nadam) [158]. These techniques are employed to iteratively adjust the weights and biases, thereby enhancing network performance [159].

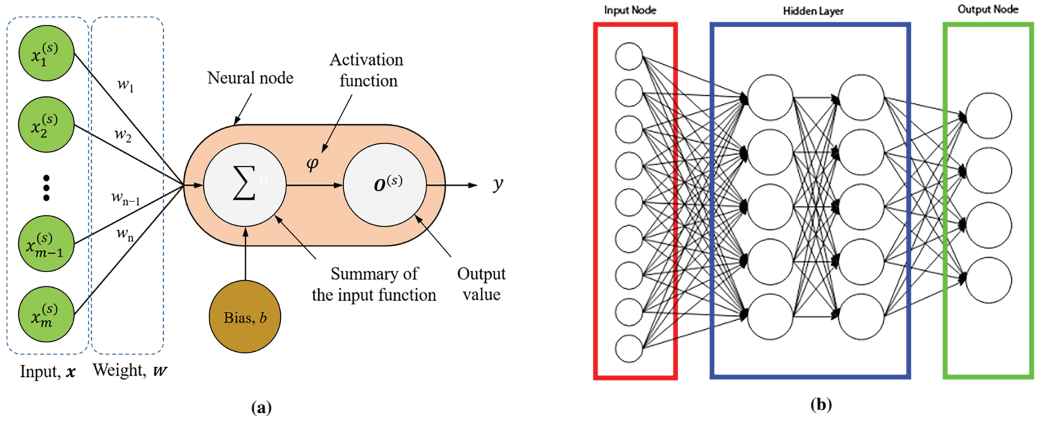

Figure 6: (**a**) Single-neuron perceptron model [151]. (**b**) Structure of the MLP [160]

Choosing the right hyperparameters is essential for constructing a neural network capable of achieving high accuracy. The performance of the network is significantly influenced by these hyperparameter settings. For example, if the number of training iterations is set too high, it may result in overfitting [161]. This occurs when the model is overly tuned to the training data, capturing noise and irrelevant patterns, and thus performs poorly on new, unseen data. Furthermore, the learning rate is a critical hyperparameter that influences the speed of convergence during training. A learning rate set too high can cause the network to converge too rapidly, potentially overlooking the global minimum of the loss function. Conversely, a learning rate set too low may lead to a protracted convergence process. Therefore, finding an optimal balance of hyperparameters is crucial to maximizing the network's performance [150].

In the field of human activity recognition (HAR) within multi-resident smart environments, researchers have widely adopted the Multi-Layer Perceptron (MLP) model for its efficacy in classification tasks, as evidenced by numerous studies [160,162].

In a novel approach, Rustam et al. [163] introduced the Deep Stacked Multilayered Perceptron (DS-MLP) model. This model leverages a meta-learning framework using a neural network as the meta-learner and five MLP models as base learners. The DS-MLP model was evaluated using the UCI-HAR and HHAR datasets, which comprise data collected from accelerometer and gyroscope sensors on smartphones. Experimental results demonstrated impressive accuracy rates of 99.4% and 97.3% for the respective datasets.

Further, another study [164] employed an MLP neural network to recognize various human activities using data from devices worn on the wrist and ankle. The authors proposed a novel data collection technique capable of distinguishing nine different activity categories. This technique utilizes an ultra-low-power STM32L4 series MCU, supplemented by various communication modules (including Bluetooth, Wi-Fi, NFC, Sub-RF, and a second Bluetooth module) and embedded Micro-Electro-Mechanical Systems (MEMS) sensors. These sensors measure mechanical movements (such as rotation and acceleration) and environmental parameters (like temperature, humidity, and proximity) [165]. The study achieved a remarkable classification accuracy exceeding 98% across all activity categories.

Further, another study [164] employed an MLP neural network to recognize various human activities using data from devices worn on the wrist and ankle. The authors proposed a novel data collection technique capable of distinguishing nine different activity categories. This technique utilizes an ultra-low-power STM32L4 series MCU, supplemented by various communication modules (including Bluetooth, Wi-Fi, NFC, Sub-RF, and a second Bluetooth module) and embedded Micro-Electro-Mechanical Systems (MEMS) sensors. These sensors measure mechanical movements (such as rotation and acceleration) and environmental parameters (like temperature, humidity, and proximity) [165]. The study achieved a remarkable classification accuracy exceeding 98% across all activity categories.

In a different exploration, Natani et al. [166] adapted the MLP architecture for multi-resident activity recognition in environments equipped with ambient sensors. By incorporating dual output layers in the MLP design, the modified architecture avoids the complexities of multi-label techniques or combined label strategies traditionally used in multi-resident activity modeling. The study, which utilized the ARAS dataset, reported accuracy rates of 69.92% for house A and 89.7% for house B.

Shi et al. [167] proposed a smartphone-assisted HAR method utilizing a residual MLP (Res-MLP) structure. This model features two linear layers with Gaussian Error Linear Unit (GELU) activation functions [168] and employs residual connections to enhance learning. Tested on the public UCI HAR dataset, the method achieved a high classification accuracy of 96.72%, underscoring the potential of Res-MLP in activity recognition tasks.

Mao et al. [169] presented a novel Human Activity Recognition (HAR) technique that incorporates an MLP neural network and extracts Euler angles from inertial measurement unit (IMU) sensors. This method initially calculates Euler angles for determining precise attitude angles, which are further refined using data from gyroscopes and magnetometers. To enhance data representation from a time domain to a frequency domain, the Fast Fourier Transform (FFT) is employed for feature extraction, thereby increasing the practical utility of the data. Moreover, the introduction of a Group Attention Module (GAM) facilitates enhanced feature fusion and information sharing. This module, termed the Feature Fusion Enrichment Multi-Layer Perceptron (GAM-MLP), effectively amalgamates features to yield accurate classification results. The technique demonstrated impressive accuracy rates of 93.96% and 96.13% on the self-created MultiportGAM and the publicly available PAMAP2 datasets, respectively.

Wang et al. [170] proposed an all-MLP lightweight network architecture tailored for HAR, which is distinguished by its simplicity and effectiveness in handling sensor data. Unlike Convolutional Neural Networks (CNNs), this architecture relies solely on MLP layers equipped with a gating unit, enabling straightforward processing of sensor data. By partitioning multi-channel sensor time series into non-overlapping patches, the model locally analyzes sensor patches to extract features, thereby reducing computational demands. Tested across four benchmark HAR datasets including WISDM, OPPORTUNITY, PAMAP2, and USC-HAD, showed that the model not only necessitated fewer Floating-Point Operations per Second (FLOPs), and parameters compared to convolutional architectures but also matched their classification performance.

Two MLP-based models, MLP-Mixer and gMLP, were utilized by Miyoshi et al. [171] for sensor-based human activity recognition. The MLP-Mixer construction [172] consists of Mixer layers that combine two different kinds of MLPs: channel-mixing MLP, and token-mixing MLP. Through the spatial mixing of information across several tokens, the token-mixing MLP can extract features. channel-mixing MLP is able to extract features by combining data in the channel direction into a single token. gMLP [173] comprises L blocks of uniform size and structure, integrating a Spatial Gating Unit (SGU) alongside conventional MLPs. The accuracy of both models was discovered in this study when the model parameters were gradually reduced, and the results indicated that accuracy is proportional to the model parameters. The experimental results show that MLP-based models can compete with current CNNs and do not perform worse.

5.2 Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs) are distinguished as a leading class of deep learning algorithms, predominantly utilized for their robust performance in image processing and pattern recognition tasks. Unlike conventional machine learning techniques, CNNs excel at autonomously extracting features from unprocessed data and discerning patterns to classify behaviors [149]. The design of CNN draws significant inspiration from the mechanisms of visual perception observed in biological organisms [147]. The architecture of CNN is fundamentally characterized by its use of convolutional layers for feature extraction, distinguishing it from traditional Artificial Neural Networks (ANNs).

There are several pivotal advantages that set CNNs apart from ANNs: 1) Local Connections: In CNNs, each neuron is linked only to a small subset of neurons in the preceding layer, rather than to every neuron. This localized connectivity reduces the overall number of parameters, expediting the learning process and convergence. 2) Weight Sharing: This feature involves reusing the same weights for multiple connections within the network. By reducing the number of unique parameters, weight sharing simplifies the model and enhances its ability to generalize across different data sets. 3) Dimensionality Reduction via Pooling: CNNs integrate pooling layers to down sample the input data effectively. These layers capitalize on the principle of local correlation, reducing the spatial dimensions of feature maps while retaining essential information. By

omitting less significant features, pooling also decreases the parameter count, contributing further to model efficiency [149,174].

These distinctive characteristics collectively position CNNs as highly efficient and capable algorithms within the realm of deep learning, making them particularly suitable for a variety of complex computational tasks. As illustrated in Fig. 7, a CNN typically comprises several layers including convolutional layers, pooling layers, fully connected layers, and nonlinear activation functions, each contributing to the streamlined and effective processing of data through the network.

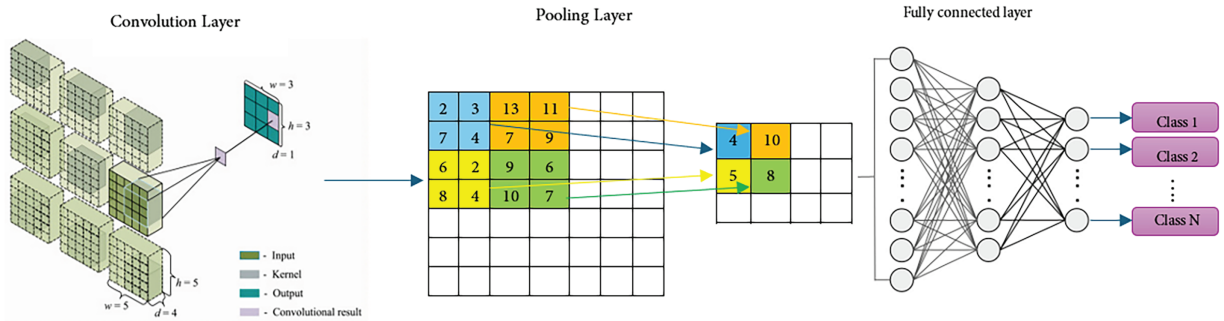

Figure 7: The pipeline of a Convolutional Neural Network

Convolution Layer: The convolution layer is a pivotal component of Convolutional Neural Networks (CNNs) and plays a critical role in the extraction of spatially invariant features through the convolution operation, utilizing shared kernels. Unlike fully connected neural networks, the convolutional layers are adept at detecting local dependencies, with each filter targeting a specific receptive field. Although the kernel of each layer covers only a modest subset of input neurons, the deployment of multiple layers results in neurons in higher layers possessing larger and more expansive receptive fields.

This hierarchical arrangement facilitates the transformation of local, low-level features into abstract, high-level semantic information [19]. Convolution layers consist of learnable convolution kernels, typically structured as weight matrices that maintain equal length and width, and are commonly odd-numbered (e.g., $3 \times 3, 5 \times 5$, or 7×7). These kernels slide across the input feature maps, performing convolution operations that are instrumental in extracting significant spatial features from the data [175].

Pooling Layer: Situated typically after the convolution layer within a CNN, the pooling layer is designed to perform dimensionality reduction and down-sampling, effectively minimizing the number of network connections. This reduction is crucial for alleviating computational demands and combating overfitting. By decreasing the spatial dimensions of the feature maps generated by the convolution layer, the pooling layer not only compresses the learned representations but also preserves essential features, enhancing the model's efficiency and robustness [176]. Various pooling techniques exist, each with specific advantages and use cases. Some common pooling techniques including Max Pooling: Selects the maximum value from each patch of the input feature map, and captures prominent features. Average Pooling: Calculates the average value across each patch, providing a smoothed feature representation [177].

Mixed Pooling: Combines methods like max and average pooling to leverage the benefits of both [178]. Stochastic Pooling: Randomly selects a maximum value from the input patch, introducing variability and promoting exploration in learning [179]. Spatial Pyramid Pooling [180] and Multi-Scale Order-Less Pooling [181]. These advanced techniques adapt the pooling process to handle varying input sizes and capture features at multiple scales. Each of these pooling strategies offers unique benefits, allowing developers to tailor the approach to meet the demands of CNN architecture and the task at hand.

Fully Connected Layer: The fully connected (FC) layer, positioned at the end of the CNN architecture, is integral to the network's decision-making process. Following the principles of a Multi-Layer Perceptron (MLP), each neuron in the FC layer is interconnected with every neuron from the previous layer. The input to this layer typically comes from the last pooling or convolutional layer and involves a flattened vector of the feature maps. This layer synthesizes the extracted features into final outputs, forming the basis for high-level reasoning and classification in the CNN [149].

Activation Functions: Activation functions are essential components in Convolutional Neural Networks (CNNs), critical for introducing non-linearity into the network. This non-linearity is vital for CNN's ability to model complex patterns and relationships within the data, enabling it to perform tasks beyond mere linear classification or regression. Without these non-linear activation functions, a CNN would merely perform linear operations, significantly limiting its capability to accurately represent the intricate, non-linear behaviors typical of many real-world phenomena [182].

Fig. 8 would typically illustrate how these activation functions modulate input signals to produce output, emphasizing the non-linear transformations applied to the input data across different regions of the function curve. In this figure, x_i denotes the input feature, while w_{ij} denotes the weight associated with the connection between the input feature x_i and neuron j. The figure illustrates that a neuron j receives n features simultaneously. The output from the neuron j is indicated by y_j, and its internal state, or bias, is denoted by b_j. The activation function, shown as $f(.)$, could be one of several types such as the Rectified Linear Unit (ReLU), hyperbolic tangent (Tanh), Sigmoid function, or others [153,154].

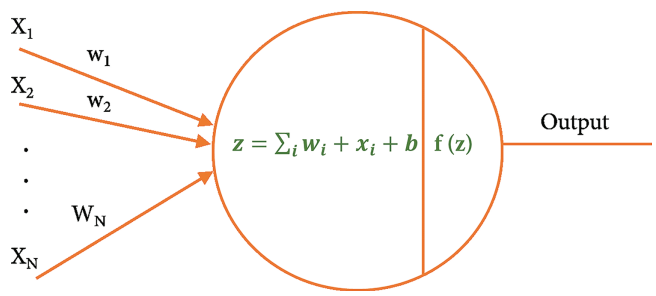

Figure 8: The general structure of activation functions

Tanh and sigmoid functions are often referred to as saturating nonlinearities due to their behavior when inputs are very large or small. Specifically, the Sigmoid function approaches values of 0 or 1, whereas the Tanh function trends towards −1 or 1 as described in reference [174]. To mitigate issues related to these saturating effects, various alternative nonlinearities have been introduced, such as the Rectified Linear Unit (ReLU) [152], Leaky ReLU [183], Parametric Rectified Linear Units (PReLU) [184], Randomized Leaky ReLU (RReLU) [185], S-shaped ReLU (SReLU) [186], and Exponential Linear Units (ELUs) [187]. Among the activation functions, ReLU (Rectified Linear Unit) is particularly popular in contemporary CNNs due to its simplicity and effectiveness in addressing the vanishing gradient problem during training. Mathematically, ReLU is defined as follows [188]:

$$ReLU\,(x) = \max\,(0, x) = \begin{cases} x, & if\ x \geq 0 \\ 0, & if\ x < 0 \end{cases} \tag{2}$$

In these definitions, x represents the input to the neuron. This function essentially "turns off" any negative input values while preserving positive values, allowing CNNs to learn complex features more effectively and preventing neurons from becoming saturated during training.

Conversely, the sigmoid function is defined as:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

where x represents the neuron's input. While sigmoid is valuable for binary classification due to its characteristic S-shape and its ability to compress real numbers into a range between 0 and 1, its tendency to saturate can impede training by leading to the vanishing gradient problem in deep neural networks.

Similarly, the hyperbolic tangent (tanh) function, like sigmoid, maps real numbers to a range between −1 and 1, providing non-linearity to the model. However, it can also suffer from issues related to vanishing gradients in deep neural networks [182].

Convolutional Neural Networks (CNNs) offer significant advantages for time series data classification in applications such as Human Activity Recognition (HAR). These advantages primarily arise from their ability to exploit local dependencies and achieve scale invariance. Local dependencies imply that neighboring time series data points are often correlated, enabling CNNs to effectively discern spatial relationships and patterns. Scale invariance, meanwhile, ensures consistent performance despite variations in scale or frequency within the input data [8].

Convolutional Neural Networks (CNNs) have become highly effective deep learning tools for Human Activity Recognition (HAR) systems in sensor-based contexts, thanks to their capability to leverage local dependencies and maintain scale invariance. While some studies have employed the simple CNN models for HAR [189–191], others have explored advanced variants, including DenseNet [192], CNN-based autoencoders [193,194], and CNN models enhanced with attention mechanisms [195,196], among others.

Qi et al. [197] introduced a robust framework termed Fast and Robust Deep Convolutional Neural Network (FR-DCNN) designed for HAR using smartphone sensors. This framework enhances data efficiency through an integrated signal selection module and various signal processing algorithms, all operating on data from inertial measurement units (IMUs). A complementary data compression module facilitates a swifter construction of the DCNN classifier, significantly optimizing computational speed.

Another novel approach is the DeepConvAttn model developed by Zhang et al. [198], which merges the convolutional neural network with an attention mechanism. Building on the established DeepConvLSTM model, comparative studies on benchmark HAR datasets have shown that DeepConvAttn delivers superior performance, attributed to its parallel processing of attention mechanisms, thereby accelerating data throughput.

Otebolaku et al. [89] developed two distinct Convolutional Neural Network (CNN) models to address the challenges of class imbalance in Human Activity Recognition (HAR) systems. The first model processed only inertial sensor data, while the second model enhanced the input feature set by integrating both inertial and ambient sensor signals. This integration leverages the CNN's inherent capabilities for scale invariance and local dependency to better capture the complex dynamics of the combined sensor data. Specifically, ambient sensor data, such as noise levels and lighting conditions, when fused with inertial sensor data, contribute significantly to the richness of the input features. This fusion leads to markedly improved recognition accuracy, as demonstrated through the evaluation and analysis of the system using datasets characterized by imbalanced class distributions. This study highlights the potential of multimodal sensor integration in enhancing the performance of HAR systems under challenging real-world conditions.

In the context of multi-resident activity recognition, Cao et al. [199] introduced an advanced end-to-end framework for Multi-Resident Activity Recognition utilizing Tree-Structured Convolutional Neural Networks (CNNs). This architecture is designed to capture temporal dependencies among sensor readings in

close proximity, which facilitates the automatic extraction of relevant temporal features. The integration of a fully connected layer processes these features to concurrently classify both the residents and their activities. This comprehensive approach not only streamlines the recognition process but also enhances the accuracy of activity predictions across multiple residents within the same environment.

Bouchabou et al. [200] developed a novel HAR system for smart homes, employing an end-to-end architecture that merges frequency encoding with a fully convolutional network (FCN). This method eliminates the need for manual feature engineering and extensive preprocessing, leveraging frequency-based embedding to enhance input data representation directly from raw signals.

Ata et al. [201] introduced a novel Convolutional Neural Network (CNN) model tailored for human activity recognition. They employed an innovative stream-based CNN architecture, incorporating various optimization techniques to enhance model performance. The focus of their research was on a spectrum of human activities, which were represented through signal data collected from gyroscope and accelerometer sensors attached to multiple body parts. This sensor data served as the input for the proposed networks. By using this approach, they were able to efficiently process the complex patterns inherent in the sensor signals, potentially leading to more accurate and reliable recognition of diverse human activities.

Ataseven et al. [202] have devised a sophisticated system for real-time physical activity recognition by employing deep transfer learning techniques. Utilizing acceleration data from Inertial Measurement Units (IMUs), they adapted a pre-trained GoogLeNet convolutional neural network model [203] to suit their needs. To effectively integrate IMU data into the GoogLeNet architecture, which was not originally designed for handling such data, they introduced three innovative data transform techniques based on continuous wavelet transform. These are Horizontal Concatenation (HC), which aligns data streams side-by-side; Acceleration-Magnitude (AM), which focuses on the magnitude of acceleration vectors; and Pixelwise Axes-Averaging (PA), which averages data across different sensor axes at each pixel point. These transformations are critical for rendering IMU data compatible with the input requirements of GoogLeNet, thereby enabling the refined model to process and recognize physical activities with enhanced accuracy. This approach not only leverages the powerful feature extraction capabilities of GoogLeNet but also tailors it to the unique characteristics of sensor-based activity data, significantly advancing the field of wearable technology-based activity recognition.

Another novel approach for Human Activity Recognition (HAR) named MLCNNwav was presented by Dahou et al. [204], which integrates the strengths of residual convolutional neural networks (CNNs) and one-dimensional trainable discrete wavelet transform (DWT). This innovative architecture utilizes a multilevel CNN to capture global features essential for recognizing a wide range of activities, while simultaneously employing the wavelet transformation to learn activity-specific features. The integration of DWT allows for a more refined feature extraction process, capturing both time and frequency domain information that is crucial for accurately classifying complex human movements. The synergy between the deep learning capabilities of residual CNNs and the precise feature delineation afforded by DWT enhances both the representation and generalization power of the model. This dual approach ensures that MLCNNwav effectively handles the variability and intricacies of sensor-based activity data, leading to significant improvements in HAR performance.

Kobayashi et al. [205] introduced MarNASNets, a deep learning model designed for sensor-based Human Activity Recognition (HAR), utilizing the innovative approach of Neural Architecture Search (NAS) [206,207]. They implemented Bayesian optimization to methodically explore various architectural options, focusing on optimal configurations for the convolution process. Key parameters varied included kernel size, types of skip operations, the number of convolutional layers, and the number of output filters. Experimental findings underscore the efficiency of MarNASNets, demonstrating that they achieve

comparable accuracy to conventional CNN models while requiring fewer parameters. This attribute makes MarNASNets particularly suitable for on-device applications where computational resources are limited.

Lafontaine et al. [208] proposed an unsupervised deep convolutional autoencoder, utilized specifically for denoising the scattering matrix data from Ultra-WideBand (UWB) radar [209], an advanced ambient sensor technology. The autoencoder focuses on isolating and removing unique background noise patterns from the data, a critical step facilitated by the encoder's restricted component. Following min-max normalization, the model employs a mean squared error (MSE) loss function during training, which aids in reconstructing the primary features from the noisy input. The effectiveness of this CNN-based autoencoder in enhancing HAR data quality through unsupervised filtering has been validated through rigorous testing.

A new analytical method called sensor data contribution significance analysis (CSA), which assesses the impact of various sensors on behavior recognition in HAR systems, was developed by Li et al. [210]. This approach utilizes a novel metric based on the frequency-inverse type frequency of sensor status. To enhance data reliability and context awareness, they also designed a spatial distance matrix that considers the physical arrangement of ambient sensors. The culmination of this research is the proposal of the HAR_WCNN algorithm, which integrates wide time-domain convolutional neural networks with data from multiple environmental sensors to recognize daily behaviors effectively.

Furthermore, an online HAR framework designed for real-time processing on streaming sensor data was presented by Najeh et al. [211]. This framework incorporates stigmergy-based encoding and CNN2D classification, paired with real-time dynamic segmentation. This segmentation process determines whether consecutive sensor events should be grouped as a single activity segment. The encoded features, structured in a multi-dimensional format suitable for CNN2D input, leverage a directed weighted network (DWN) to account for overlapping actions and capture the spatiotemporal trajectories of human movement, enhancing the real-time analytical capabilities of the system.

In another study, Tan et al. [212] developed a novel convolutional neural network architecture enhanced with a multi-head attention mechanism (CNN-MHA). This architecture features several attention heads, each autonomously determining the attention weights for different input segments. The integration of these weights occurs at a fully connected layer that synthesizes the final attention representation. The multi-head attention design enables the network to preserve long-term dependencies within the input data while focusing on salient features, significantly improving the precision and effectiveness of the processing.

A multi-input model for Human Activity Recognition (HAR) using hybrid CNNs was introduced by Lai et al. [213] that incorporates fundamental CNN structures with advanced squeeze-and-excitation (SE) blocks [214] and residual blocks [215]. The model uses multiple transformed datasets derived from the UCI-HAR and MHEALTH databases through fast Fourier, continuous wavelet, and Hilbert-Huang transform methods. These datasets serve as heterogeneous inputs to a system that synergizes one- and two-dimensional CNNs with SE and residual blocks, enhancing the model's ability to handle complex time-frequency domain data effectively.

Sezavar et al. [216] proposed DCapsNet, a deep neural network that merges traditional convolutional layers with a capsule network (CapsNet) framework [217,218]. This architecture is specifically designed for classifying and extracting activity features from integrated sensor data. The convolutional layers of the model are better suited for processing temporal sequences, producing scalar outputs but not capturing equivariance. To enhance the model's classification efficiency, a dynamic routing method is used to train the capsule network (CapsNet), enabling it to detect equivariance in terms of both magnitude and orientation.

The practical application of a HAR system based on photoplethysmography (PPG) signals was explored by Ryu et al. [219]. They collected PPG data from 40 participants engaged in daily activities to build a dataset,

which was then used to train a 1D CNN model to classify five distinct activities. Their analysis determined that a 10-s window size was optimal for the input signal, showcasing the feasibility of PPG-based HAR systems for real-world applications through comprehensive performance evaluations.

A novel HAR model proposed by Yu et al. [220], called ASK-HAR, makes use of several convolution kernels that offer reliable characteristics and enable each neuron to adaptively choose the right receptive field size based on the input content. Standard CNN architecture strives to keep the kernel size constant within the same feature layer. The suggested model, however, uses the softmax attention mechanism rather than linear splicing to merge several branches with different kernel sizes. Each layer's convolution output is then integrated and supplied to the Convolutional Block Attention Module (CBAM) [221]. The CBAM seeks to alleviate the limitations of traditional convolutional neural networks by handling data with different dimensions, shapes, and orientations.

These summaries provide a view of the latest advancements in HAR technologies, reflecting the diverse CNN-based approaches employed to tackle different aspects of activity recognition. Each study contributes uniquely to the field, showcasing the potential of deep learning technologies in enhancing the accuracy and efficiency of HAR systems. The summary of the proposed CNN-based HAR systems is shown in Table 2.

Table 2: A summary of CNN-based HAR systems

Reference	Year	Proposed CNN method	Dataset(s)	Performance (Accuracy)	Issues handled/achievements
Qi et al. [197]	2019	C1-BN1-C2-BN2-C3-BN3-C4-BN4-MP-D-FC-S	Self-recorded data with 12 activities	94.18%	Fast computations
Zhang et al. [198]	2020	I-C1-C2-C3-C4-A1-A2-S	OPPORTUNITY Skoda	90.6% 96.5%	Improve processing speed because attention layers can be processed in parallel
Otebolaku et al. [89]	2020	C1-P1-C2-P2-C3-P3-D-F-FC1-FC2-S	self-collected data using inertial sensors self-collected data using inertial and ambient sensors	93.6% 98.9%	Combined inertial and ambient sensors data
Cao et al. [199]	2020	Tree-structure CNN (C1-C2-C3-. . .Ck-1)-FC	CASAS: Kyoto multi-resident	Resident: 79.22% Activity: 58.48%	Capturing multi-resident issue
Bouchabou et al. [200]	2021	(1DC1-BN1 + Relu)-(1DC2-BN2 + Relu)-(1DC3-BN3 + Relu)-1D GAP-S	CASAS: Aruba CASAS: Milan	92.44% 90.86%	Improv performance and do not need heavy pre-processing because of frequency encoding

(Continued)

Table 2 (continued)

Reference	Year	Proposed CNN method	Dataset(s)	Performance (Accuracy)	Issues handled/achievements
Ata et al. [201]	2022	I-C1-MP1-D1-C2-MP2-D2-C3-MP3-D3-C4-MP4-D4-F-FC1-FC2-FC3	WISDM	97.5%	Compared different optimizers and show NAdam as best them
Dahou et al. [204]	2023	Initial Block(C1-BN1 + Relu-MP)-An ensemble of ConvBlock (C1-BN1 + Relu-C2-BN2-MP-DWT-CE)-CE-FC-BN + D-S	OPPORTUNITY UCI-HAR UNIMIB-SHAR WISDM	93.19% 95.52% 92.46% 99.14%	Recorded high-accuracy rates on all datasets
Kobayashi et al. [205]	2023	multiple convolutional blocks (MBConv)	HASC UCI-HAR WISDM	92.6% 95.33% 90.62%	Achieve reasonable accuracy with fewer parameters
Lafontaine et al. [208]	2023	C1-C2-P-F-FC1-FC2-TC1-TC2-TC3	Self-collected data using UWB radars with 14 activities	69.9%	Use of UWB radar as an ambient sensor for collecting data in order to more informative data
Li et al. [210]	2023	EM-C-P-FC	CASAS: Cairo CASAS: Milan	91.99% 95.35%	Use of a spatial distance matrix based on the layout of ambient sensors for context-awareness.
Najeh et al. [211]	2023	C-GAP-F-FC	CASAS: Aruba	81.6%	Dynamic segmentation of sensor events for real-time HAR
Tan et al. [212]	2023	C1-MP1-C2-MP2-C3-MP4-MHA-D-FC1-FC2	PAMAP2	95.7%	Solve long-term dependencies using multi-head attention mechanism

(Continued)

Table 2 (continued)

Reference	Year	Proposed CNN method	Dataset(s)	Performance (Accuracy)	Issues handled/achievements
Lai et al. [213]	2024	Ensemble of model a [C1-C2-MP-FC1-FC2] and model b[C1-C2-SE(GAP + FC + FC)-MN-Residual(C + C)-Residual(C + C)-C3-C4-GAP-FC1-FC2]-CE-FC-FC-FC	UCI-HAR UCI-HAR and MHEALTH	95% 96%	Improving performance using a hierarchical modeling process
Sezavar et al. [216]	2024	C1-BN1-C2-BN2-C3-BN3-C4-BN4-C5-BN5-Primary Capsule Layer-Activity Capsule Layer	UCI-HAR WISDM	98.43% 95.26%	The model able to extract more discriminative features
Ryu et al. [219]	2024	C1-C2-MP1-D1-C3-C4-MP2-D2-C5-C6-MP3-D3-C7-C8-MP4-D4-C9-C10-MP5-D5-FC1-FC2-FC3-FC4-FC5	Self-collected data using PPG signal with 5 activities	95.14%	Use of photoplethysmography (PPG) as a single biometric signal for practical HAR applications
Yu and Alqaness [220]	2025	4-branch SK (Selective Kernel) multikernel convolution + CBAM	UCI-HAR USC-HAD WISDM PAMAP2 DSADS	97.25% 89.40% 98.46% 94.67% 89.42%	Aggregating multi-scale data from multiple kernels of different sizes, to effectively handle data with diverse dimensions, forms, and orientations.

Note: I-Input Layer, C-Convolution Layer, A-Attention Layer, S-SoftMax Layer, BN-Batch Normalization, P-Polling Layer, MP-Max Polling Layer, D-Dropout Layer, FC-Fully Connected (Dense) Layer, F-Flattened Layer, EM-Embedding Layer, GAP-Global Average Polling, CE-Concatenate Layer, TC-Transposed Convolutional Layer, MHA-Multi-head Attention Layer, CBAM-Convolutional Block Attention Module.

5.3 Recurrent Neural Network (RNN)

A Recurrent Neural Network (RNN) is a specialized class of neural networks designed to process time-series or sequential data effectively. RNNs are unique in their ability to utilize the outcomes of previous computations as inputs into subsequent stages, thereby capturing temporal dependencies inherent in the data sequence [159]. In an RNN, each node, or module, repetitively takes both the current and the previous timestep's outputs as inputs, using these along with a set of trainable parameters to compute its output. Typically, the output of each module is determined through the hyperbolic tangent (tanh) activation

function, which helps to regulate the information flow by maintaining output values within a normalized range of −1 to 1 [222].

A simple recurrent neural network is shown in Fig. 9, where the internal memory $(h_t)$of an RNN module at time step t is calculated using the following equation [223,224]:

$$h_t = g\left(Wx_t + Uh_{t-1} + b\right) \tag{4}$$

where x_t is the input vector at timestep t, W and U represent the weight matrices for the input x and hidden state h, respectively, b denotes the bias, and $g()$ is the activation function, typically tanh.

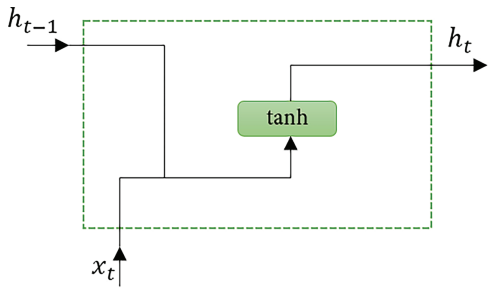

Figure 9: Standard RNN unit

In the field of human activity recognition (HAR) using sensor data, Recurrent Neural Networks (RNNs) have shown remarkably effective results. For instance, a study referenced as [225] implemented an RNN model to detect activities in a multi-resident setting, capitalizing on distinctive patterns derived from individual residents' past activities, such as interaction, frequency, duration, and general behavior. This study involved modifications to the ARAS dataset, primarily the amalgamation of similar activity categories into broader classes, achieving impressive model accuracies of 90.97% and 91.82% for two separate households.

Similarly, Ramos et al. [105] engineered a tailored RNN architecture for HAR, which judiciously partitions incoming data. In their model, sensor data is initially processed through a dedicated time series analysis layer, while temporal data elements are reserved for processing in a subsequent dense layer. To counteract overfitting, the model incorporates three strategically placed dropout layers that introduce varying degrees of randomness at different stages of the model. The output from the RNN layer is then integrated with time-related activity data and processed through a dense layer of neurons, which facilitates dynamic adjustment of the model's internal weights. This model was validated using the SDHAR-HOME dataset, which includes data from two residents, a pet, and an occasional visitor, categorized into eighteen distinct activity types over two months. The performance outcomes were robust, with an accuracy of 89.59% for the first user and an accuracy of 86.26% for the second.

Despite their proficiency in handling temporal sequences, traditional RNNs are limited by their relatively short memory span, which restricts their capacity to maintain information across extended sequences [226]. To overcome this challenge, advanced variants such as Long Short-Term Memory (LSTM) networks, bidirectional LSTM, Gated Recurrent Unit (GRU), and bidirectional GRU have been developed.

5.4 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks [227,228], as an enhancement of recurrent neural networks (RNNs), address the challenge of long-term dependencies, a prevalent issue in sequence data processing. These networks are adept at solving the vanishing gradient problem, effectively maintaining

information across extensive sequences, as noted in recent studies [229]. The LSTM network processes the output from the previous time step and the current input at a given time step, producing an output sent to the subsequent time step. Classification usually makes use of the final hidden layer of the last time step [230]. The Long Short-Term Memory (LSTM) network architecture encompasses a memory cell denoted by c, a hidden state represented by h, and three pivotal gates: the input gate (i), the forget gate (f), and the output gate (o). These gates are integral to the LSTM's functionality, as they regulate the flow of information within the network, crucial for the operations of reading, retaining, and writing information [231]. At each time step t, the LSTM performs the following sequence of operations: [232,233].

Input Gate: Determines how much of the new input is stored in the cell. It is calculated using the formula:

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \tag{5}$$

Forget Gate: Decides the amount of information to discard from the previous cell state. It is computed as:

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{6}$$

Cell State Update: Updates the cell state by combining the input gate's output and the forget gate's decision, alongside a non-linear transformation of the input:

$$c_t = f_t \times c_{t-1} + i_t \times \tanh\left(W_c x_t + U_c h_{t-1} + b_c\right) \tag{7}$$

Output Gate: Controls the contribution of the cell state to the hidden state output. It is expressed as:

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \tag{8}$$

Hidden State Update: The new hidden state, which serves as the output for the current time step, will also be used in the computations for the next time step:

$$h_t = o_t \times \tanh\left(c_t\right) \tag{9}$$

The activation functions used in these operations include the sigmoid function $\sigma(x)$, which moderates the gate activation, allowing for a non-linear transformation that prevents activations from saturating at extreme values. The symbol $\times$ denotes element-wise multiplication. The parameters W (weights for the input x), U (weights for the hidden state h), and b (biases) are associated with each gate and the memory unit, influencing the network's learning capabilities.

This computational framework enables LSTMs to learn and remember over long sequences, making them highly effective for tasks involving sequential data where temporal dynamics are complex and extended memory is crucial. The update mechanism is visually represented in Fig. 10, which illustrates the flow of data through an LSTM unit, emphasizing the role of each gate and state update in the processing of sequential information.

While Long Short-Term Memory (LSTM) networks are highly effective for many sequence modeling tasks, they are not without limitations that may impede their broader application. A significant constraint arises from their inherent sequential computation model, which does not lend itself to parallel processing. This limitation is particularly challenging in contexts where computational efficiency is paramount. Additionally, the intricate structure of LSTM networks demands substantial computational resources, which can be a barrier in resource-constrained environments [234].

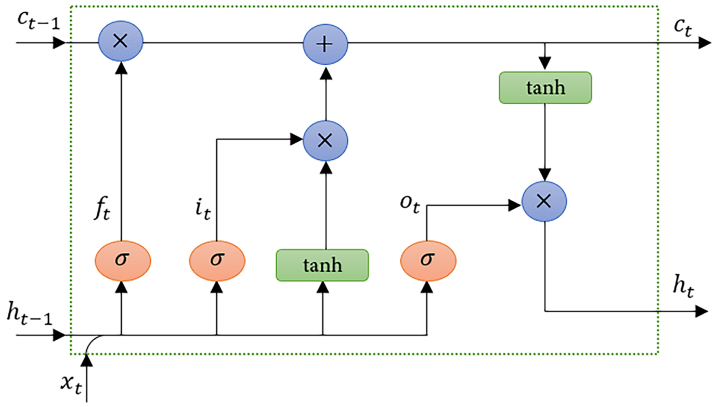

Figure 10: The inner structure of an LSTM unit

To address these limitations, a variant known as Bidirectional Long Short-Term Memory (Bi-LSTM) [235] has been developed. Bi-LSTM enhances the traditional LSTM by processing data in both forward and backward directions, thereby providing a richer understanding of the sequence by integrating contextual information from both past and future states [236]. This dual-directional approach is implemented using two parallel LSTM layers: forward Layer, where processes the input sequence from the beginning to the end, effectively capturing forward contextual information, and backward Layer, where simultaneously processes the sequence from end to beginning, capturing backward contextual details.

The bi-directional processing allows for a more comprehensive understanding of the sequence, as it incorporates insights from both ends of the data spectrum [234]. During the training phase, the forward and backward layers operate independently, extracting features and adjusting their internal states based on the sequential input they process. At every time step, both layers generate their prediction scores, which are subsequently integrated using a weighted sum approach to produce the final output. This integration enables Bi-LSTM networks to capture temporal dependencies more effectively and provide a broader context, thus enhancing the predictive accuracy and contextual awareness of the model. By leveraging insights from both the past and future contexts, Bi-LSTM models overcome some of the fundamental limitations of traditional LSTMs, offering enhanced capabilities for complex sequence modeling tasks where understanding the entire temporal context is crucial.

Recurrent Neural Network (RNN) structures, particularly Long Short-Term Memory (LSTM) networks, and their variants are notably proficient in handling time-series data due to their capability to capture temporal dependencies. In the domain of Human Activity Recognition (HAR) using sensor data, LSTM has shown exceptional utility. Its inherent memory function allows it to preserve historical data, which is crucial for recognizing patterns over time. Consequently, numerous studies have employed simple LSTM models [237–239], Bi-LSTM models [240,241], and advanced LSTM variants [242–244] to improve the performance of HAR systems.

Hu et al. [245] have notably advanced the use of LSTM in HAR by introducing a novel loss function termed 'harmonic loss' to enhance the model's efficacy in sensor-based environments. This approach begins with a label replication strategy in a many-vs.-one LSTM setup, allowing the true labels to be replicated across each sequence step, which in turn enables local output generation and error calculation at each of these steps. Subsequently, the harmonic loss, inspired by the harmonic series and the Ebbinghaus memory curve [246], is applied to assign differential weights to these local errors, emphasizing the correction of more significant errors based on their impact on model performance.

In another study, Liciotti et al. [247] explored various sophisticated LSTM architectures to further enhance HAR systems in a sensor-fused multi-residential environment. Their research began with a standard LSTM architecture and extended into more complex configurations such as bidirectional LSTM (Bi-LSTM), cascade LSTM (Casc-LSTM), and two ensemble approaches: ensemble2LSTM (Ens2-LSTM) and cascade-ensemble LSTM (CascEns-LSTM). These advanced models aim to improve the robustness and accuracy of activity detection by capturing a broader spectrum of temporal and sensor-derived data.

Gajjala and Chakraborty [248] introduced a groundbreaking approach that synergizes Long Short-Term Memory (LSTM) neural networks with Particle Swarm Optimization (PSO) [249], a novel application aimed at enhancing the performance of LSTM models specifically in Human Activity Recognition (HAR) systems. This hybrid model leverages the strengths of both methodologies to optimize LSTM parameters, a crucial factor in the effectiveness of predictive modeling.

In research conducted by the authors of [250], they proposed a deep learning model utilizing bidirectional LSTM (Bi-LSTM) for recognizing human activities in smart homes equipped with non-intrusive sensors. The architecture of the model is composed of multiple layers. The Split Layer initially segments the input data into two parts, isolating the time-related information from other sensor data. The Bidirectional LSTM Layers, which are central to the model, process the data in both forward and backward directions to capture both past and future contexts, enhancing the understanding of temporal dependencies. Dropout Layers are included to prevent overfitting by randomly omitting a percentage of neurons during the training process. The Batch Normalization Layer is designed to correct internal covariate shift, which occurs when the distribution of parameters changes across different layers of the model. The Concatenate Layer merges the temporal components with the sensor data processed by the bidirectional LSTM layers, setting the stage for the final analysis. In the last stage, two Fully Connected (Dense) Layers are employed to carry out the classification task.

A synchronous LSTM technique (sync-LSTM) was developed by Thapa et al. [251], designed to handle multiple concurrent input sequences and produce numerous synchronized output sequences in parallel. This method obviates the need for manual feature extraction or selection by automatically and directly structuring and learning spatiotemporal features from raw sensor data. The sync-LSTM algorithm improves learning efficacy and is well-suited for complex real-world Human Activity Recognition (HAR) scenarios involving simultaneous activities. The algorithm is noted for its satisfactory accuracy and computational complexity.

An innovative model for human activity recognition that eliminates the need for pre-processing sensor data was proposed by Benhaili et al. [252]. This model employs stacked Long Short-Term Memory (LSTM) layers to recognize various daily activities directly from raw sensor inputs. The approach outperformed traditional methods that rely heavily on feature design, effectively obviating the need for feature engineering. This advancement not only simplifies the deployment process but also enhances the model's ability to generalize across different sensor setups and activity types.

Cao et al. [253] introduced a model that integrates sensor signals and activity patterns into a unified framework, termed the Graph LSTM and Metric Learning model (GLML). This model incorporates multiple construction graph fusions to effectively model sensor-aspect signals alongside graph-aspect activities. GLML is designed as a multi-task classification system that addresses challenges such as imbalanced class distributions, heterogeneous multimodal data, and the presence of distinct multimodal sensor signals generated by several users within the same activity category. GLML operates on a semi-supervised co-training architecture, which leverages unlabeled data through multiple iterative pseudo-label sampling processes. This approach enhances the model's ability to learn from a larger dataset without the need for complete labeling, thus improving its generalization capabilities. Additionally, they suggest incorporating multiple graph interactions for various sensor signals using a graph attention model and attention mechanisms. This

innovation allows the system to dynamically prioritize and process sensor inputs based on their relevance and impact on activity recognition, thereby optimizing the detection and classification processes.

To enhance the performance of action identification systems and improve the accuracy of human activity recognition, Mekruksavanich et al. [254] proposed an advanced version of the long short-term memory network, referred to as RLSTM, which incorporates a squeeze-and-excitation (SE) module [214]. The SE block is a computational unit designed to adapt to any given transformation, enhancing the model's capability to focus on relevant features within the data. It utilizes a global average pooling process, referred to as the "Squeeze" operation, which aggregates background data from a broader context to generate channel-specific statistics. Following this, the "Excitation" operation analyzes these statistics to determine channel-wise dependencies, thereby optimizing the network's response to varying inputs by recalibrating feature channels according to their relevance.

A semi-supervised adversarial learning approach that employs LSTM networks for human activity recognition was introduced by [255]. This method leverages semi-supervised learning frameworks combined with adversarial learning techniques to enhance the model's ability to handle errors during the training of both annotated and unannotated data. Unlike traditional models, it does not rely on prior knowledge or historical context, which allows it to adapt to changes in human activity patterns and recognize new activities efficiently. Moreover, the model is configured to operate over time, enabling it to handle heteroscedastic uncertainty. This capability leads to enhanced prediction reliability and increased robustness of the model under diverse conditions.

In another study, Tehrani et al. [256] proposed a sophisticated multilayer bidirectional long-short term memory (Bi-LSTM) architecture designed specifically for human activity recognition. This Bi-LSTM model comprises two sub-models: one processes the input data sequence in its original order, and the other learns from the reverse sequence. This dual-path approach contrasts with traditional single-sequence LSTM models, offering a more comprehensive analysis by capturing temporal dependencies from both directions of the sequence. To further enhance model performance, particularly the average F1 score, they introduced an innovative postprocessing technique that utilizes windowing and voting mechanisms in the final stage of the model.

In the last study, El Ghazi and Aknin [257] developed an LSTM-based deep learning model incorporating batch normalization, which was further refined through hyperparameter tuning via Bayesian Optimization specifically tailored for human activity recognition (HAR) using wearable sensors in smart home settings. The integration of batch normalization helps to stabilize the learning process by reducing internal covariate shift, while Bayesian Optimization efficiently explores the hyperparameter space to find optimal settings. This approach has demonstrated superior performance over other methods, underscoring the significant role that precise hyperparameter tuning plays in enhancing the accuracy and reliability of activity classification.

These advancements collectively demonstrate the ongoing evolution of LSTM-based models to meet the diverse and complex demands of HAR systems in smart environments. By integrating innovative architectural features and learning strategies, these models offer increased accuracy, efficiency, and applicability across various real-world settings. Table 3 provides a summary of the proposed LSTM-based HAR systems in sensor-based environments.

Table 3: A summary of LSTM-based HAR systems

Reference	Year	Proposed LSTM method	Dataset(s)	Performance (Accuracy)	Issues handled/achievements
Hu et al. [245]	2020	LSTM with harmonic loss, LSTM with ensemble learning strategy	UCI-HAR Opportunity	92.98%, 93.79% 89.57%, 90.36%	Use of a novel loss function to improve LSTM performance
Liciotti et al. [247]	2020	Bi-LSTM, Cascade LSTM, Ensemble LSTM, Cascade-ensemble LSTM	CASAS: Milan CASAS: Cairo CASAS: Kyoto2 CASAS: Kyoto3 CASAS: Kyoto4	94.12%, 92.55% 94.24%, 93.60% 86.90%, 79.94% 86.23%, 83.66% 74.37%, 66.92% 74.22%, 70.06% 93.25%, 85.42% 93.38%, 88.47% 85.82%, 83.97% 86.02%, 85.20%	Use of sophisticated LSTM variants to improve performance of HAR system in multi-resident environments
Gajjala et al. [248]	2021	LSTM with PSO algorithm LS1-LS2-F-FC-S	WISDM	97.19%	Used of PSO algorithm for the optimization of the LSTM parameters
Ramos et al. [250]	2021	SP-BLS-D1-BN-CE-FC1-FC2-D2	CASAS: Milan	95.42%	The model can perform real-time prediction
Thapa et al. [251]	2021	New adapted LSTM (Sync-LSTM) SLS1-D-SLS2-D-FC1-FC2	CASAS: Kyoto3 Kasteren: HouseB	98.38% 97.42%	New adapted LSTM for recognizing concurrent activities
Benhaili et al. [252]	2022	LS1-LS2-LS3-D-F-FC	UCI-HAR WISDM	97.15% 99%	Recognizing activity without any pre-processing
Cao et al. [253]	2022	Graph LSTM and Metric Learning model (GLML)	PAMAP2 OPPORTUNITY UCI-HAR	F-score = 94.41% F-score = 83.78% F-score = 83.07%	Addresses the problems of imbalanced class distribution, multi-modal heterogeneous data, and different users performing

(Continued)

Table 3 (continued)

Reference	Year	Proposed LSTM method	Dataset(s)	Performance (Accuracy)	Issues handled/achievements
Mekruksavanich et al. [254]	2022	LS-BN-SEM-GAP-FC-S	Realworld16	98.04%	Improving accuracy
Yang et al. [255]	2022	Adversarial Sync LSTM	CASAS: Milan Kasteren: House-C CASAS: Aruba	97.26% 98.01% 98.34%	Ability to recognize complex and unannotated human activity
Tehrani et al. [256]	2023	I-BLS-LS-F-FC1-BN1-D1-FC-BN2-D2-FC3-BN3-D3-FC4	AReM MHEALTH PAMAP2	F-score = 95.46% F-score = 95.79% F-score = 93.41%	Improve performance using postprocessing approach based on windowing and voting
El Ghazi and Aknin [257]	2024	LSTM with batch normalization and hyperparameter tuning	PAMAP2	97.71%	Highlighting the crucial role of hyperparameter optimization

Note: I-Input Layer, LS-LSTM Layer, BLS-Bidirectional LSTM Layer, S-SoftMax Layer, BN-Batch Normalization, D-Dropout Layer, FC-Fully Connected (Dense) Layer, F-Flattened Layer, CE-Concatenate Layer, SP-Split Layer, SLS-Sync-LSTM, SEM-Squeeze-and-Excitation Module, GAP-Global Average Polling.

5.5 Gated Recurrent Unit (GRU)

The Gated Recurrent Unit (GRU) [258] is a variation of recurrent neural networks (RNN) designed to address short-term memory challenges. Unlike the LSTM, which uses a three-gate system and retains a cell state, the GRU employs a simpler two-gate structure, eliminating the cell state and thus reducing computational complexity for quicker learning. By merging the input and forget gates of the LSTM into one update gate, the GRU enhances efficiency [259]. It identifies long-term dependencies in data sequences through its update gate, reset gate, and memory content. The selective use of earlier data is facilitated by these gates, allowing for effective data modification and utilization. Fig. 11 illustrates the architecture of a GRU unit [260,261].

At time t, the activation of a GRU cell, represented as h_t, is computed by blending its prior activation (h_{t-1}), with a candidate activation ($\tilde{h}_t$). This process can be described as follows [262]:

$$h_t = (1 - z_t)\, h_{t-1} + z_t \tilde{h}_t \tag{10}$$

here, an update gate, represented by z_t, controls how much the unit modifies its activation or content. The equation for this gate is formulated as follows:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{11}$$

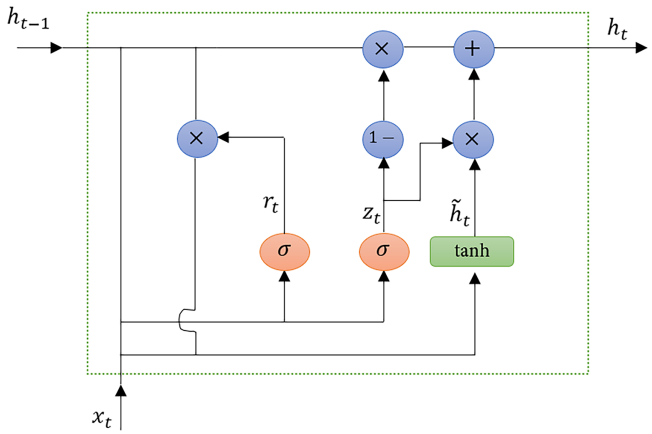

Figure 11: The inner structure of a GRU unit

This process calculates a linear combination of the current state and a newly generated state, a technique similar to that used in LSTM units. However, unlike LSTMs, GRUs do not have a mechanism to regulate the amount of state exposed; they reveal their entire state at each update. The candidate activation, denoted as $\tilde{h}_t$, is computed similarly to that in a conventional recurrent unit.

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t h_{t-1}) + b_h) \tag{12}$$

In this context, r_t represents a collection of reset gates. When r_t approaches 0, signifying an "off" state, the reset gate effectively makes the unit behave as if it is processing the initial symbol of an input sequence, thus enabling it to forget the previously computed state. The calculation of the reset gate, denoted as r_t, is similar to the process used for the update gate.

$$r_t = \sigma (W_r x_t + U_r h_{t-1} + b_r) \tag{13}$$

GRU models typically train faster than LSTMs due to their reduced complexity. However, whether a GRU or an LSTM is the better choice depends on the specific use case and requirements [231,263].

The Bidirectional Gated Recurrent Unit (Bi-GRU) [264] enhances the traditional GRU architecture by integrating contexts from both past and future in sequential modeling tasks. Unlike the standard GRU, which processes input sequences only in a forward direction, the Bi-GRU handles sequences in both forward and backward orientations [259]. This is achieved by utilizing two parallel GRU layers: one processes the input data forward, while the other processes it in reverse [265].

Several studies have employed Gated Recurrent Unit (GRU) models to advance human activity recognition systems in sensor-driven settings, as highlighted in recent literature [266–269]. Haque et al. [270] introduced a sophisticated two-layer stacked GRU architecture, meticulously designed to optimize the processing of hidden state outputs from each recurrent layer. This architecture enables the extraction of complex features from sensor data by leveraging the capabilities of stacked layers. To construct a hierarchical context vector, attention scores derived from both layers are concatenated and subsequently integrated into dense connected layers, where context-specific and simplified attention scores are computed independently. The proposed model has demonstrated superior performance over existing state-of-the-art deep learning algorithms across both balanced and imbalanced class distributions.

In a related study, the authors of [271] explored various Recurrent Neural Network (RNN) variations on the ARAS dataset, evaluating their performance using data synthesized by a Generative Adversarial Network

(GAN) [18], in addition to datasets with diverse sample sizes. The results from this study indicated that the GRU model surpassed the Long Short-Term Memory (LSTM) model in recognizing activities involving multiple residents. This comparison underscores the effectiveness of GRU models in handling complex, sequential data in multi-agent environments.

Jian et al. [272] introduced the Attention Augmented Sequential Classification (AASC) method, a robust real-time human activity recognition system that leverages an attention-augmented Gated Recurrent Unit (GRU) using radar range profiles. This approach employs an attention mechanism specifically designed to capture the temporal correlations inherent in range profile signatures, thus enabling the model to learn the long-term temporal correlations associated with human activities without expanding the recurrent neural network's size or complexity. The adaptive generation of attention weights from features processed by the GRU enhances the model's performance, showing marked improvements in accuracy and robustness compared to conventional GRU-based systems in real-time scenarios.

In another study, the Att-BiGRU model, an attention-based bidirectional gated recurrent unit designed to enhance the capabilities of recurrent neural networks, was proposed by Mekruksavanich and Jitpattanakul [273]. This innovative approach facilitates the extraction of temporal-dependent properties, essential for effective complex activity recognition by allowing the processing of sequences in both forward and reverse directions. The employment of an attention mechanism further allows the model to pinpoint and utilize key temporal features, substantially increasing the accuracy in recognizing complex human activities.

Pan et al. [274] developed a novel approach using Gated Recurrent Units (GRU) that integrates both temporal and channel attention mechanisms for human activity recognition, sidestepping the need for assuming an independent identical distribution (I.I.D) of data. Traditional methods typically assume I.I.D and overlook the variability presented by different individual volunteers. Their model employs GRU and temporal attention to synthesize and emphasize significant motion moments, aiming to reduce the number of model parameters. Concurrently, channel attention is utilized to mitigate sensor data bias, enhancing the overall efficacy and adaptability of the system in diverse real-world environments.

Sun et al. [275] developed CapsGaNet, a novel architecture for spatiotemporal multi-feature extraction in Human Activity Recognition (HAR) that integrates capsule networks (CapsNet) [217,218] with GRU and attention mechanisms. This architecture includes a temporal feature extraction layer utilizing GRU and attention, coupled with a spatial feature extraction layer composed of capsule blocks, leading to a robust output layer. Despite certain limitations, such as challenges in accurately recognizing similar activities, CapsGaNet offers significant advantages in adaptability, practicability, and recognition accuracy.

In another study, a deep novel learning approach called Bi-GRU-I was proposed by Tong et al. [276], which leverages multi-channel information through Bi-GRU layers for temporal feature extraction and the Inception-v3 block [277] for spatial feature extraction. Experimental results indicate that Bi-GRU-I is highly effective for HAR systems that rely on inertial sensors, showcasing its utility in capturing complex human activities.

Mohsen [278] explored an optimized GRU-based architecture for classifying daily human activities, focusing on hyperparameter tuning to enhance test accuracy. They found that the performance of the GRU model is significantly influenced by various hyper-parameters, including batch size, optimizer type, dropout rate, number of training epochs, learning rate, and the configuration of activation and loss functions, as well as neuron counts in specific layers.

A two-stacked GRU model with attention mechanisms was utilized by Abraham and James [279], aiming to create lightweight, efficient models suitable for devices with limited computational resources. This

model focuses on streamlined neural network topologies characterized by fewer layers and a reduced neuron count, offering an effective solution without compromising performance.

In another work, Mim et al. [280] proposed a method called GRU-INC, which integrates Inception-Attention architecture with GRU to effectively leverage both temporal and spatial aspects of time-series data. They employed the Inception module [277] to streamline parameters and enhance efficiency. Furthermore, they augmented this with a CBAM (Convolutional Block Attention Module) block [221] to enhance the features extracted from temporal and spatial data.

Fährmann et al. [281] introduce a technique for detecting occupancy and multi-occupancy within a smart environment using the Bi-GRU model. This occupancy detector, based on Bidirectional Gated Recurrent Units (Bi-GRU), demonstrates proficiency in capturing intricate patterns in sensory data by considering both past and future contexts. The research includes a thorough comparison of Bi-GRU models against other sophisticated RNN algorithms, highlighting the superior performance of Bi-GRU in discerning complex patterns.

These developments collectively illustrate the continuous evolution of GRU-based models, designed to address the varied and intricate requirements of Human Activity Recognition (HAR) systems within smart environments. Table 4 summarizes the proposed GRU-based HAR systems in sensor-based settings.

Table 4: A summary of GRU-based HAR systems

Reference	Year	Proposed GRU method	Dataset(s)	Performance (Accuracy)	Issues handled/achievements
Haque et al. [270]	2019	(G1 + AC + SA) − (G2 + AC + SA) − FC1 − FC2 − FC3	UCI-HAR	94.16%	Enhancement performance using hierarchical attention mechanism
Jian et al. [272]	2021	FC-G-AR-MLP-S	Self-collected data with 4 activities	F-score = 91.08%	Improving performance using attention-augmented
Mekruksavanich and Jitpattanakul [273]	2022	BG-AT-FC-S	WISDM UCI-HAR PAMAP2 OPPORTUNITY	92.42% 99.00% 95.82% 88.24%	Extracted critical temporal characteristics from complex human activities using attention mechanism
Pane et al. [274]	2022	GRU with channel attention and temporal attention	Ulster	90.2%	Handling data variability of different individual volunteers using dual attention mechanisms

(Continued)

Table 4 (continued)

Reference	Year	Proposed GRU method	Dataset(s)	Performance (Accuracy)	Issues handled/achievements
Sun et al. [275]	2022	G2 + AC + SA)-CE-FC-S	WISDM Self-collected data (DAAD) with 8 activities	96.80% 96.90%	A spatiotemporal multi-feature extraction method for abnormal data
Tong et al. [276]	2022	BG1-BG2-IB1-IB2-IB3-GAP-FC	UCI-HAR WISDM Self-collected data	95.42% 98.25% 99.60%	Effective approach for the inertial sensor-based HAR
Mohsen [278]	2022	I-G1-G2-S	WISDM	97.08%	Effect of hyper-parameter tuning on performance
Abraham and James [279]	2023	I-G1-G2-AT-D-FC1-FC2	UCI-HAR	99.00%	Design lightweight model to be implemented in resource-constraint gadgets
Mim et al. [280]	2023	G1-AT-G2-IB1-IB2-MP-D-CBAM-IB3-GAP-D-S	UCI-HAR OPPORTUNITYPAMAP2 WISDM	F-score = 96.26% F-score = 90.05% F-score = 90.30% F-score = 99.12%	Achieve good performance for complex activities recognition
Fährmann et al. [281]	2024	I-BG1-D1-BG2-D2-BG3-D3-BG4-FC1-FC2-S	UCI Kasteren CASAS: HH112 CASAS: HH111	98.55% 85.23% 87.29% 90.82%	Use of occupancy detector in multi-resident setting

Note: I-Input Layer, G-GRU Layer, BG-Bidirectional GRU Layer, S-SoftMax Layer, AR-Attention Reinforcement, MLP-Multilayer Perceptron, C-Convolution Layer, GAP-Global Average Polling, D-Dropout Layer, FC-Fully Connected (Dense) Layer, AC-Attention with Context, SA-Simplified Attention, AT-Attention Layer, CE-Concatenate Layer, PC-Primary Caps, DC-Digit Caps, IB-Inception Block, MP-Max Pooling, CBAM-Convolutional Block Attention Module.

5.6 *Transformer*

Transformer architecture [282] was first presented for machine translation in 2017, and it has since grown to become a key paradigm in deep learning. The transformer is an encoder-decoder design that uses self-attention. Each of the encoder's stack of identical layers is made up of two sublayers. The initial layer is a multi-head self-attention mechanism, while the second layer is a position-wise fully connected feed-forward network. Additionally, the inputs and outputs of the multi-headed self-attention module are connected via a normalizing layer [283] and residual connections [168]. The output sequence is then created by a decoder using the representation that the encoder generated. In order to manage multi-head attention over the encoder stack's output, the decoder appends a third sub-layer to each encoder layer in addition to the first two. Fig. 12 shows the overall Transformer design with multi-head attention layer [284].

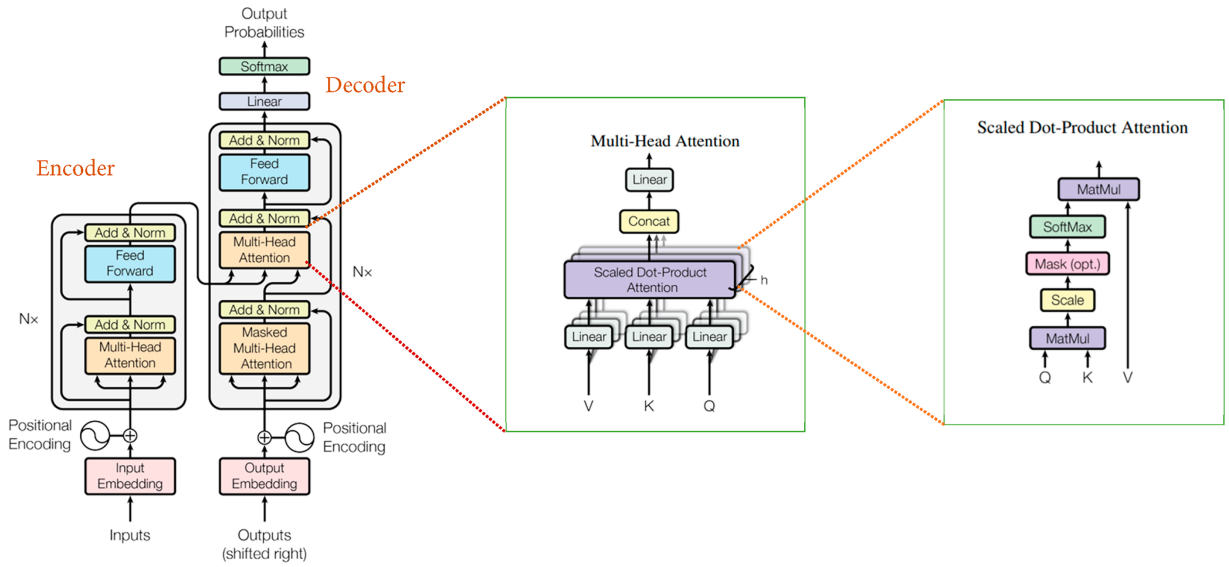

Figure 12: The structure of the Transformer model [282]

When compared to recurrent layers, the transformer architecture's self-attention mechanism provides a number of benefits, including improved model parallelism and less inductive bias than convolution networks. This approach enables the model to create pairwise correlations, represent long-range connections between input data pieces, and flexibly focus on different parts of the input sequence [285,286]. Additionally, Transformers can manage a large number of parameters without suffering from the vanishing or exploding gradient issues common in traditional models, making it possible to enhance performance by scaling up the model size [287].

Multi-Head Attention: To improve the Transformer model's capacity to identify dependencies among sequence elements, a multi-headed self-attention mechanism is used. The fundamental idea behind the attention mechanism is that each token in the sequence has the ability to aggregate data from other tokens, which improves the model's comprehension of contextual relationships. This is accomplished by creating an attention function by mapping a query, a collection of key-value pairs, and an output (all of which are represented as vectors). The outputs are calculated as a weighted sum of the data, with the compatibility function between the query and its matching key determining the weights. Several different scaled dot-product attentions (self-attention) can be combined to create multi-head attention. It can efficiently compute

the three vectors (Q, K, and V) in parallel to combine and calculate the final result. The formula can be seen in Eq. (14).

$$\begin{cases} MultiHead\ (Q, K, V) = Concat\ (head_1, \ldots, head_2)\ W^O \\ where\ head_i = Attention\ (QW_i^Q, KW_i^K, VW_i^V) \end{cases} \tag{14}$$

where the projections are parameter matrices $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{model} \times d_V}$, $and\ W^O \in \mathbb{R}^{hd_v \times d_{model}}$.

Scaled dot-product attention (self-attention), the transformer's primary component, makes use of the weight of each piece of data in the input vector, which is represented by

$$Attention\ (Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{15}$$

In scaled dot-product attention, the first step is to turn the input data into an embedding vector. From the embedding vectors, the three vectors of query vector (Q), key vector (K), and value vector (V) are then extracted. A score is then calculated for each vector, and the score is equal to $Q \cdot K$. To ensure gradient stability, score normalization is applied by dividing by $\sqrt{d_k}$. After that, the softmax activation function is used to process the score, and the weighted score v for each input vector is made by taking the softmax dot product value v. The final result is generated after summing [288].

Position-Wise Feed-Forward Networks: Along with attention sub-layers, each encoder and decoder layer have a fully connected feed-forward network. The feed-forward network is applied in the same way and independently to every point. Two linear transformations joined by a ReLU activation comprise this.

$$FFN\ (x) = \max\ (0, xW_1 + b_1)\ W_2 + b_2 \tag{16}$$

Positional Encoding: The Transformer model needs a method to record the relative or absolute locations of tokens inside a sequence in order to make use of the sequence's order, as it does not rely on recurrence or convolution. Positional encoding is implemented at the input level of the encoder and decoder stacks in order to remedy this. The input embeddings are supplemented with these positional encodings as they have the same dimensionality, d_{model}. Because of this combination, the model can better comprehend the sequential nature of the data by including positional information. In transformer design, positional encodings were accomplished by employing sine and cosine functions of different frequencies:

$$\begin{cases} PE_{(pos,2i)} = \sin(pos/10{,}000^{2i/d_{model}}) \\ PE_{(pos,2i+1)} = \cos(pos/10{,}000^{2i/d_{model}}) \end{cases} \tag{17}$$

where pos is the position and i is the dimension. There is a sinusoidal connection between each positional encoding dimension. The wavelengths in geometric development vary from 2π to $10{,}000 \cdot 2\pi$. Since PE_{pos+k} can be written as a linear function of PE_{pos} for any fixed offset k, this function was chosen because it would make it easier for the model to learn how to attend to relative locations.

Recent research has emphasized the use of Transformer models in a number of experiments to improve human activity recognition systems in sensor-driven environments [289–292]. A Two-stream Transformer Network (TTN) based on self-attention was proposed by Xiao et al. [293] to capture the spatial–temporal dependency of multimodal sensory signals. In order to extract sensor-over-time and time-over-sensor information from temporal and spatial channels, they implemented the two-stream architecture. These two streams of characteristics are complementary because the time-over-sensor features can convey

supplementary information that cannot be immediately obtained from sensor data. An attention block is added to the spatial channel in order to highlight the contributions of various axis-readings. To include the location information into the temporal sequence, positional encoding is utilized.

Authors of [288] introduced a new transformer model combined with a bidirectional GRU mechanism, termed TRANS-BiGRU, for recognizing various activity types in multi-resident environments. The TRANS-BiGRU model has demonstrated significant performance enhancements over existing deep learning models in HAR systems.

A HAR system that combines a transformer model with an activity-based sliding windows method was suggested by Huang and Zhang [294]. By successfully integrating various sensor data aspects, this system greatly improves the accuracy of activity identification. Both activity-driven sliding windows, which identify the window during which user action takes place and segment the series appropriately, and the transformer, which can process data in parallel and learn complex patterns, have shown themselves to be effective tools for sensor data analysis.

Saidani et al. [295] suggested a HAR system that makes use of a Transformer model and data augmentation methods. To increase training data and enhance model generalization, methods including temporal shifting, domain adaptation, and Gaussian noise were employed. The system's discriminative capacity is improved by the suggested composite features of this model, which can extract information from both high-level and low-level sensor data. The model's ability to extract long-range relationships with fewer layers has increased accuracy, allowing it to identify intricate patterns in human activities.

A transformer-based deep reverse attention method that considers the side residual outputs was proposed by Pramanik et al. [296]. They direct the side residual outputs using the feature outputs from the layer that follows, which are decoded using an LSTM-Deconvolution-based decoder. To ensure that the model is not biased towards any specific data, they also take a number of steps to dynamically calibrate the learning and minimize overfitting. They enrich the original dataset with noisy samples, adding Gaussian noise to the original time series data to provide variety in the learning and decrease overfitting. Additionally, they add a gaussian dropout at the end of each block with a dropout probability of 0.1 to lessen the chance of overfitting. The first reverse attention block receives the output from block 1 as the primary branch and the features from block 2 as the support branch. The identical process is used with block-2 and block-3 for the second reverse attention block. Through the employment of transformers, each of these reverse attention blocks seeks to connect the lower-level spatial information with the higher-level semantic information in order to increase the relevance of the features.

Suh et al. [297] present a unique Transformer-based Adversarial learning framework that takes into consideration individual sensor orientations as well as spatial and temporal data for the purpose of recognizing human activity utilizing wearable sensors via Self-KnowledgE Distillation (TASKED). The primary goal of the model is to jointly optimize adversarial learning between the topic discriminator and transformer-based feature extractor in order to learn cross-domain feature representations. The transformer architecture is designed to capture both temporal and spatial representations from time-series sensor data. To generalize feature distributions across different subject domains, an adversarial learning scheme was employed alongside the multi-kernel maximum mean discrepancy (MK-MMD) regularization method. Moreover, the teacher-free self-knowledge distillation approach was used to enhance training process stability and avoid a bias towards feature generalization that is regularized by adversarial learning and MMD loss.

The binarization of Deep Vision Transformer (DeepViT) for effective human activity recognition was studied by Luo et al. [298]. They begin by converting sensor inputs to spectra using the Continuous

Wavelet Transform (CWT), then stacking the spectra for multiple sensor fusion. They employ a transformer called DeepViT [299] to categorize the wavelet spectra of various activities in order to classify activities. They then binarize DeepViT using three different methods, including simple binarization, Libra Parameter Binarization [300], and Uniq binarization [301], and compare the results.

Muniasamy and Asiri [302] offer a unique human activity recognition model that leverages a Residual Transformer Network (RTN) in conjunction with a CNN. First, the incoming signal undergoes preprocessing to create a spectrogram. Both the frequency and temporal components of sensor signals are present in spectrogram images. CNN provides a solid basis for spatial analysis by effectively extracting spatial elements from spectrograms. The CNN model used in this work is the EffcientNet-B3 [303] model. Simultaneously, the transformer is capable of simulating long-range temporal dependencies that are essential for identifying complicated human activities since it has residual connections with various input scales and sparse attention. In order to overcome issues like disappearing gradients and maintain steady information flow during training, residual connections are incorporated. By combining the data from the CNN and Transformers, the adaptive attention fusion process maximizes their combined impact. This integration enhances the robustness of the model to differentiate between different human activities by optimizing feature extraction.

All of these advancements show how Transformer models, which are intended to meet the complex and varied needs of Human Activity Recognition (HAR) systems in smart settings, are always evolving. The transformer-based HAR systems in sensor-based environments are compiled in Table 5.

Table 5: A summary of transformer-based HAR systems

Reference	Year	Proposed transformer method	Dataset(s)	Performance (Accuracy)	Issues handled/achievements
Xiao et al. [293]	2022	I-stream of 1 [AT-C-TR without PE-GT], stream of 2 [C-TR with PE-GT]-CE-S	PAMAP2 Opportunity USC-HAD Skoda	F-score = 98% F-score = 69% F-score = 57% F-score = 95%	The Transformer processes two parallel streams to capture both temporal and spatial dependencies simultaneously.
Chen et al. [288]	2022	Transformer with Bidirectional GRU	CASAS: Kyoto Multi Resident ARAS: houseA ARAS: houseB	92.86% 89.48% 90.59%	Enhance performance in multi-resident setting.
Saidani et al. [295]	2023	Transformer model with data augmentation techniques	WISDM PAMAP2 UCI-HAR	97.3% 98.2% 98.6%	Extracting a diverse set of features to capture various patterns and information from sensor data, thereby enhancing the model's robustness.

(Continued)

Table 5 (continued)

Reference	Year	Proposed transformer method	Dataset(s)	Performance (Accuracy)	Issues handled/achievements
Huang and Zhang [294]	2023	Transformer model with activity-based sliding windows segmentation	CASAS: Aruba	F-score = 96.4%	The use of transformers to capture complex patterns in sensor sequences and segment the sequences can lead to improved performance.
Pramanik et al. [296]	2023	Block 1[C-BN-C-BN-GD]-RAB1-Block 2[C-BN-C-BN-GD]-RAB2-Block 3[C-BN-C-BN-GD]-CE-FC	MHEALTH USC-HAD WHARF UTD-MHAD1 UTD-MHAD2	100% 95.14% 92.92% 64.46% 84.21%	Using reverse attention block to amplify the relevance of the features by establishing a connection between the higher-level sematic information and the lower-level spatial information. Add Gaussian noise and introduce an additional gaussian dropout to reduce the risk of overfitting.
Suh et al. [297]	2023	Transformer model with adversarial learning	Opportunity PAMAP2 MHEALTH RealDISP	75.83% 83.04% 95.00% 86.49%	Learning cross-domain feature representations by jointly optimizing adversarial learning between a transformer-based feature extractor and a subject discriminator.
Luo et al. [298]	2025	Deeper Vision Transformer (DeepViT) with Continuous Wavelet Transform (CWT) and binarization techniques	MHEALTH UCI-HAR USC-HAD	98.83% 99.02% 91.4%	Investigating the impact of binarizing various components and applying latent binarization in Deep Vision Transformer on the performance of sensor-based HAR systems.
Muniasamy and Asiri [302]	2025	Combine of Residual Transformer Network (RTN) and EfcientNet-B3	HARTH KU-HAR HuGaDB	94.82% 98.94% 96.33%	The proposed method can distinguish between activities with similar overall patterns by capturing subtle differences.

Note: I-Input Layer, AT-Attention Layer, C-Convolution Layer, TR-Transformer, GT-Global Attention Layer, PE-Positional Encoding, S-SoftMax Layer, BN-Batch Normalization, GD-Gaussian Dropout, RAB-Reverse Attention Block, FC-Fully Connected (Dense) Layer, CE-Concatenate Layer.

5.7 Deep Belief Network (DBN)

The Deep Belief Network (DBN) stands as a form of deep generative model utilized in unsupervised learning endeavors, particularly geared towards uncovering patterns within extensive datasets. Comprising multiple layers of hidden units, DBNs excel at discerning intricate patterns and extracting features from data. In comparison to discriminative models, DBN displays heightened resistance to overfitting problems, rendering them apt for extracting features from unlabeled data [15]. A foundational component of DBNs is the stack of Restricted Boltzmann Machines (RBMs), which operate in an unsupervised learning paradigm. Each RBM within a DBN encompasses a visible layer representing observable data features and a hidden layer capturing latent representations [304]. RBMs undergo layer-by-layer training, initially trained individually and subsequently fine-tuned collectively as part of the entire DBN. RBMs are symmetric bipartite graphs trained through numerous forward and backward passes between the visible and hidden layers.

During the forward pass, activations depict the likelihood of an output given a weighted input, while in the backward pass, the outcome is an estimation of the likelihood of inputs given the weighted activations. Through iterative RBM training within a DBN, joint probability distributions of activations and inputs are attained [261,305]. The schematic representation of the Deep Belief Network (DBN) structure is illustrated in Fig. 13.

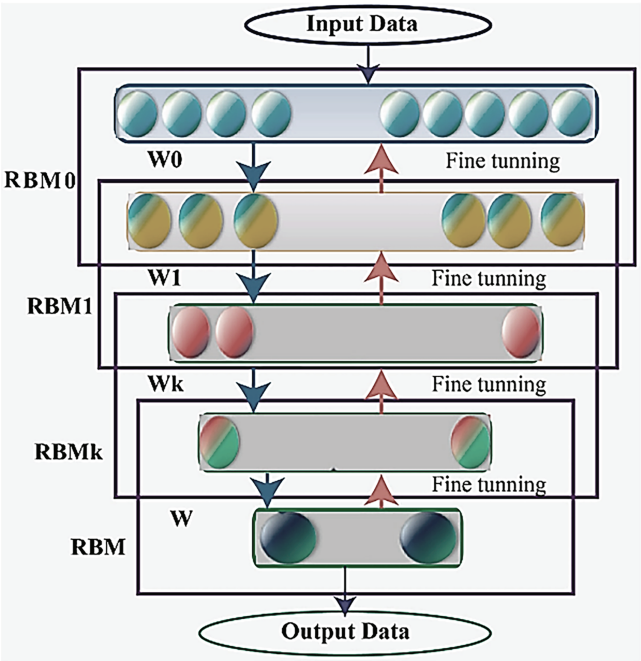

Figure 13: The structure of a DBN model [304]

Deep Belief Networks (DBNs) are formidable models renowned for their prowess in unsupervised learning endeavors and have showcased remarkable effectiveness across a spectrum of applications. Particularly within HAR systems, DBNs have been instrumental in recognizing activities from raw sensor data. This involves the independent pretraining of each layer in an unsupervised fashion, followed by comprehensive network refinement via backpropagation.

In a study conducted by the authors of [306], a DBN model was employed to classify three activity categories within a multi-resident setting. This novel approach was compared with conventional machine learning techniques including support vector machine (SVM) [307] and back-propagation algorithm

(BPA) [308], utilizing the publicly available CASAS database. The findings revealed superior accuracy of the DBN model in contrast to SVM and BPA, thereby reinforcing the efficacy of DBN in activity recognition endeavors.

In another study, Rossi et al. [80] proposed a hierarchical Deep Belief Network (DBN) architecture tailored for group activity recognition. The first meta-layer of the model discerns actions executed by individual users, while the second meta-layer seeks shared temporal and spatial dynamics indicative of group activity formation. Evaluation of the suggested method within a controlled laboratory environment, where participants utilized a mobile phone application to categorize daily activities, demonstrated enhanced performance in activity detection. Ambient sensor data, complementing mobile phone data, further bolstered activity recognition accuracy, outperforming SVM-based models as evidenced by empirical results on the experimental dataset.

Huang et al. [309] harnessed Deep Belief Networks (DBNs) to predict discrete patterns in the daily routines of elderly individuals via knowledge-based approaches. Unsupervised daily activities, ranging from sleeping to meditation, were observed and tracked using motion sensors, actuators, and surveillance systems to mitigate potential safety risks. Leveraging gathered data, DBNs were employed to identify actions associated with various health and self-care concerns, exhibiting superior performance in activity recognition compared to support vector machines (SVMs) according to experimental results.

In the last study, Kumar and Murugan [310] introduced a novel approach, named the CSODBN-HAR model, which integrates Deep Belief Networks (DBN) and the Cuckoo Search algorithm [311] for human activity recognition. The Cuckoo algorithm, a heuristic optimization technique renowned for its versatility across diverse applications, employs mechanisms derived from local optimization. In the presented CSODBN-HAR model, a standard scalar approach is employed as a pre-processing step. Subsequently, the DBN model is employed for the identification and categorization of human activities. Notably, the Cuckoo Search algorithm is leveraged to fine-tune the hyperparameters of the DBN model, leading to a significant enhancement in recognition rates and underscoring the novelty of the approach. Extensive comparative analyses demonstrate the superiority of the CSODBN-HAR model over state-of-the-art human activity recognition (HAR) systems.

Nevertheless, despite their demonstrated efficacy, deep belief networks (DBNs) are less prevalent in human activity recognition (HAR) tasks involving sensor data compared to architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). This disparity arises from DBNs' disparate settings for each input feature, rendering them inherently challenged in learning discriminative patterns from time series data. In contrast, CNNs leverage pooling layers to introduce partial invariance to local translations, while RNNs maintain weight consistency across sequential time steps, rendering them better suited for time series classification tasks [261].

5.8 Hybrid Models

Hybrid deep learning architectures, which combine elements from diverse deep learning architectures, exhibit promising efficacy in enhancing performance for human activity recognition (HAR) tasks. By amalgamating the strengths inherent in distinct architectures, such hybrid models proficiently encapsulate both spatial and temporal dependencies within the data domain. A plethora of investigations have embraced hybrid deep learning models to recognize human activities, including CNN + LSTM [312–315], CNN + Bi-LSTM [316–319], CNN + GRU [320–322], and CNN + Bi-GRU [323], among others.

In a study conducted by Xia et al. [324], a deep neural network amalgamating Long Short-Term Memory (LSTM) units with convolutional layers was introduced to recognize human activities. The model

autonomously extracts activity features and performs classification with minimal parameters. The authors meticulously explored the impact of hyperparameters, including batch size, optimizer type, and filter count, on model performance, culminating in the selection of optimal hyperparameters for the final model design.

Dua et al. [325] introduced a multi-input Convolutional Neural Network-Gated Recurrent Unit (CNN-GRU) model for HAR, leveraging the prowess of GRU in classifying time series data and the feature extraction proficiency of CNN. The suggested temporally and spatially deep multi-input CNN-GRU architecture outperformed existing DL models in HAR tasks. The model's multi-input framework adeptly captures both shallow and deep characteristics, facilitating more precise activity prediction by effectively managing long-term dependencies in sequence data with GRU layers while concurrently capturing local features with convolutional layers.

Kim et al. [326] advocated a stacking ensemble deep learning model integrating one GRU head, two 1D-CNN heads, and a rudimentary meta-learner. Additionally, they introduced a wearable Inertial Measurement Unit (IMU) system to assess patients using the Berg Balance Scale (BBS). The BBS is an extremely trustworthy balancing test for stroke patients and the elderly [327].

In another study, a hybrid model amalgamating a convolutional neural network (CNN) with a long short-term memory (LSTM) network was proposed by Natani et al. [166], enabling the exploitation of spatial information extracted by CNN in tandem with temporal dependencies captured by LSTM. Their experimentation on the ARAS dataset evinced that the CNN + LSTM model outperformed other deep learning and machine learning models, including traditional sequential models like Hidden Markov Model (HMM) [79] and Conditional Random Fields (CRF) [328].

Challa et al. [329] proposed a multibranch CNN+ Bi-LSTM architecture directly operating on raw data acquired from wearable sensors with minimal or no preprocessing. This model adeptly captures both short-term and long-term dependencies in sequential data by harnessing the benefits of both CNN and Bi-LSTM. To enhance feature extraction, the proposed architecture employs multiple convolutional filter sizes to capture diverse local dependencies, thus proficiently discerning both simple and complex activities such as walking, sitting, jogging, Nordic walking, vacuuming, and ironing.

Khatun et al. [330] conducted an evaluation of a deep convolutional neural network (CNN) coupled with long short-term memory (LSTM) and Self-Attention mechanisms for classifying daily activities leveraging wearable sensor data. Their model integrated inputs from gyroscopes, raw accelerometers, and linear acceleration data from smartphones. The synergy of these components empowered the network to discern intricate patterns in human activities.

In a bid to enhance the efficacy of Inertial Measurement Unit (IMU)-based human activity recognition systems, Kim et al. [331] proposed a novel deep learning architecture alongside a data augmentation strategy. This augmentation approach aimed to rectify data imbalances by equilibrating the dataset across activity classes, thereby augmenting the total dataset volume. The foundation of the data augmentation algorithm utilized in this study lies in the oversampling technique. Several oversampling algorithms, including Synthetic Minority Oversampling Technique (SMOTE) [332], Borderline SMOTE [333], and Random Over Sampling Examples (ROSE) algorithms [334], were explored to identify the most suitable augmentation approach. It was observed that the SMOTE data augmentation strategy yielded optimal performance, resulting in the classification model achieving its peak accuracy.

Perumal et al. [335] developed an Internet of Things (IoT) centric multi-activity recognition system within a smart home environment. This innovative system integrated a gyroscope and tri-axial accelerometer into a microprocessor chip, thus constituting a scalable Inertial Measurement Unit (IMU) device. To ensure real-time data acquisition and optimal sensor placement, IMU devices were strategically affixed to subject

locations. Throughout the experimental phase, six distinct multi-activities were collected from the IMU sensors and transmitted to the IoT-Edge Server for processing. To enhance classification accuracy, sensor data underwent fusion, and artifacts were mitigated using the Kalman filtering technique. Subsequently, the preprocessed sensor-fused data were subjected to classification utilizing both convolutional neural network (CNN) and long short-term memory (LSTM) models, synergistically employed to discern activity patterns with heightened precision.

Agac and Durmaz Incel [336] employed the channel and spatial attention modules derived from the Convolutional Block Attention Module (CBAM) [221] to advance sensor-based human activity recognition. Their investigation meticulously scrutinized three distinct attention mechanisms: channel attention, spatial attention, and channel-spatial attention. These attention modules were strategically applied to the convolutional layers of a hybrid sensor-based human activity recognition (HAR) model, specifically a combination of convolutional neural network (CNN) and long short-term memory (LSTM) architectures, across different depths. This comprehensive approach aimed to enhance model interpretability and performance by effectively highlighting relevant features within the input data at various spatial and channel levels.

In other study, a novel fuzzy convolutional attention-based GRU network model for human activity recognition from sequential data was proposed by Khodabandelou et al. [337]. The rationale behind this approach is rooted in the understanding that a model trained to track temporal activity evolution is inherently adept at capturing activity patterns, thereby enhancing prediction accuracy and adaptability to diverse environments. Consequently, they advocate for the application of a temporal and sequential modeling approach to analyze activity data. The model introduces an innovative adaptive kernel mechanism to enhance the representation of activity sequences based on data characteristics derived from fuzzy rules applied to input sequences. Subsequently, local subsequences are extracted from the entire sequence using a convolutional neural network (CNN) to identify patterns within the convolution window. Finally, an attention-based GRU module is incorporated into the model to extract significant portions of the time-series sequences, further augmenting its ability to discern nuanced activity patterns.

In order to recognize human behavior in edge computing contexts, Huang et al. [338] proposed the O-Inception Element-Wise-Attention GRU network (OI-EleAttG-GRU) architecture. This architecture operates in two main phases: first, to alleviate local computing demands, human activity recognition tasks are dynamically assigned to suitable servers using a Dynamic Scheduling Technique (DTS). Subsequently, the OI-EleAttG-GRU deep learning architecture, installed on the server, executes the human activity recognition tasks. It identifies the corresponding actions for each task and returns the recognition results to the local device. The architecture integrates an element-wise attention gate, merging a GRU block with a novel O-Inception convolutional structure. This convolutional structure achieves a balance between identification accuracy and training speed, while the GRU block with EleAttG enhances prediction accuracy through an efficient attention mechanism. By leveraging the strengths of both components, the proposed model facilitates rapid and accurate identification of human activities in edge computing scenarios.

Choudhury and Soni [339] introduced a highly efficient adaptive batch size-based CNN-LSTM model tailored for recognizing a multitude of human activities in uncontrolled environments. This model strategically utilizes adjustable batch sizes, spanning from 128 to 1024, during iterative training and validation stages. By dynamically adjusting batch sizes, the model effectively addresses class imbalances and handles non-normalized data, thus enhancing its robustness and performance in challenging real-world settings.

Ali Imran et al. [340] introduced HARDenseRNN, a novel model designed for human activity recognition, seamlessly integrating components from convolutional neural networks (CNN) and recurrent neural networks (RNN). This model is composed of two CNN modules featuring multiple types of kernels and residual connections, followed by a 128-unit bidirectional gated recurrent unit (Bi-GRU). Leveraging two

InceptionResNet modules [341], which combine the structural elements of InceptionNet modules [203] and ResNet modules [215], the architecture exhibits superior performance compared to existing models while simultaneously inheriting the strengths of both approaches.

To address the challenges inherent in multi-resident activity recognition within ambient sensor-equipped households, Ramanujam and Perumal [342] propose the Multi-label Multi-output Hybrid Sequential Model (MLMO-HSM). This model aims to mitigate issues related to both the generation of deep learning models, particularly the training time, and the multi-output classification process by leveraging a feature engineering approach. Through this strategy, the suggested model has demonstrated improved performance, underscoring the effectiveness of feature engineering in enhancing its capabilities.

Zhao et al. [343] implemented an LSTM-1DCNN algorithm for human activity recognition (HAR) utilizing data from a triaxial accelerometer. The algorithm employs two branches to capture the temporal and spatial characteristics of the measured data. Through training and evaluation on both publicly available datasets and self-built datasets, the method's performance is compared against five other algorithms: Decision Tree, Random Forest, Support Vector Machine, 1DCNN, and LSTM. Evaluation measures indicate that the recognition performance achieved by the LSTM-1DCNN algorithm surpasses that of the other five algorithms, highlighting its efficacy in HAR tasks.

In another study, Praba and Suganthi [344] proposed the Human Activity Recognition Neural Network (HARNet) architecture, which integrates CNN and LSTM networks with SVM transfer learning for automated human activity recognition. The selection of hybrid CNN and LSTM approaches from a multitude of Deep Learning (DL) algorithms is motivated by their capability to extract independent and discriminative features. These features, in turn, contribute to enhancing the SVM classifier's accuracy in activity classification.

Lalwani and Ramasamy [345] introduce a powerful classification approach that integrates a bidirectional gated recurrent unit (Bi-GRU), a convolutional neural network (CNN), and a bidirectional long short-term memory (Bi-LSTM). This methodology excels in extracting crucial insights from raw sensor data, effectively and efficiently. By synergizing CNN, Bi-GRU, and Bi-LSTM components, the model demonstrates prowess in discerning both short-term patterns and long-term relationships within sequential data. Furthermore, the incorporation of multiple filter sizes expedites the feature extraction process, enabling the model to capture a diverse array of temporal local relationships.

In order to teach the model sequential and spatial patterns, Thakur et al. [346] introduce a Hybrid deep learning model based on Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) techniques. Also, by using optimization approaches, the selection of characteristics has been improved. This study introduces a hybrid optimization strategy that combines the Grey Wolf Optimizer (GWO) [347] with the Whale Optimization Algorithm (WOA) [348]. The hybrid strategy improves the optimization process overall by integrating and leveraging the advantages of both approaches.

These diverse efforts demonstrate the rich landscape of hybrid models for human activity recognition (HAR) and their adaptability across different scenarios. By merging various deep learning techniques, these models harness the strengths of each architecture, effectively navigating the complexities of human activity data. Table 6 provides a concise summary of the proposed hybrid deep learning models used in HAR systems with sensor-based data, clearly illustrating their distinctive features and unique contributions to advancing the field. The table serves as a comprehensive resource, highlighting key innovations and providing comparative insights into the performance of different hybrid models.

Table 6: A summary of hybrid deep learning models in HAR systems

Reference	Year	Proposed method	Dataset(s)	Performance (Accuracy)	Issues handled/achievements
Xia et al. [324]	2020	LSTM + CNN (LS1-LS2-C1-MP-C2-GAP-BN-FC)	UCI-HAR WISDM Opportunity	95.78% 95.85% 92.63%	Explored the impact of some hyper-parameters on model performance
Dua et al. [325]	2021	Three heads of CNN + GRU (C1-C2-D-MP-F-G1-G2-D)-CE-FC-BN-S	UCI-HAR WISDM PAMAP2	96.20% 97.21% 95.27%	Perform well on datasets comprising various features and can recognize a varied range of activities
Kim et al. [326]	2021	Ensemble of head1(C-MP-F), head2(C-MP-F), head3(G-FC-F)-CE-D-FC1-FC2	Self-collected data	98.4%	Useful for elderly or brain-disease patients who are at risk of falling
Natani et al. [166]	2021	CNN1D + LSTM CNN2D + LSTM (C1-C2-FC1-LS-FC2)	ARAS: House A ARAS: House B	86.31% 83.28% 87.97% 88.26%	Handling the multi-resident issue
Challa et al. [329]	2022	Three parallel branches of CNN (C1-C2-D-MP-F)-CE-BG1-BG2-FC1-BN-FC2	UCI-HAR WISDM PAMAP2	96.05% 96.37% 94.29%	Able to identify complex activities
Khatun et al. [330]	2022	CNN + LSTM (C-LS1-AT-LS2-AT-LS3-FC1 -BN1-D1-FC2-BN2-D2-FC3)	Self-collected data with 4 activities MHEALTH UCI-HAR	99.93% 98.76% 93.11%	Enhance HAR performance in smartphone users' activities
Kim et al. [331]	2022	Ensemble of head1(C1-MP1-C2-MP2-C3-MP3-F), head2(C1-MP1-C2-MP2-C3-MP3-F), head3(BG1-BG2-BG3-BG4-FC-F)-CE-D-FC1-FC2	WISDM UCI-HAR	99.49% 99.48%	Using data augmentation algorithm to performance improvement

(Continued)

Table 6 (continued)

Reference	Year	Proposed method	Dataset(s)	Performance (Accuracy)	Issues handled/achievements
Perumal et al. [335]	2022	CNN + LSTM	Self-recorded data with 6 activities from 10 users	98.2%	Handling the multi-resident issue and using the Kalman filtering process to improve the performance
Khodabandelou et al. [337]	2023	AF-C1-MP1-C2-MP2-C3-MP3-AT-G-FC	Self-collected data with 5 activities	99.78%	Handling dynamic activities data
Huang et al. [338]	2023	C1-P1-C2-P2-C3-P3-C4-P4-IB1-IB2-F-G-S	WISDM	97.52%	Enabling accurate and fast human activity recognition
Agac et al. [336]	2023	C1-CBAM1-C2-CBAM2-C3-CBAM3-C4-CBAM4-LS1-LS2-FC	PAMAP2 Opportunity	F-score = 96% F-score = 77%	Optimizing wearables resource consumption by analyzing the attention modules parameters
Choudhury et al. [339]	2023	CNN + LSTM (I-C1-C2-D1-MP-F-LS-D2-FC1-FC2)	Self-recorded data with 6 activities MHEALTH	98% 99.5%	Handling the class imbalance problem
Ali Imran et al. [340]	2023	IRN1-D1-BN1-IRN2-D2-BN2-BLS-BN3-F-D3-FC-D4-S	WISDM	98.81%	Evaluating the impact of segmentation window size and step size
Ramanujam and Perumal [342]	2023	Ensemble of Block1(LS-D), and Block2(C1-C2-C3-C4-BN-D)-CE-FC1-FC2-two parallel S	ARAS: House A ARAS: House B	Resedent1: 86.59% Resedent2: 93.06% Resedent1: 94.61% Resedent2: 95.29%	Handling the multi-resident issue, and addressing the training time challenge
Zhao et al. [343]	2023	Ensemble of (LS1-LS2), and (C1-BN1-MP1-C2-BN2-MP2-C3-BN3-MP3-C4-BN4-MP4-C5-BN5-MP5-GAP)-CE-FC1-FC2-FC3	UCI WISDM MotionSense PAMAP2 MHEALTH Self-recorded data with 7 activities	96.31% 98.49% 97.57% 91.61% 99.11% 98.16%	Use of various datasets and recording data for experiments and analysis

(Continued)

Table 6 (continued)

Reference	Year	Proposed method	Dataset(s)	Performance (Accuracy)	Issues handled/achievements
Praba and Suganthi [344]	2024	CNN + LSTM (C1-BN1-C2-BN2-LS-FC1-FC2)	MHEALTH	99.8%	Enhance performance using transfer learning technique
Lalwani and Ramasamy [345]	2024	C1-MP1-C2-MP2-D1-BLS-MP3-BG-D2-FC	WISDM	99.32%	Enhance HAR performance using hybrid models
Thakur et al. [346]	2025	CNN + LSTM with hybrid GWO–WOA optimization	Collected data by [349]	95%	Using hybrid GWO –WOA optimization to select appropriate features that have the potential to improve model efficacy.

Note: I-Input Layer, C-Convolution Layer, LS-LSTM Layer, BLS-Bidirectional LSTM Layer, G-GRU Layer, BG-Bidirectional GRU Layer, P-Pooling, G-GRU Layer, S-SoftMax Layer, MP-Max-Pooling, BN-Batch Normalization, D-Dropout Layer, FC-Fully Connected (Dense) Layer, AT-Attention Layer, F-Flattened Layer, CE-Concatenate Layer, IB-Inception Block, GAP-Global Average Polling, AF-Adaptive Fuzzy Kernel, CBAM-convolutional block attention module, IRN-InceptionResNet modules.

6 Federated Learning in HAR

Federated Learning (FL) has emerged as a promising approach for Human Activity Recognition (HAR), enabling the joint training of machine learning models on distributed devices while ensuring data privacy. By using the processing power and locally stored data of potentially millions of smart devices for parallel training, FL increases training efficiency in addition to improving privacy. Additionally, it provides substantial communication advantages by lowering the amount of data sent over the network, which eases congestion and lowers communication expenses. This section starts by describing the basic ideas of Federated Learning (FL) and looking at its main design factors, architectural elements, and several frameworks that support FL implementations. After that, we give a thorough summary of current research that has used FL in HAR systems. The goal of this section is to provide a comprehensive grasp of the fundamental structural elements needed for FL to be successfully integrated into HAR systems.

6.1 Federated Learning (FL)

Federated Learning (FL), a distributed collaborative AI paradigm, eliminates the need for raw dataset sharing by coordinating several devices with a central server to facilitate model training [350]. Actually, FL is a safe machine learning method that ensures that private data stays local while executing training algorithms across several edge devices or servers [351]. Each client trains the model locally rather than exchanging data, and parameter updates rather than direct data transmission are used to communicate with the server. The primary function of the server is to aggregate these parameters in order to update the global model [352]. The basic framework of FL is introduced in Fig. 14.

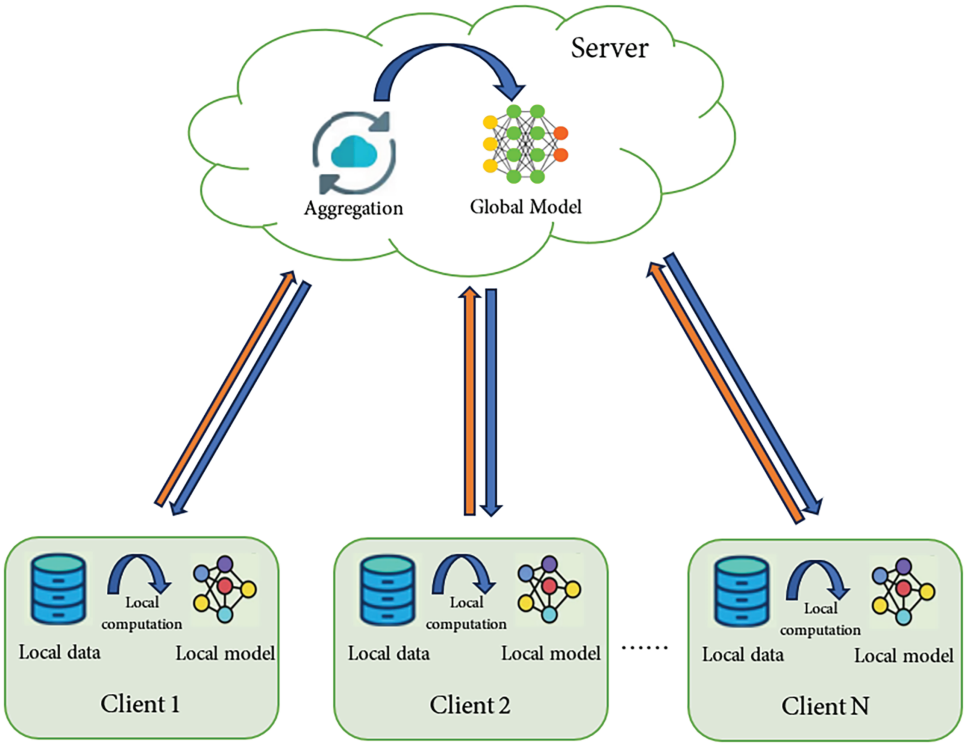

Figure 14: The basic framework of federated learning (FL)

Federated Learning (FL) successfully tackles issues with performance, network connectivity, and privacy. Since FL does not require raw data to be centralized for training, the risk of exposing private user information is greatly reduced, thereby ensuring a higher level of data privacy. Moreover, FL reduces communication delays typically caused by data offloading, as raw data do not need to be transmitted to the server [353]. FL can enhance both the learning accuracy and the convergence rate of the overall training process by leveraging diverse datasets and substantial computational resources across a network of devices. This may not be possible with centralized approaches that have limited computational power and insufficient data [354].

6.2 Type of FL Systems

FL techniques are typically separated into three categories based on the degree of overlap of data characteristics in the client dataset: Horizontal Federated Learning (HFL), Vertical Federated Learning (VFL), and Federated Transfer Learning (FTL). Each of these categories is appropriate for addressing a distinct set of real-world issues.

6.2.1 Horizontal Federated Learning (HFL)

In Horizontal Federated Learning (HFL), multiple clients hold distinct data samples while sharing the same feature space. The clients collaboratively train a global model without requiring any raw data to leave their local devices [355].

6.2.2 Vertical Federated Learning (HFL)

In Vertical Federated Learning (VFL), often called feature-based FL, clients possess different sets of features for the same or overlapping data samples, in contrast to HFL. For instance, two businesses can employ VFL to collaboratively build a machine learning model using data from the same group of individuals, each contributing distinct feature sets [356].

6.2.3 Federated Transfer Learning (FTL)

To build an effective shared machine learning model in HFL and VFL, all participants must either share the same feature space in HFL or the same sample space in VFL. However, in practical scenarios, the datasets held by different participants often vary significantly. Federated Transfer Learning (FTL) extends the capabilities of FL to scenarios where client datasets differ in both feature space and sample space [357]. Depending on how comparable the data or models are among participants, FTL models can learn from the source domain to the target domain. The idea entails knowledge transfer between domains, enabling learning from a variety of datasets. In some cases, it might be beneficial for consumers with less data to learn from those with more data, even if the characteristics or samples are different [352].

6.3 FL Architectures

The architecture of federated learning establishes the structure of models, data, and training procedures among clients and servers. The main FL architecture includes centralized FL, decentralized FL, and hierarchical FL.

6.3.1 Centralized FL (CFL)

One of the most popular FL architectures is centralized FL, sometimes referred to as server-oriented FL. To execute a FL model, a CFL system consists of a central server and a number of clients. During a single training cycle, each client uses their own datasets to train a network model in parallel. The trained parameters are then sent by each client to the central server, which combines them using a weighted averaging technique like Federated Averaging (FedAvg) [350]. After that, all clients receive the calculated global model for use in the following training cycle. Each client receives both their customized model and the same global model at the conclusion of the training procedure [353]. The data flow in the centralized architecture is frequently asymmetric, meaning that the server sends back training results after aggregating information from other clients. The server and local clients may communicate synchronously or asynchronously [358,359].

HAR can be implemented using the centralized FL (CFL) architecture, which comprises a central server and a diverse set of clients, many of which are smart devices such as smartphones or wearable sensors. These client devices separately update the common model throughout each training cycle by interpreting their localized data, such as gyroscope or accelerometer readings. After local training is complete, each client sends its model parameters to the central server for aggregation [360].

6.3.2 Hierarchical FL (HFL)

Despite its many advantages, centralized federated learning (FL) has several serious disadvantages, such as (i) a single point of failure that compromises robustness, (ii) latency because it uses a lot of communication resources, and (iii) convergence because of system and statistical heterogeneity [361]. Heterogeneous environments provide a serious difficulty since the diversity of data generated from many sources results in non-independent and non-identically distributed (non-IID) data, which means that the size and distribution of the data vary across each client. The FL model may become divergent as a result,

which would impair its functionality. As the number of clients rises, this issue gets even worse. Additionally, the FL server is usually situated in the cloud for model aggregation. Using the cloud server as a FL server has a number of challenges, such as time delays and communication expenses. To overcome these limitations, researchers have utilized both edge and cloud servers, giving rise to a new approach known as Hierarchical Federated Learning (HFL) [362]. Hierarchical FL does not rely on a single central server but rather introduces several layers of aggregation. The edge servers in the intermediate layer serve as aggregation locations for the local models. These edge servers execute "sub-global" model aggregations, which are then sent to cloud servers at the top tier for global model aggregation. Similar to CFL, hierarchical FL does not need the transfer of sensitive, raw data in order for clients to take part in the training of a shared global model. HFL is especially helpful when clients are spread out geographically or when communication expenses are considerable [363].

Hierarchical federated learning (HFL) is especially well-suited for HAR systems because of its layered structure, which offers a number of benefits such as scalability, improved data privacy, and efficient use of network resources. [364].

6.3.3 Decentralized FL (DFL)

Decentralized FL (DFL) is a network architecture that does not have a central server to manage the training process. DFL's main concept is to use peer-to-peer communication between individual clients in place of contact with the server. Within this framework, each client trains the model on its own local data and communicates updates to neighboring devices. DFL removes the central server, which lowers communication overhead and the possibility of a single point of failure, in contrast to centralized federated learning [365]. A connected graph is used to depict the communication topology, with nodes standing in for the clients and edges denoting a channel of communication between two clients. In contrast to the server-client architecture's star graph, the network graph is usually optimized to be sparse with a modest maximum degree, meaning that each node only needs to send and receive messages to a limited number of peers. Each client in a decentralized algorithm completes a round by updating locally and communicating with their graph neighbors [366].

6.4 FL Optimization Algorithm

During the FL training process, clients first optimize their local models using procedures like the gradient descent (GD) algorithm and then send the trained models to the central server for aggregation. The Federated Learning (FL) optimization algorithm is responsible for updating local models on client devices and coordinating their aggregation on the central server to improve the global model. The FL optimization algorithm method selection has a major impact on robustness, fairness, and convergence, particularly in heterogeneous situations where clients have varying data distributions [352]. Some of the common optimization methods are presented below.

***Federated Averaging (FedAvg)*:** FedAvg [350] is among the aggregation methods in FL that are used most frequently. In this method, Local models are trained on client devices using their own data, and the updates are sent to a central server, which averages them to update the global model. By extending local training times to shorten the communication round, the FedAvg algorithm can lower the communication cost. Nevertheless, extending training times is no longer able to resolve the communication cost issue after the model converges to a certain point [352].

***Federated Proximal Gradient Descent (FedProx)*:** FedProx [367] addresses the statistical heterogeneity present in federated networks with a distributed optimization methodology. In FedAvg, every client transmits updates to the server after undergoing a brief period of local training. However, local models may

diverge significantly from the global model when clients' data differs greatly. FedProx fixes this by keeping local models closer to the global model during training by including a *proximal* term in the local objective function. This method can manage diverse data and system environments and assist in reducing problems such as non-IID data and model poisoning.

***Federated Matched Averaged (FedMa)*:** FedMA [368] is a layer-wise federated learning algorithm for modern CNNs and LSTMs that employs Bayesian nonparametric techniques to adapt to data heterogeneity. Through layer-by-layer matching and averaging of hidden elements such as neurons for fully connected layers, hidden states for LSTM, and channels for convolution layers with comparable feature extraction signatures, FedMA builds the shared global model. FedMA's ability to effectively use educated local modals is a feature that many federated learning applications seek. This approach outperforms FedAvg and FedProx after a few training cycles and yields good results with heterogeneous learners.

***Stochastic Controlled Averaging for Federated Learning (Scaffold)*:** Scaffold [369] is a federated learning optimization algorithm that uses control variates estimates and corrects the client drift at each client, which is the difference between the local and global model update deviations. Each client tracks the direction of its local model updates with a control variate. After aggregating these local control variates to form a global control variate, each client computes its drift as the difference between its local and the global control variate [370]. This minimizes differences between local models during aggregation, enhancing global convergence. SCAFFOLD is unaffected by client sampling or data heterogeneity and necessitates a drastically reduced number of communication rounds. Additionally, SCAFFOLD can use the client's data's similarity to its advantage for quadratics, resulting in even faster convergence.

6.5 Applications FL on HAR

Deep learning has driven significant advancements in high-accuracy HAR classifiers. However, training these models on real-world data from diverse smart devices introduces several major challenges, which can be mitigated through federated learning (FL). This section discusses the benefits of FL in addressing these challenges.

6.5.1 FL for Privacy Improvement

When centralized ML-based solutions are deployed, HAR can expose data to privacy issues through open data sharing with the cloud or data centers. Despite the fact that cloud servers' powerful computing capacities allow for efficient data training and analysis, a centralized ML-based solution for HAR comes with serious privacy risks [371]. FL enhances user data privacy by storing the original data on the local device and only sharing modifications to the model parameters [372]. Secure aggregation, which guarantees that model updates are safely and secretly aggregated on the server, is a crucial part of FL for HAR.

Secure aggregation can be achieved through the use of Differential Privacy (DP) [373]. Differential privacy (DP) provides a mathematical guarantee of privacy by adding carefully calibrated noise to the data or computation. This ensures that even if the update improves the global model, the server won't be able to infer specific information about the local data on a device from the model change. Secure multiparty computation (SMPC) is an additional technique for safe aggregation in FL [374]. SMPC is a branch of cryptography that allows for the cooperative calculation of a function using inputs from several participants while protecting the privacy of such information. In order to guarantee that the server can only compute the aggregate update and is unable to access the individual model updates provided by the devices, SMPC may be utilized in the FL context.

6.5.2 FL for Reducing Communication

A centralized training approach for HAR systems may lead to high communication costs because of data transfer, bandwidth requirements, and energy use. To be more precise, centralized models require that all collected data be transferred to a single location for training. This can lead to significant data transfer volumes, higher network use, and associated costs, depending on the number of devices, data volume, and data complexity. Furthermore, regular training data transfers to the central server require a significant amount of bandwidth in order to obtain updates and results. More network congestion and higher communication costs could result from this, especially in areas with limited bandwidth or high data unit costs. Finally, data transmission over a network uses energy as well [375,376].

In federated learning (FL), model training is performed locally on client devices, and communication with the central server occurs through parameter exchanges rather than direct data sharing. The server updates the global model solely through basic parameter aggregation. This approach protects local user data while preserving the server's processing and storage capacity. FL can also acquire a global model with better performance through a client-server connection. This method differs from conventional centralized training, which gathers all distinct local datasets and sends them to a central server for model training [352].

6.5.3 FL for Scalability

A centralized machine learning model may face scalability issues in HAR systems as the number of devices increases, due to the large volume of data that must be transmitted and processed. In contrast, federated learning (FL) distributes the learning process among the devices themselves, which greatly simplifies scaling to a large number of devices. Federated learning (FL) is therefore a popular choice for large-scale HAR applications, such as those in smart cities or extensive healthcare systems [360].

6.5.4 FL for Real-Time

In centralized learning systems, latency often occurs because data must be transmitted to a central server for processing [377]. This latency may cause issues for applications that need to be completed quickly, like medical emergency detection in HAR systems. For real-time HAR systems to function effectively, resource-constrained devices such as wearables, smartphones, and smart home sensors require fast, low-latency predictions.

Federated Learning processes data locally on each device, significantly reducing latency and supporting real-time HAR. Since each device can train the model and generate predictions based on its data without waiting for the server model, decisions may be made more swiftly and efficiently, enhancing the HAR system's real-time responsiveness [360]. Model compression [378], asynchronous and lightweight FL protocols [379], and personalized updates [380] are important facilitators that guarantee low-latency predictions on edge devices. Hierarchical FL architectures further minimize latency by performing aggregation closer to the data source.

6.6 FL for Human Activity Recognition

Innovative deep learning methods for human activity recognition leverage large amounts of sensor data to achieve high levels of accuracy. However, there are certain disadvantages to training such models at a data center utilizing information gathered from smart devices, namely potential privacy risks and high connection costs. To address these issues, federated learning can be used to train a global classifier by aggregating multiple local models trained on data from different clients [371]. By utilizing FL's capabilities, data from several users can be combined to produce more resilient models without sacrificing user privacy or incurring

significant computational costs. Consequently, recent studies have explored integrating FL into HAR systems to achieve improved performance.

The authors of [381] implemented three FL algorithms including FedAvg, FedMA, and FedPer [382] for the HAR task and evaluated their performance against a centralized training approach. In the federated learning experiments, each subject in the REALWORLD dataset is treated as a separate client with its own data, resulting in a total of 15 clients. They had expected that while maintaining a high level of generalization, this learning approach would result in a high degree of device adaptation. The results indicate that the FedAvg method indeed exhibits this behavior and outperforms other more complex algorithms. The results also highlight the limitations of other, more complex FL algorithms for the HAR task. Notably, although FedMA and FedPer are designed to enhance the personalization of local models, they still rely on averaging client parameters to some extent. Furthermore, The study [383] investigated how well federated learning (FL) techniques work to recognize human activity from smartphone sensor data while protecting privacy. They applied five different FL aggregating methods including Vanilla Averaging, Weighted Averaging, Selective Averaging, Best-Model Update, and Continual Federated Learning, across two model architectures of MLP and CNN. These evaluations provide insight into how FL frameworks balance model architecture, inference time, and correctness. In particular, the CNN model that employed the Selective Averaging technique outperformed the other configurations that were examined.

Federated learning is one of the most widely recognized and commonly used approaches for multi-user, sensor-based HAR with multi-adaptation capabilities. Through the integration of the subjects' local models into a center model and the sharing of the center model with each subject for local model performance improvement, federated learning helps numerous subjects with heterogeneous data [384]. Anicai and Shakir [385] introduced a cutting-edge approach that combines federated learning (FL) and genetic mutation (GM) for activity recognition in multi-resident settings. Federated learning enhances privacy by allowing multiple devices to collaboratively train a model without data sharing, while genetic mutation introduces variability in model weights prior to federated averaging, mimicking evolutionary processes. For the activity classification task, two recurrent neural network models, LSTM and GRU, were employed, and a genetic mutation operator was used on the weights of the models before federated averaging. This combination has proven effective in enhancing model performance while maintaining user privacy and facilitating early model deployment.

To adapt FL models for heterogeneous devices, the authors of [386] dynamically adapt model layers and sizes to enable heterogeneous devices to participate effectively in FL. They suggested FedDL, a federated learning system for HAR that can dynamically develop personalized models for various users by capturing the underlying user relationships. In particular, they created an iterative, bottom-up, layer-wise dynamic layer sharing technique that learns the similarity of users' model weights to create the sharing structure and merges models appropriately. Based on the dynamic sharing approach, FedDL combines local models, greatly accelerating convergence while preserving excellent accuracy. In the same direction, the issues of heterogeneity, privacy preservation, label scarcity, and real-time in HAR based on multi-modality sensor data were discussed by the authors in [387]. To get over all of the aforementioned challenges, they suggested a personalized federated HAR architecture called FedHAR. FedHAR was the first to introduce hierarchical attention architecture for aligning features at different levels. Subsequently, a semi-supervised online learning approach was suggested for online HAR tasks. This approach included a semi-supervised learning loss to aggregate gradients from all label clients and unlabeled clients, a novel algorithm for computing unsupervised gradients under the consistency training proposition, and an unsupervised gradients aggregation strategy to address the concept drift and convergence instability issue in online learning. Along the same lines, to facilitate co-training among heterogeneous devices, Wang et al. [388] proposed Hydra, a hybrid-model

federated learning approach that allows devices to train models tailored to their individual computational capabilities. To be more precise, Hydra uses BranchyNet [389] to create a large-small global hybrid model that allows heterogeneous devices to train the appropriate model components based on their computational power. It proposes a matching strategy for efficient co-training between high- and low-performance devices and groups devices according to model similarity to lessen the effects of data heterogeneity. Furthermore, a large-to-small knowledge distillation algorithm is suggested for devices with limited resources in order to maximize the effectiveness of information transfer from large models to tiny models, hence addressing the low accuracy of small models.

Despite providing HAR with a certain amount of privacy, FL has many drawbacks, especially when training on low-power and low-computational resource devices like wearable sensors. For energy-efficient and privacy-preserving HAR, the authors of [390] suggested the hybrid neuromorphic federated learning (HNFL) approach, a federated learning architecture that combines spiking neural networks (SNNs) with long short-term memory (LSTM) networks. The hybrid spiking-LSTM (S-LSTM) model synergistically combines the sequence modeling capabilities of LSTMs with the event-driven efficiency of SNNs. This fusion establishes the S-LSTM as a pioneering concept for HAR in a federated environment by providing a harmonious balance between accuracy and efficiency. Furthermore, Shaik et al. [391] presented FedStack, a unique federated learning architecture that facilitates the ensembling of different architectural client models. This study avoids model compilation issues in the global model and gets around the drawbacks of typical federated design, where clients may have different local model architectures.

In order to construct reliable FL-based HAR models, a balance between generalization and personalization must be struck. In order to tackle this problem, the author of [392] suggested FedAR, a new hybrid strategy for HAR that leverages the advantages of both federated and semi-supervised learning. In fact, FedAR is an FL framework for HAR that addresses personalization while also taking data scarcity into account. FedAR uses label propagation and active learning to semi-automatically annotate local streams of unlabeled sensor data. It also employs FL to build a global activity model that is both scalable and privacy-aware. Additionally, FedAR incorporates a transfer learning approach to optimize the global model for every user. The findings show that label propagation and active learning together produce recognition rates that are on par with fully supervised learning. In the same direction, The study [393] introduced FedCLAR, a novel federated clustering approach for HAR. FedCLAR integrates transfer learning techniques with federated clustering for FL-based HAR in order to address the non-IID issue. FedCLAR only chooses a subset of the model weights that each client shares in order to calculate a similarity score and create user groups using a hierarchical clustering technique. The chosen weights describe the subject-specific activity patterns in an intuitive manner. Similar to a typical federated learning environment, FedCLAR will employ a generic global model that has been trained by all participating users for users who are not able to be included in any cluster. Additionally, FedCLAR further enhances personalization by fine-tuning activity recognition on an individual basis through transfer learning techniques.

Traditional FL techniques frequently make the assumption that clients' data is independently and identically distributed (IID). However, Human activity in real-world scenarios exhibits variations, resulting in skewness where identical activities are executed uniquely across clients. This can cause local model objectives to diverge from the global model objective, thereby hindering overall convergence. FedCoad, a novel strategy created to handle non-IID amongst various clients in real-world contexts, was presented by [394]. The FedCoad model employs contrastive learning to reduce the representation gap between global and local models, thereby facilitating global model convergence. Adaptive control variables penalize local model updates according to the model weights and the control variables' rate of change. Experimental

results show that FedCoad outperforms other methods on benchmark datasets, particularly under skewed (non-IID) data settings.

Table 7 presents a concise summary of recent studies that integrate FL into sensor-based HAR systems, highlighting their distinctive features and unique contributions to advancing the field.

Table 7: A summary of recent studies on Federated Learning in HAR systems

Reference	Year	FL architecture	Dataset(s)	Methodology/Contribution
Tu et al. [386]	2021	CFL	LiDAR (collected data) UWB [395] HARBox [395]	A dynamic layer-sharing system that iteratively, layer-by-layer, merges models based on similarities between users' model weights to create the sharing structure.
Shen et al. [396]	2022	CFL	HHAR, PAMAP2, ExtraSensory, SmartJLU	Utilizing an attention module for every client in order to protect data privacy while learning both client-specific and globally associated information.
Presotto et al. [392]	2022	CFL	MobiAct, WISDM	A new hybrid approach for HAR integrates federated learning (FL) with semi-supervised learning (e.g., active learning) to leverage the strengths of both techniques
Presotto et al. [393]	2022	Hierarchical FL	MobiAct, WISDM	The server clusters and merges local models based on the similarity of their weights
Shaik et al. [391]	2022	Hierarchical FL	MHEALTH	A heterogeneous stacked FL architecture is proposed to support diverse local model architectures across different devices
Yu et al. [387]	2023	CFL	RealWorld, HAR-UCI	A personalized federated HAR framework leveraging semi-supervised online learning
Khan et al. [390]	2023	CFL	UCI-HAR, RealWorld	A neuromorphic FL-based model that integrates the strengths of both LSTM and SNN within a federated learning setting
Wang et al. [388]	2024	Hierarchical FL	HHAR, MobiAct, HARBox [395]	A hybrid-model FL mechanism that adapts training to device capabilities, clusters devices by model similarity, and pairs high- and low-performance devices for efficient co-training
Zhang et al. [397]	2024	CFL	WISDM	A novel framework extending the multi-level FL architecture with three specialized methods to address statistical, device, and model heterogeneities

(Continued)

Table 7 (continued)

Reference	Year	FL architecture	Dataset(s)	Methodology/Contribution
Iwan et al. [394]	2025	CFL	WISDM, MotionSense, HHAR	A federated contrastive learning approach designed to address client skewness in real-world settings
Albogamy [398]	2025	CFL	UCI-HAR, HARTH, HAR7+	Integrating federated learning with a hybrid LSTM-GRU model to achieve accurate human activity recognition while preserving user privacy and minimizing computational costs on edge devices
Kabir et al. [383]	2025	CFL	UCI-HAR	Implementing five distinct FL aggregation techniques across two model architectures to provide insight into how FL frameworks balance model design, inference time, and accuracy.

7 Explainable AI (XAI) in HAR

7.1 Explainable AI (XAI)

Deep Learning (DL) models are successful because of their large parametric capacity and effective learning algorithms. Since these models often have millions of parameters and hundreds of layers, they are frequently thought of as intricate black-box systems. A prevailing view is that focusing on performance reduces both transparency and interpretability. As the use of deep learning systems continues to grow, it has become crucial to understand their inner mechanisms and derive meaningful insights from their outputs. As a result, model performance and transparency need to be balanced [399,400].

Explanations of model decisions are crucial in many applications to maintain transparency and confidence. This is especially important in the medical field, as physicians need to comprehend how an AI system, for instance, identifies a disease from a CT scan. Given that AI models are inherently fallible, interpretability is crucial not only for fostering trust but also for preventing potentially life-threatening errors [401].

A new field of artificial intelligence study called Explainable AI (XAI) has emerged in response to the need for explainability. In complicated black-box AI systems, XAI tackles the issues of explainability, interpretability, and transparency [402]. It seeks to create models that not only retain a high degree of predicted accuracy but also offer clear and interpretable explanations for their findings. In essential applications, this transparency is especially crucial since it helps end users better understand and trust AI decisions. Furthermore, XAI supports the ongoing enhancement of prediction performance by helping system developers find problems in their models. Additionally, visualization techniques improve user understanding and facilitate more productive interactions with AI systems [403].

XAI has drawn a lot of interest in a variety of application areas, which has caused a sharp rise in the tools and methods that have been offered in both academia and industry. Existing XAI techniques include a wide range of aspects and capabilities, from straightforward exploratory data analysis to techniques intended for deciphering intricate AI models. The two main categories of explainable AI (XAI) models are intrinsic

and post-hoc models. Intrinsic models are inherently interpretable, as their structures and components are designed to be easily understood. Examples of such models include linear regression, decision trees, and rule-based systems. These models naturally shed light on the process of making predictions and have a comparatively straightforward structure. Conversely, post-hoc models analyze a trained model externally, frequently with the aid of another model, to offer reasons. Post-hoc explanations use outside techniques to derive interpretable information from the original model [404].

Post-hoc methods can be further divided into two categories: model-specific and model-agnostic. Model-specific strategies are designed for a certain kind of model. They produce explanations by utilizing the internal organization or functions of that model. Model-agnostic methods don't make any assumptions about how the model functions inside. Rather, they examine the input-output behavior of the model as a "black box". These methods can be used with a variety of AI models. For instance, by methodically altering the model's inputs, perturbation-based approaches produce and understand outcomes.

In the following, we explore the most used XAI models including Local Interpretable Model-agnostic Explanations (LIME) [405], SHapley Additive exPlanations (SHAP) [406], Anchors [407], Integrated Gradients [408], and Gradient-weighted Class Activation Mapping (Grad-CAM) [409].

7.1.1 LIME

Local Interpretable Model-agnostic Explanations (LIME) [405] is a model-agnostic XAI method that explains how individual attributes affect the prediction for a particular case. It offers the contribution of individual features, which are frequently displayed in plots, as well as the probability that the instance belongs to each class in classification tasks. LIME approximates the complex model in the instance's local neighborhood using a simple, interpretable linear model. It accomplishes this by altering the target instance's feature values to create a collection of synthetic samples surrounding it. The original model assesses these samples, and a kernel function based on proximity to the original instance is used to weight the outputs. After that, a local linear model is trained on this neighborhood, making sure it closely resembles the behavior of the original model by optimizing a fidelity function. Lastly, the linear model's coefficients show the feature weights, offering a comprehensible explanation for the prediction [402,410]

7.1.2 SHAP

Shapley additive explanations (SHAP) [406] is a popular post-hoc, model-agnostic explainability technique that works with any machine learning model. It quantifies each feature's contribution to a prediction by assigning it a numerical number known as the SHAP value. The method is based on cooperative game theory, in which the payment is determined by the model's prediction and features are viewed as players. Through the aggregate of these contributions, SHAP provides a fair and consistent feature of the importance of the metric. It enables both global interpretability and local interpretability and highlights the most important features influencing model predictions [411].

7.1.3 Anchors

Anchors [407] is a post-hoc, model-agnostic explainability technique in XAI, developed to clarify individual predictions made by machine learning models. It expands on the LIME methodology to offer explanations in a more scalable and effective way. The main concept involves pre-calculating explanations for a subset of the dataset's examples, known as "anchors," and then using these anchors to produce explanations for other cases. This method makes it possible to interpret the predictions of big and complicated models by drastically lowering the computing cost of creating explanations. As an extension of LIME, it maintains

the local accuracy of the original technique while providing explanations that are more precise and effective [412].

7.1.4 Integrated Gradients

Integrated Gradients [408] is a post-hoc, model-specific explainability method for deep learning models. It measures each input feature's contribution to the model's output in order to assign it a significance score. Two fundamental tenets of the approach are implementation invariance, which assures identical attributions for functionally equivalent models, and sensitivity, which guarantees non-zero attributions for inputs that influence predictions. The drawbacks of conventional gradient-based techniques are addressed by Integrated Gradients, which aggregate gradients along a path from a baseline to the input. This method offers a feature importance metric that is both theoretically valid and computationally effective [413].

7.1.5 Grad-CAM

Gradient-weighted Class Activation Mapping (Grad-CAM) [409] is a post-hoc, model-specific explainability method that is specifically used to Convolutional Neural Networks (CNNs). It offers visual explanations for any CNN-based model without the need for retraining or architectural changes. As is well known, High-level semantics and spatial information are both well-captured by CNN layers. Building on this characteristic, an ideal representation for extracting important information is frequently found in the last convolutional layer. Grad-CAM uses gradient information backpropagated to the last convolutional layer to award significance scores to each neuron for a certain target class. The target class and the input data are taken into account by the procedure. The input is sent to CNN, which applies task-specific computations, to determine the raw score for a particular category. Every gradient, with the exception of those related to the target class, is set to zero during backpropagation. In order to create the Grad-CAM heatmap for the target class, the non-zero gradients are propagated back to the pertinent feature maps [414].

7.2 XAI for HAR Systems

Explainable Artificial Intelligence (XAI) has the potential to improve system interpretability, transparency, and confidence in Human Activity Recognition (HAR), especially in the healthcare and ambient assisted living (AAL) domains. Human-comprehensible explanations for every prediction are the goal of XAI, which is essential when important decisions depend on the recognition of Activities of Daily Living (ADL). For example, in healthcare systems that monitor senior citizens, ADL detection supports higher-level behavior analysis, which helps physicians diagnose and treat patients. XAI can enhance clinicians' trust in decision support systems that rely on ADL recognition. Explanations also enable data scientists to improve systems by fine-tuning sensors, optimizing algorithms, or enhancing training data [415].

Despite these advantages, there is presently little study on DL-based HAR using XAI since current methods, while accurate, frequently lack interpretability. In order to recognize explainable sensor-based ADLs, Arrotta et al. [415] introduced DeXAR, a system that uses CNN classification of semantic images obtained from raw sensor data. DeXAR creates a heat map linking each pixel of the input semantic image to a relevance value by obtaining an explanation for each prediction. The heat map intuitively illustrates the reasoning behind the CNN-based classification of a semantic image by utilizing the actual learnings of the model during training. They investigated three potential XAI strategies: LIME, Model Prototypes [416], and Grad-CAM. Their findings show that the suggested white-box XAI method based on prototypes works effectively. Nevertheless, a significant drawback of DeXAR is that it only records snippets of larger events, producing activity classifications and explanations during little time periods. These window-level explanations are helpful for improving the system, but they are more difficult for non-technical users to

understand in the context of the full activity. Consequently, the study by [417] suggested DeXAR++, a unique technique for employing deep learning classifiers to generate explanations of examples of human activity. This method uses a computer-vision XAI technique to encode sensor data time windows as images and generate explanations for each window. These window-level explanations are then combined by DeXAR++ to produce a single, thorough explanation for every instance of an approximated activity. It also presents a new visualization technique that is intended to be understandable by non-expert users.

The author of [418] created visual explanations of the Human Activity Recognition by adapting gradient-weighted class activation mapping (grad-CAM) to one-dimensional convolutional neural network (CNN) architectures. The efficacy of SHAP, LIME, and Anchors for human activity recognition using smart home data was assessed in the study by [419]. Additionally, they enhanced LIME to create LIME+, which pinpoints the windows of time when sensors have the biggest impact on classifying an instance. The study also suggested ways to use the outputs of these models to produce understandable plain language descriptions. The authors used the CASAS: Milan dataset to train an LSTM model and produced explanations using LIME, LIME+, SHAP, and Anchors in order to evaluate the advantages of XAI. Their results revealed that SHAP provided superior explanations compared to LIME and Anchors. A feature importance analysis in XAI was proposed by the authors of [420] in order to identify which features have a bigger influence on the model's decisions. Additionally, they use SHAP and LIME in conjunction to conduct an interpretability analysis of a CNN-LSTM algorithm-based HAR model.

Jeyakumar et al. [421] introduced X-CHAR, an explainable model for complex human activity recognition that dispenses with the requirement for precise low-level activity annotations. X-CHAR maintains the performance of end-to-end deep learning on time series data while representing activities as sequences of high-level, human-understandable ideas and offering explanations through concept sequences and counterfactual instances. Three parts make up the model: the Sensor Fusion Module, which extracts and aligns features; the Temporal Bottleneck Module, which finds patterns and predicts concepts over time; and the Classification Module, which labels the end activity.

Table 8 provides a concise summary of recent studies that used XAI methods in sensor-based HAR systems, highlighting their key features and unique contributions to advancing the field.

Table 8: Summary of XAI approaches applied in HAR systems

Reference	Year	DL Model	XAI Technique	Dataset(s)	Achievements
Arrotta et al. [415]	2022	CNN	Dexar (LIME, Grad-CAM, Model Prototypes)	CASAS: Milan	The propose white-box XAI method is suitable for ADLs recognition when sensor data are represented as images.

(Continued)

Table 8 (continued)

Reference	Year	DL Model	XAI Technique	Dataset(s)	Achievements
Arrotta et al. [417]	2022	CNN	Dexar++ (LIME, Grad-CAM, Model Prototypes)	MARBLE	DeXAR++ demonstrated improved explainability over the original DeXAR, achieving explanation scores of 0.85 with LIME, 0.80 with Grad-CAM, and 0.96 with Model Prototypes, compared to 0.80, 0.65, and 0.94, respectively, in DeXAR.
Aquino et al. [418]	2022	CNN	Grad-CAM	UniMiB-SHAR	Application of the grad-CAM approach to detect signal discontinuities and other database bias issues.
Das et al. [419]	2023	LSTM	LIME, LIME+, SHAP, Anchors	CASAS: Milan	The approach achieved explanation scores of 64.1, 73.1, 92.2, and 87.4 for LIME, LIME+, SHAP, and ANCHORS, respectively.
Jeyakumar et al. [421]	2023	X-CHAR	X-CHAR	Opportunity, Nursing Activity [422]	Experimental results demonstrated that the proposed model delivers more interpretable explanations than existing methods for complex activity recognition.
Liu et al. [420]	2024	CNN + LSTM	SHAP, LIME	UCI-HAR	The experimental results demonstrate that LIME offers a clear and interpretable breakdown of feature contributions for individual predictions

8 Research Directions and Future Aspects

The analytical framework proposed in this study offers a thorough perspective on key aspects of Human Activity Recognition (HAR) within smart environments. In particular, we have provided a comprehensive overview of the various deep learning methodologies employed in HAR systems. Despite these advances, several promising research avenues remain unexplored. Addressing these gaps could significantly improve the state of the art and accelerate the adoption of novel HAR approaches in both current and future

technologies. In light of the literature reviewed, this section highlights critical issues and challenges that warrant further investigation and development.

8.1 Data Acquisition

Data is the cornerstone of artificial intelligence, and the efficacy of deep learning models is closely tied to the quality and quantity of the data used for training. In the context of Human Activity Recognition (HAR), developing models with high generalizability necessitates meticulous attention to data collection. A sufficiently large and diverse training dataset is essential for capturing the variations in activities across different contexts and populations. However, acquiring high-quality data in large volumes, particularly for HAR, presents significant challenges due to the labor-intensive process of gathering raw data from various wearable devices, often making it prohibitively expensive. One potential approach to mitigate these costs is crowdsourcing, where individuals contribute their own data. While crowdsourcing can reduce data collection costs, it introduces the challenge of ensuring data quality. Developing strategies to incentivize participation while maintaining high data standards is a critical area of research [10].

Novel methodologies for collecting high-quality data must be explored and implemented to further the progress of HAR research. Several strategies can effectively address the issue of data scarcity. Transfer Learning (TL) is one such approach, which involves leveraging pre-trained models or model components to build new models, thereby reducing the need for extensive new data. TL is particularly valuable in scenarios where labeled data is scarce, as it allows for the reuse of knowledge gained from related tasks.

Data augmentation is another powerful technique to enhance the generalizability of trained models. It is especially useful in situations with small datasets or imbalanced data, which are common challenges in real-world HAR applications. By augmenting the dataset with synthetic instances, classification accuracy can often be improved [423]. For instance, Jeong et al. [424] proposed a novel data augmentation method for IMU sensor signals, employing techniques such as time-warping and data masking. Time-warping distorts time-series data along the temporal axis, while data masking selectively removes portions of the input signals, thereby creating a more varied dataset for training. Wang et al. [425] introduced another innovative approach by modifying the sampling frequency of sensor data to enhance the coverage of the sample space, aiming to replicate more realistic activity data. This technique ensures that the augmented data better represents the diversity of human activities.

The recent advancements in deep generative models offer additional solutions to the problem of sensor data scarcity. While Autoencoders (AEs) have been explored for generating synthetic sensor data, they have not gained widespread acceptance due to the lower quality of the synthetic data produced. In contrast, Generative Adversarial Networks (GANs) have emerged as a more promising approach due to their ability to generate highly realistic synthetic instances [426–428]. Jimale and Mohd Noor [429] proposed a robust GAN architecture that integrates convolutional layers with fully connected networks in both the generator and discriminator, resulting in higher-quality samples. Hu [430] suggest Balancing Sensor Data Generative Adversarial Networks (BSDGAN) in order to create sensor data for minority human activities. BSDGAN uses an autoencoder to initialize model training, ensuring that the characteristics of minority activities are adequately learned. The generated data is then combined with the original dataset to balance the representation of different activity classes. Another innovative approach is the hierarchical multi-modal GAN model (HMGAN), proposed by [431], which is specifically designed for sensor-based HAR. HMGAN comprises multiple modal generators, a hierarchical discriminator, and an auxiliary classifier. This architecture allows the model to learn complex multi-modal data distributions from sensor data, providing discrimination outputs for both low-level modal and high-level overall consistency losses. This balance

between modality details and global consistency enhances the model's ability to generalize across different sensor modalities and activity types.

These advancements in data acquisition and augmentation techniques are critical to overcoming the challenges of data scarcity in HAR and ensuring the development of robust, generalizable models. As research continues to evolve in this area, the potential for more accurate and reliable HAR systems in smart environments will significantly increase.

8.2 Data Annotation

Labeling data is a critical prerequisite for deep supervised learning, particularly in the context of Human Activity Recognition (HAR). While labeling image and audio data can often be straightforward, thanks to visual or auditory confirmation, annotating time-series data captured by sensors in HAR settings is significantly more challenging. This complexity arises because accurate labeling of human activities typically requires additional sensing sources and expert intervention. The process is not only labor-intensive but also costly and time-consuming, particularly when dealing with large datasets. Consequently, a pressing challenge in HAR research is to develop methods that maintain high performance while minimizing the need for extensive data annotation efforts [432]. Several strategies have been proposed to address this challenge, including data augmentation, semi-supervised learning, unsupervised learning, and active learning. Each of these approaches aims to reduce the dependency on labeled data while still enabling the model to learn effectively.

Semi-Supervised Learning leverages both labeled and unlabeled data to learn more generalizable feature representations. For instance, Zeng et al. [433] introduced two semi-supervised techniques that utilize Convolutional Neural Network (CNN) to learn discriminative hidden features from both labeled and unlabeled raw sensor data. This semi-supervised CNN model is designed to conduct feature learning directly on raw data, enhancing their ability to generalize. In another study, Qu et al. [434] proposed a novel Context-Aware Mutual Learning technique for semi-supervised HAR. This method employs a mutual learning framework where two networks learn from each other, reducing overfitting. Additionally, batch-level supervision and contextual information are introduced via Distribution-Preserving Loss and Context-Aware Learning, which are extracted from raw sequences for the main network, further improving performance. Yun and Wang [435] developed a semi-supervised strategy that enhances the performance of HAR algorithms by extracting embedded human behavioral information. This method computes pseudo-labels for a broader set of unlabeled radar data, improving recognition accuracy. To refine the model-generated pseudo-labels and mitigate the impact of incorrect pseudo-labels on performance, this approach also integrates contrastive learning concepts.

Unsupervised Learning is traditionally used in exploratory data analysis to identify patterns in data. More recently, its application in sensor-based HAR has gained attention as a way to exploit unlabeled data [436]. Sheng and Huber [437] proposed an unsupervised embedding learning approach that projects activity data into an embedding space. This model is based on an autoencoder framework and leverages temporal coherence and location preservation, which are inherent characteristics of human activity. Takenaka and Hasegawa [438] introduced a novel unsupervised representation learning technique that combines segment discrimination (SD), autoencoder (AE), and feature-independent SoftMax (FIS) to enable deep learning models to obtain feature representations from accelerometer data without requiring activity labels. This method demonstrates the potential of unsupervised learning to generate meaningful features even in the absence of labeled data.

Active Learning is a specialized form of semi-supervised learning that focuses on selecting unlabeled data based on an objective function that identifies instances with low prediction confidence for human

annotation. This approach dynamically prioritizes data that is most likely to improve model performance when labeled. Recent research has attempted to combine active learning with deep learning (DL) to harness the power of DL's classification capabilities while reducing annotation requirements [19]. For instance, the authors of [439] proposed an activity recognition technique based on Dynamic Active Learning. Unlike traditional active learning methods that select samples based on a predefined label set, this approach dynamically identifies novel activities outside the existing label set while simultaneously choosing informative samples from known classes. Another innovative framework proposed by Bi et al. [439] combines semi-supervised learning and active learning. The core idea is to use unlabeled examples in a semi-supervised manner while actively selecting the most informative samples for annotation. This dual approach minimizes annotation costs while maintaining model performance, offering a promising direction for future HAR research.

These advanced data annotation strategies are essential for scaling HAR systems to more complex and diverse real-world environments, where manual labeling is impractical. As these methods continue to evolve, they will likely play a crucial role in enabling more efficient and accurate HAR systems.

8.3 Imbalanced Data Handling

Data imbalance is a prevalent issue in real-world applications, particularly in domains like Human Activity Recognition (HAR). Imbalanced data refers to scenarios where certain classes are significantly underrepresented compared to others, leading to biased model performance that favors the majority class. Addressing this imbalance is crucial for ensuring accurate and fair predictions across all classes. Three primary strategies have been identified for handling imbalanced data: data-level techniques, algorithm-level techniques, and hybrid approaches. Data-Level Techniques aim to balance the class distribution by either adding or removing samples from the training dataset. This can be achieved through oversampling, where additional samples are generated for the minority class, or undersampling, where samples from the majority class are removed to achieve a more balanced dataset. Techniques like Synthetic Minority Oversampling Technique (SMOTE) are commonly used for oversampling. Algorithm-Level Techniques modify existing learning algorithms to reduce bias towards the majority class. These approaches typically require a deep understanding of both the learning process and the specific application domain. For instance, cost-sensitive learning adjusts the algorithm to penalize misclassifications of minority class instances more heavily, thus encouraging the model to focus on correctly classifying these underrepresented cases. Hybrid Approaches combine both data-level and algorithm-level strategies to enhance classification performance. By integrating these methods, hybrid approaches can address the limitations of using either strategy alone, leading to more robust and accurate models in the presence of imbalanced data [440,441].

Several studies have specifically addressed the challenge of imbalanced class distributions in HAR [442–445]. Guo et al. [446] proposed an enhanced version of SMOTE tailored for HAR, which addresses class imbalance by generating new synthetic minority class activities in regions near existing minority class instances. This method adjusts the distribution of activity classes by leveraging the Euclidean distance between minor activity instances, rather than relying solely on linear interpolation. Singh et al. [447] explored and compared three different approaches to managing class imbalance in HAR datasets: cost-sensitive learning, undersampling, and oversampling. Their study focused on enhancing the transparency and reliability of deep learning models by systematically experimenting with both data-level and algorithmic solutions to overcome class imbalances. Their findings highlight the importance of explicability in deep learning models, particularly when dealing with imbalanced datasets, as it directly impacts the model's transparency and reliability.

The ongoing exploration of these techniques underscores the importance of effectively managing data imbalance in HAR and similar fields. As research in this area progresses, more sophisticated and adaptive methods are likely to emerge, further improving the performance and fairness of deep learning models in the presence of imbalanced data.

8.4 Data Segmentation

Data segmentation is a crucial preprocessing step in Human Activity Recognition (HAR), where continuous sensor data streams are divided into meaningful segments representing distinct activities. This process is vital for real-time applications in smart environments, as it enables the accurate classification and recognition of activities by isolating different activities within the data stream. Effective segmentation enhances the accuracy of classification models by providing clear boundaries for the start and end of each activity instance.

The most widely used method for segmenting sensor data streams is the sliding window approach. This technique involves dividing the continuous sequence of sensor events into overlapping or non-overlapping segments. Two primary configurations of the sliding window method are commonly employed: Time-Based Sliding Windows, and Sensor-Based Sliding Windows. The Time-Based Sliding Windows approach divides the entire sensor event sequence into segments of equal time intervals, irrespective of the sensor events themselves. This method is simple to implement and works well in scenarios where activities have consistent durations [448,449]. In contrast, the Sensor-Based Sliding Windows configuration segments the sensor event sequence based on the number of sensor activations, ensuring that each segment contains an equal number of sensor events. This method is particularly useful when the frequency of sensor activations varies significantly over time [450,451].

While these traditional methods are effective in many scenarios, they have limitations, particularly when dealing with dynamic or overlapping activities. To address these challenges, several advanced segmentation techniques have been proposed. For instance, Wan et al. [452] introduced a dynamic segmentation technique that uses sensor and time correlations to identify basic activities. They calculated the Pearson product-moment correlation (PMC) between sensor pairs to determine their association, allowing for more accurate segmentation of complex activities. Another innovative approach is presented by Chen et al. [453], who developed a hybrid segmentation technique combining Fuzzy C-Means (FCM) clustering with Change Point Detection (CPD). This method first classifies sensor events based on their locations using FCM, and then applies CPD to detect transition actions between different segments. The resulting segmentation sequence reflects the natural transitions between activities, improving the performance of HAR systems. Kim et al. [454] proposed a correlation-based real-time segmentation system designed for multi-user collaborative activities. Their system uses a time-based window approach to group concurrent or overlapping events into a single event set. Transition points between these sets are identified by comparing duration and historical correlation values, allowing the system to segment activities accurately even in complex, multi-user environments. Najeh et al. [455] introduced a real-time event segmentation method specifically for recognizing human activities in building settings. This approach involves two key steps: calculating the sensor and temporal correlation and identifying the triggering sensor for each action. The segmentation is performed dynamically by determining the Pearson product-moment correlation (PMC) coefficient between sensor events, which allows for real-time segmentation in environments where activities occur in rapid succession.

These advanced segmentation techniques are critical for improving the robustness and accuracy of HAR systems, particularly in real-world applications where activities are often complex, overlapping, or occurring in dynamic environments. As HAR technology continues to evolve, the development of more sophisticated

and context-aware segmentation methods will be essential for enabling more accurate and reliable activity recognition in smart environments.

8.5 Generalizing Cross-Dataset

Generalizing cross-dataset refers to the ability of a model to accurately recognize human activities on datasets different from the one on which it was originally trained. This is a critical challenge in Human Activity Recognition (HAR) due to the issue of domain shift, which occurs when there are differences in factors such as user groups, device types, and sensor locations between the source and target datasets. These variations create a significant obstacle for models, as they need to adapt to different conditions without prior exposure to them. Domain shift manifests in two main scenarios: cross-domain and cross-dataset. In cross-domain activity recognition, the training and testing data are from different domains within the same dataset. For example, training data might come from one user (User A), while testing data comes from another (User B) within the same dataset. Cross-dataset activity recognition, on the other hand, involves training a model on one dataset (the source dataset) and evaluating it on a completely different dataset (the target dataset), typically collected by different research teams under different conditions. This scenario is particularly challenging because it requires models to generalize across diverse environments without prior exposure to the target data [456].

Several studies are conducted with the goal of bridging the gap between source and target domains in order to address the cross-domain issue [457–461]. By matching the representation spaces of the various domains, this method can improve model performance. Nevertheless, in many real-world situations, this approach could not be feasible since it frequently implies the availability of target domain data during training.

In real-world applications, models often need to generalize across various situations without prior exposure to data from these specific contexts. This scenario leads to cross-dataset activity recognition, where models are trained on one dataset (the source dataset) and evaluated on a different dataset (the target dataset). Cross-dataset activity recognition presents two major challenges. The first challenge lies in selecting the appropriate source domain. Identifying a source domain that closely resembles the target domain while avoiding negative transfer is difficult, especially when multiple potential source domains exist for a given target task. The second challenge involves accurately transferring knowledge from the selected source domain to the target domain.

To address these challenges in cross-dataset HAR, Qin et al. [462] propose an Adaptive Spatial-Temporal Transfer Learning (ASTTL) method. ASTTL adaptively evaluates the relative importance of marginal and conditional probability distributions to learn spatial features in transfer learning. Additionally, it incorporates incremental manifold learning to capture temporal information. As a result, ASTTL can be used for both precise activity transfer and effective source domain selection by learning adaptive spatial-temporal features for cross-dataset HAR. They initially employed a Random Forest (RF) method, training the model on the original source domain and testing it on the target domain. The experimental results demonstrate that the highest accuracy is obtained when the model is trained and tested on the same dataset. For example, using the DSADS dataset as both the source and target achieves an accuracy of approximately 98%. In contrast, when DSADS is used as the source and other datasets serve as the target, the accuracy decreases substantially: UCI-HAR achieves around 58%, USC-HAD around 52%, and PAMAP around 55%. These findings indicate that cross-domain discrepancies significantly reduce model performance. To address this issue, the proposed ASTTL method improves classification accuracy when the source and target domains differ. Specifically, when trained on DSADS and tested on UCI-HAR, ASTTL achieves an accuracy of approximately 78.5%.

In another study, Hong et al. [456] presented CrossHAR, a novel HAR model designed to enhance model performance on unseen target datasets. CrossHAR operates in three key phases: It begins by exploring the principles of sensor data generation to enrich raw sensor data and diversify the data distribution. Next, it uses the enriched data in a hierarchical self-supervised pretraining approach to create a generalizable representation. Finally, CrossHAR fine-tunes the pretrained model using a small set of labeled data from the source dataset to optimize its performance in cross-dataset HAR. Due to the inherent challenges of cross-dataset activity recognition, a significant performance decline typically occurs when transferring knowledge from a source dataset to a target dataset. To validate this hypothesis, an empirical study was conducted using two widely examined datasets: UCI and HHAR. When the UCI dataset was employed as both the source and target, the LSTM model achieved an accuracy of approximately 82%. In contrast, when UCI was used as the source and HHAR as the target, the accuracy decreased to 68.15%. Notably, the proposed model mitigates this performance drop, improving the accuracy to 77.27%.

Presotto et al. [463] introduced an innovative approach that combines publicly available datasets to develop a generalized HAR model, which can then be fine-tuned with a small amount of labeled data from an unseen target domain. Their findings indicate that pre-training on multiple datasets significantly improves performance in data-scarce scenarios.

These advancements represent significant steps forward in addressing the complexities of cross-dataset generalization in HAR. As HAR applications continue to expand into diverse and dynamic environments, the ability of models to adapt to new, unseen data will be crucial for the development of reliable and robust activity recognition systems.

9 Conclusion

This study offers an exhaustive examination of state-of-the-art deep learning methodologies applied in sensor-based human activity recognition. We embark on our inquiry with a foundational exploration of the human activity recognition domain, elucidating the types of sensors commonly deployed in practical scenarios, delineating prevalent challenges, and shedding light on both foundational and emerging applications, accompanied by openly accessible datasets. Subsequently, to a concise overview of mainstream deep learning paradigms, we delve into recent research endeavors harnessing deep learning frameworks for human activity recognition from sensor data. The innate capability of deep learning to assimilate hierarchical data representations and employ sophisticated methodologies positions it adeptly to address the inherent complexities of human activity recognition tasks. Our analysis classifies relevant research across multi-layer perceptron (MLP), convolutional neural network (CNN), recurrent neural networks (RNN), long short-term memory (LSTM), gated recurrent units (GRU), Transformer, deep belief networks (DBN), and hybrid architectures integrating diverse components. Moreover, we undertake a comparative evaluation of various deep learning models, assessing their performance in terms of accuracy, architectural intricacies, and achievements within the domain. Beyond centralized deep learning approaches, we explore the role of Federated Learning (FL) in Human Activity Recognition (HAR), providing a summary of current applications and research trends. Additionally, we discuss the importance of Explainable Artificial Intelligence (XAI) in sensor-based HAR systems that utilize deep learning and provide an overview of recent studies employing XAI methods in HAR.

Despite the promising insights of our study, we acknowledge its limitations. Several deep learning models, including Temporal Convolutional Networks (TCN), Spiking Neural Networks (SNNs), and Kolmogorov-Arnold Networks (KAN), have shown effectiveness in sensor-based human activity recognition but were not included in our analysis.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Farhad Mortezapour Shiri, Thinagaran Perumal; methodology, Farhad Mortezapour Shiri; investigation, Farhad Mortezapour Shiri, Norwati Mustapha; data curation, Farhad Mortezapour Shiri, Raihani Mohamed; supervision, Thinagaran Perumal; writing—original draft preparation, Farhad Mortezapour Shiri; writing—review and editing, Thinagaran Perumal, Norwati Mustapha, Raihani Mohamed, Farhad Mortezapour Shiri. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: All data generated or analyzed during this study are included in this published article.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

Abbreviations

HAR	Human Activity Recognition
MLP	Multi-Layer Perceptron
CNN	Convolutional Neural Network
RNN	Recurrent Neural Networks
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Units
Bi-LSTM	Bidirectional Long Short-Term Memory
Bi-GRU	Bidirectional Gated Recurrent Units
DBN	Deep Belief Networks
IoT	Internet of Things
AAL	Ambient Assisted Living
ML	Machine Learning
DL	Deep Learning
SHA	Simple Human Activity
CHA	Complex Human Activity
ANN	Artificial Neural Network
RBM	Restricted Boltzmann Machine
IMU	Inertial Measurement Unit
SMOTE	Synthetic Minority Oversampling Technique
GWO	Grey Wolf Optimizer
WOA	Whale Optimization Algorithm
GAN	Generative Adversarial Network
AE	Autoencoder
CPD	Change Point Detection
PIR	Passive infrared
FPGA	Field-Programmable Gate Array
SGD	Stochastic Gradient Descent
GELU	Gaussian Error Linear Unit
FFT	Fast Fourier Transform
ReLU	Rectified Linear Unit
ROSE	Random Over Sampling Examples

FC	Fully Connected
ELU	Exponential Linear Unit
PreLU	Parametric Rectified Linear Unit
AM	Acceleration-Magnitude
UWB	Ultra-Wideband
SE	Squeeze-And-Excitation
GM	Genetic Mutation
BPA	Back-Propagation Algorithm
CWT	Continuous Wavelet Transform
HMM	Hidden Markov Model
CRF	Conditional Random Fields
CBAM	Convolutional Block Attention Module
PMC	Pearson product-moment correlation
GAR	Group Activity Recognition
EMG	Electromyogram
EEG	Lectroencephalogram
ECG	Electrocardiogram
EOG	Electrooculogram
PPG	Photoplethysmography
ADL	Activities of Daily Living
FL	Federated Learning
HFL	Horizontal Federated Learning
VFL	Vertical Federated Learning
FTL	Federated Transfer Learning
CFL	Centralized Federated Learning
HFL	Hierarchical Federated Learning
DFL	Decentralized Federated Learning
SNN	Spiking Neural Network
XAI	Explainable AI
LIME	Local Interpretable Model-agnostic Explanations
SHAP	SHapley Additive exPlanations
Grad-CAM	Gradient-weighted Class Activation Mapping

References

1. Park H, Lee GH, Han J, Choi JK. Multiclass autoencoder-based active learning for sensor-based human activity recognition. Future Gener Comput Syst. 2024;151:71–84. doi:10.1016/j.future.2023.09.029.
2. Rezaee MR, Abdul Hamid NAW, Hussin M, Ahmad Zukarnain Z. Fog offloading and task management in IoT-fog-cloud environment: review of algorithms, networks, and SDN application. IEEE Access. 2024;12:39058–80. doi:10.1109/access.2024.3375368.
3. Yang J, Liao T, Zhao J, Yan Y, Huang Y, Zhao Z, et al. Domain adaptation for sensor-based human activity recognition with a graph convolutional network. Mathematics. 2024;12(4):556. doi:10.3390/math12040556.
4. Diraco G, Rescio G, Siciliano P, Leone A. Review on human action recognition in smart living: sensing technology, multimodality, real-time processing, interoperability, and resource-constrained processing. Sensors. 2023;23(11):5281. doi:10.3390/s23115281.
5. Bouchabou D, Nguyen SM, Lohr C, LeDuc B, Kanellos I. A survey of human activity recognition in smart homes based on IoT sensors algorithms: taxonomies, challenges, and opportunities with deep learning. Sensors. 2021;21(18):6037. doi:10.3390/s21186037.
6. Jung IY. A review of privacy-preserving human and human activity recognition. Int J Smart Sens Intell Syst. 2020;13(1):1–13. doi:10.21307/ijssis-2020-008.

7. Ali Hamad R, Woo WL, Wei B, Yang L. Overview of human activity recognition using sensor data. In: 21st UK Workshop on Computational Intelligence; 2022 Sep 7–9; Sheffield, UK. doi:10.1007/978-3-031-55568-8_32.
8. Wang J, Chen Y, Hao S, Peng X, Hu L. Deep learning for sensor-based activity recognition: a survey. Pattern Recognit Lett. 2019;119:3–11. doi:10.1016/j.patrec.2018.02.010.
9. Shiri FM, Perumal T, Mustapha N, Mohamed R, Bin Ahmadon MA, Yamaguchi S. Measuring student satisfaction based on analysis of physical parameters in smart classroom. In: 2024 12th International Conference on Information and Education Technology (ICIET); 2024 Mar 18–20; Yamaguchi, Japan. doi:10.1109/ICIET60671.2024.10542750.
10. Gu F, Chung MH, Chignell M, Valaee S, Zhou B, Liu X. A survey on deep learning for human activity recognition. ACM Comput Surv. 2022;54(8):1–34. doi:10.1145/3472290.
11. Dang LM, Min K, Wang H, Jalil Piran M, Lee CH, Moon H. Sensor-based and vision-based human activity recognition: a comprehensive survey. Pattern Recognit. 2020;108:107561. doi:10.1016/j.patcog.2020.107561.
12. Arshad MH, Bilal M, Gani A. Human activity recognition: review, taxonomy and open challenges. Sensors. 2022;22(17):6463. doi:10.3390/s22176463.
13. Serpush F, Menhaj MB, Masoumi B, Karasfi B. Wearable sensor-based human activity recognition in the smart healthcare system. Comput Intell Neurosci. 2022;2022:1391906. doi:10.1155/2022/1391906.
14. Demrozi F, Pravadelli G, Bihorac A, Rashidi P. Human activity recognition using inertial, physiological and environmental sensors: a comprehensive survey. IEEE Access. 2020;8:210816–36. doi:10.1109/access.2020.3037715.
15. Chen K, Zhang D, Yao L, Guo B, Yu Z, Liu Y. Deep learning for sensor-based human activity recognition. ACM Comput Surv. 2022;54(4):1–40. doi:10.1145/3447744.
16. Ramanujam E, Perumal T, Padmavathi S. Human activity recognition with smartphone and wearable sensors using deep learning techniques: a review. IEEE Sens J. 2021;21(12):13029–40. doi:10.1109/JSEN.2021.3069927.
17. Bank D, Koenigstein N, Giryes R. Autoencoders. In: Machine learning for data science handbook. Berlin/Heidelberg, Germany: Springer; 2023. p. 353–74. doi:10.1007/978-3-031-24628-9_16.
18. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM. 2020;63(11):139–44. doi:10.1145/3422622.
19. Zhang S, Li Y, Zhang S, Shahabi F, Xia S, Deng Y, et al. Deep learning in human activity recognition with wearable sensors: a review on advances. Sensors. 2022;22(4):1476. doi:10.3390/s22041476.
20. Sankar S, Srinivasan P, Saravanakumar R. Internet of Things based ambient assisted living for elderly people health monitoring. Rese Jour Pharm and Technol. 2018;11(9):3900. doi:10.5958/0974-360x.2018.00715.1.
21. Zdravevski E, Lameski P, Trajkovik V, Kulakov A, Chorbev I, Goleva R, et al. Improving activity recognition accuracy in ambient-assisted living systems by automated feature engineering. IEEE Access. 2017;5:5262–80. doi:10.1109/access.2017.2684913.
22. El murabet A, Abtoy A, Touhafi A, Tahiri A. Ambient Assisted living system's models and architectures: a survey of the state of the art. J King Saud Univ Comput Inf Sci. 2020;32(1):1–10. doi:10.1016/j.jksuci.2018.04.009.
23. Mokhtari G, Aminikhanghahi S, Zhang Q, Cook DJ. Fall detection in smart home environments using UWB sensors and unsupervised change detection. J Reliab Intell Environ. 2018;4(3):131–9. doi:10.1007/s40860-018-0065-2.
24. Ali Hamad R, Yang L, Woo WL, Wei B. Joint learning of temporal models to handle imbalanced data for human activity recognition. Appl Sci. 2020;10(15):5293. doi:10.3390/app10155293.
25. Javed AR, Fahad LG, Ahmad Farhan A, Abbas S, Srivastava G, Parizi RM, et al. Automated cognitive health assessment in smart homes using machine learning. Sustain Cities Soc. 2021;65:102572. doi:10.1016/j.scs.2020.102572.
26. Gjoreski M, Gjoreski H, Luštrek M, Gams M. How accurately can your wrist device recognize daily activities and detect falls? Sensors. 2016;16(6):800. doi:10.3390/s16060800.
27. Plasqui G. Smart approaches for assessing free-living energy expenditure following identification of types of physical activity. Obes Rev. 2017;18(Suppl 1):50–5. doi:10.1111/obr.12506.

28. Ogbuabor G, La R. Human activity recognition for healthcare using smartphones. In: Proceedings of the 2018 10th International Conference on Machine Learning and Computing; 2018 Feb 26–28; Macau, China. doi:10.1145/3195106.3195157.
29. Morales J, Akopian D. Physical activity recognition by smartphones, a survey. Biocybern Biomed Eng. 2017;37(3):388–400. doi:10.1016/j.bbe.2017.04.004.
30. Qi J, Yang P, Hanneghan M, Tang S, Zhou B. A hybrid hierarchical framework for gym physical activity recognition and measurement using wearable sensors. IEEE Internet Things J. 2019;6(2):1384–93. doi:10.1109/JIOT.2018.2846359.
31. Fosstveit SH, Lohne-Seiler H, Feron J, Lucas SJE, Ivarsson A, Berntsen S. The intensity paradox: a systematic review and meta-analysis of its impact on the cardiorespiratory fitness of older adults. Scand J Med Sci Sports. 2024;34(2):e14573. doi:10.1111/sms.14573.
32. Mahindru A, Patil P, Agrawal V. Role of physical activity on mental health and well-being: a review. Cureus. 2023;15(1):e33475. doi:10.7759/cureus.33475.
33. Alo UR, Nweke HF, Teh YW, Murtaza G. Smartphone motion sensor-based complex human activity identification using deep stacked autoencoder algorithm for enhanced smart healthcare system. Sensors. 2020;20(21):6300. doi:10.3390/s20216300.
34. Müller PN, Müller AJ, Achenbach P, Göbel S. IMU-based fitness activity recognition using CNNs for time series classification. Sensors. 2024;24(3):742. doi:10.3390/s24030742.
35. Cust EE, Sweeting AJ, Ball K, Robertson S. Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. J Sports Sci. 2019;37(5):568–600. doi:10.1080/02640414.2018.1521769.
36. Kranzinger C, Bernhart S, Kremser W, Venek V, Rieser H, Mayr S, et al. Classification of human motion data based on inertial measurement units in sports: a scoping review. Appl Sci. 2023;13(15):8684. doi:10.3390/app13158684.
37. Brock H, Ohgi Y. Assessing motion style errors in ski jumping using inertial sensor devices. IEEE Sens J. 2017;17(12):3794–804. doi:10.1109/JSEN.2017.2699162.
38. Brock H, Ohgi Y, Lee J. Learning to judge like a human: convolutional networks for classification of ski jumping errors. In: Proceedings of the 2017 ACM International Symposium on Wearable Computers; 2017 Sep 11–15; Maui, HI, USA. doi:10.1145/3123021.3123038.
39. Anand A, Sharma M, Srivastava R, Kaligounder L, Prakash D. Wearable motion sensor based analysis of swing sports. In: 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA); 2017 Dec 18–21; Cancun, Mexico. doi:10.1109/ICMLA.2017.0-149.
40. Giménez-Egido JM, Ortega E, Verdu-Conesa I, Cejudo A, Torres-Luque G. Using smart sensors to monitor physical activity and technical-tactical actions in junior tennis players. Int J Environ Res Public Health. 2020;17(3):1068. doi:10.3390/ijerph17031068.
41. Mat Sanusi KA, Mitri DD, Limbu B, Klemke R. Table tennis tutor: forehand strokes classification based on multimodal data and neural networks. Sensors. 2021;21(9):3121. doi:10.3390/s21093121.
42. Buckley C, O'Reilly MA, Whelan D, Farrell AV, Clark L, Longo V, et al. Binary classification of running fatigue using a single inertial measurement unit. In: 2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN); 2017 May 9–12; Eindhoven, The Netherlands. doi:10.1109/BSN.2017.7936040.
43. Liu Q, Mo S, Cheung VCK, Cheung BMF, Wang S, Chan PPK, et al. Classification of runners' performance levels with concurrent prediction of biomechanical parameters using data from inertial measurement units. J Biomech. 2020;112:110072. doi:10.1016/j.jbiomech.2020.110072.
44. Mo LF, Zeng LJ. Running gait pattern recognition based on cross-correlation analysis of single acceleration sensor. Math Biosci Eng. 2019;16(6):6242–56. doi:10.3934/mbe.2019311.
45. Worsey MTO, Espinosa HG, Shepherd JB, Thiel DV. An evaluation of wearable inertial sensor configuration and supervised machine learning models for automatic punch classification in Boxing. IoT. 2020;1(2):360–81. doi:10.3390/iot1020021.

46. Jayakumar B, Govindarajan N. Multi-sensor fusion based optimized deep convolutional neural network for boxing punch activity recognition. Proc Inst Mech Eng Part P J Phys Eng Technol. 2024;2024:17543371241237085. doi:10.1177/17543371241237085.
47. Jiao L, Wu H, Bie R, Umek A, Kos A. Multi-sensor golf swing classification using deep CNN. Procedia Comput Sci. 2018;129:59–65. doi:10.1016/j.procs.2018.03.046.
48. Kim M, Park S. Golf swing segmentation from a single IMU using machine learning. Sensors. 2020;20(16):4466. doi:10.3390/s20164466.
49. Jiao L, Wu H, Bie R, Umek A, Kos A. Towards real-time multi-sensor golf swing classification using deep CNNs. J Database Manag. 2018;29(3):17–42. doi:10.4018/jdm.2018070102.
50. Kautz T, Groh BH, Hannink J, Jensen U, Strubberg H, Eskofier BM. Activity recognition in beach volleyball using a Deep Convolutional Neural Network. Data Min Knowl Discov. 2017;31(6):1678–705. doi:10.1007/s10618-017-0495-0.
51. Haider F, Salim FA, Postma DBW, van Delden R, Reidsma D, van Beijnum BJ, et al. A super-bagging method for volleyball action recognition using wearable sensors. Multimodal Technol Interact. 2020;4(2):33. doi:10.3390/mti4020033.
52. Haider F, Salim F, Naghashi V, Tasdemir SBY, Tengiz I, Cengiz K, et al. Evaluation of dominant and non-dominant hand movements for volleyball action modelling. In: Adjunct of the 2019 International Conference on Multimodal Interaction; 2019 Oct 14–18; Suzhou, China. doi:10.1145/3351529.3360651.
53. Delhaye E, Bouvet A, Nicolas G, Vilas-Boas JP, Bideau B, Bideau N. Automatic swimming activity recognition and lap time assessment based on a single IMU: a deep learning approach. Sensors. 2022;22(15):5786. doi:10.3390/s22155786.
54. Wang Z, Shi X, Wang J, Gao F, Li J, Guo M, et al. Swimming motion analysis and posture recognition based on wearable inertial sensors. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC); 2019 Oct 6–9; Bari, Italy. doi:10.1109/SMC.2019.8913847.
55. Chen L, Hu D. An effective swimming stroke recognition system utilizing deep learning based on inertial measurement units. Adv Robot. 2023;37(7):467–79. doi:10.1080/01691864.2022.2160274.
56. Rana M, Mittal V. Analysis of front leg kinematics of cricket bowler using wearable sensors and machine learning. IEEE Sens J. 2022;22(22):22053–61. doi:10.1109/JSEN.2022.3205363.
57. Khan A, Nabila FH, Mohiuddin M, Mollah M, Alam A, Tanzim Reza M. An approach to classify the shot selection by batsmen in cricket matches using deep neural network on image data. In: 2022 25th International Conference on Computer and Information Technology (ICCIT); 2022 Dec 17–19; Cox's Bazar, Bangladesh. doi:10.1109/ICCIT57492.2022.10055811.
58. Salman M, Qaisar S, Qamar AM. Classification and legality analysis of bowling action in the game of cricket. Data Min Knowl Discov. 2017;31(6):1706–34. doi:10.1007/s10618-017-0511-4.
59. Abdullah MA, Ibrahim MAR, Shapiee MNA, Zakaria MA, Mohd Razman MA, Muazu Musa R, et al. The classification of skateboarding tricks via transfer learning pipelines. PeerJ Comput Sci. 2021;7:e680. doi:10.7717/peerj-cs.680.
60. Hollaus B, Westenberger E, Kreiner J, Freitas G, Fresen L. Motion based trick classification in skateboarding using machine learning. In: 2023 World Symposium on Digital Intelligence for Systems and Machines (DISA); 2023 Sep 21–22; Košice, Slovakia. doi:10.1109/DISA59116.2023.10308913.
61. Groh BH, Fleckenstein M, Kautz T, Eskofier BM. Classification and visualization of skateboard tricks using wearable sensors. Pervasive Mob Comput. 2017;40:42–55. doi:10.1016/j.pmcj.2017.05.007.
62. Saba T, Rehman A, Latif R, Fati SM, Raza M, Sharif M. Suspicious activity recognition using proposed deep L4-branched-ActionNet with entropy coded ant colony system optimization. IEEE Access. 2021;9:89181–97. doi:10.1109/ACCESS.2021.3091081.
63. Riboni D, Murru F. Unsupervised recognition of multi-resident activities in smart-homes. IEEE Access. 2020;8:201985–94. doi:10.1109/access.2020.3036226.
64. Du Y, Lim Y, Tan Y. A novel human activity recognition and prediction in smart home based on interaction. Sensors. 2019;19(20):4474. doi:10.3390/s19204474.

65. Chen L, Liu X, Peng L, Wu M. Deep learning based multimodal complex human activity recognition using wearable devices. Appl Intell. 2021;51(6):4029–42. doi:10.1007/s10489-020-02005-7.
66. Dernbach S, Das B, Krishnan NC, Thomas BL, Cook DJ. Simple and complex activity recognition through smart phones. In: 2012 Eighth International Conference on Intelligent Environments; 2012 Jun 26–29; Guanajuato, Mexico. doi:10.1109/IE.2012.39.
67. Liu L, Peng Y, Wang S, Liu M, Huang Z. Complex activity recognition using time series pattern dictionary learned from ubiquitous sensors. Inf Sci. 2016;340:41–57. doi:10.1016/j.ins.2016.01.020.
68. Liu L, Peng Y, Liu M, Huang Z. Sensor-based human activity recognition system with a multilayered model using time series shapelets. Knowl Based Syst. 2015;90:138–52. doi:10.1016/j.knosys.2015.09.024.
69. Lv M, Chen L, Chen T, Chen G. Bi-view semi-supervised learning based semantic human activity recognition using accelerometers. IEEE Trans Mob Comput. 2018;17(9):1991–2001. doi:10.1109/TMC.2018.2793913.
70. Peng L, Chen L, Wu M, Chen G. Complex activity recognition using acceleration, vital sign, and location data. IEEE Trans Mob Comput. 2019;18(7):1488–98. doi:10.1109/TMC.2018.2863292.
71. Peng L, Chen L, Wu X, Guo H, Chen G. Hierarchical complex activity representation and recognition using topic model and classifier level fusion. IEEE Trans Biomed Eng. 2017;64(6):1369–79. doi:10.1109/TBME.2016.2604856.
72. Naccarelli R, Casaccia S, Revel GM. The problem of monitoring activities of older people in multi-resident scenarios: an innovative and non-invasive measurement system based on wearables and PIR sensors. Sensors. 2022;22(9):3472. doi:10.3390/s22093472.
73. Shiri FMP, Perumal T, Mustapha N, Mohamed R, Ahmadon MA, Yamaguchi S. A survey on multi-resident activity recognition in smart environments. Evol Inf Commun Comput Syst. 2023;4(1):12–27.
74. Arrotta L, Bettini C, Civitarese G. MICAR: multi-inhabitant context-aware activity recognition in home environments. Distrib Parallel Databases. 2022:1–32. doi:10.1007/s10619-022-07403-z.
75. Lapointe P, Chapron K, Bouchard K, Gaboury' S. A new device to track and identify people in a multi-residents context. Procedia Comput Sci. 2020;170:403–10. doi:10.1016/j.procs.2020.03.082.
76. Wang T, Cook DJ. sMRT: multi-resident tracking in smart homes with sensor vectorization. IEEE Trans Pattern Anal Mach Intell. 2021;43(8):2809–21. doi:10.1109/tpami.2020.2973571.
77. Wang T, Cook DJ. Multi-person activity recognition in continuously monitored smart homes. IEEE Trans Emerg Top Comput. 2022;10(2):1130–41. doi:10.1109/tetc.2021.3072980.
78. Mohamed R, Zainudin MNS, Perumal T, Muhammad S. Adaptive profiling model for multiple residents activity recognition analysis using spatio-temporal information in smart home. In: Proceedings of the 8th International Conference on Computational Science and Technology; 2021 Aug 28–29; Labuan, Malaysia. doi:10.1007/978-981-16-8515-6_60.
79. Benmansour A, Bouchachia A, Feham M. Modeling interaction in multi-resident activities. Neurocomputing. 2017;230:133–42. doi:10.1016/j.neucom.2016.05.110.
80. Rossi S, Acampora G, Staffa M. Working together: a DBN approach for individual and group activity recognition. J Ambient Intell Humaniz Comput. 2020;11(12):6007–19. doi:10.1007/s12652-020-01851-0.
81. Qiu S, Zhao H, Jiang N, Wang Z, Liu L, An Y, et al. Multi-sensor information fusion based on machine learning for real applications in human activity recognition: state-of-the-art and research challenges. Inf Fusion. 2022;80:241–65. doi:10.1016/j.inffus.2021.11.006.
82. Tao W, Lai ZH, Leu MC, Yin Z. Worker activity recognition in smart manufacturing using IMU and sEMG signals with convolutional neural networks. Procedia Manuf. 2018;26:1159–66. doi:10.1016/j.promfg.2018.07.152.
83. Graña M, Aguilar-Moreno M, De Lope Asiain J, Araquistain IB, Garmendia X. Improved activity recognition combining inertial motion sensors and electroencephalogram signals. Int J Neural Syst. 2020;30(10):2050053. doi:10.1142/S0129065720500537.
84. Salehzadeh A, Calitz AP, Greyling J. Human activity recognition using deep electroencephalography learning. Biomed Signal Process Control. 2020;62:102094. doi:10.1016/j.bspc.2020.102094.
85. Butt FS, La Blunda L, Wagner MF, Schäfer J, Medina-Bulo I, Gómez-Ullate D. Fall detection from electrocardiogram (ECG) signals and classification by deep transfer learning. Information. 2021;12(2):63. doi:10.3390/info12020063.

86. Thilagaraj M, Arunkumar N, Ramkumar S, Hariharasitaraman S. Electrooculogram signal identification for elderly disabled using Elman network. Microprocess Microsyst. 2021;82:103811. doi:10.1016/j.micpro.2020.103811.
87. Almanifi ORA, Mohd Khairuddin I, Mohd Razman MA, Musa RM, Abdul Majeed APP. Human activity recognition based on wrist PPG via the ensemble method. ICT Express. 2022;8(4):513–7. doi:10.1016/j.icte.2022.03.006.
88. Mekruksavanich S, Jitpattanakul A. Sport-related activity recognition from wearable sensors using bidirectional GRU network. Intell Autom Soft Comput. 2022;34(3):1907–25. doi:10.32604/iasc.2022.027233.
89. Otebolaku A, Enamamu T, Alfoudi A, Ikpehai A, Marchang J, Lee GM. Deep sensing: inertial and ambient sensing for activity context recognition using deep convolutional neural networks. Sensors. 2020;20(13):3803. doi:10.3390/s20133803.
90. Tamura T. Wearable inertial sensors and their applications. In: Wearable sensors. Amsterdam, The Netherlands: Elsevier; 2014. p. 85–104. doi:10.1016/b978-0-12-418662-0.00024-6.
91. Gohel V, Mehendale N. Review on electromyography signal acquisition and processing. Biophys Rev. 2020;12(6):1361–7. doi:10.1007/s12551-020-00770-w.
92. Soufineyestani M, Dowling D, Khan A. Electroencephalography (EEG) technology applications and available devices. Appl Sci. 2020;10(21):7453. doi:10.3390/app10217453.
93. Dudeja U, Dubey SK. Decoding emotions: emotion classification from EEG brain signals using AI. In: 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON); 2023 Dec 1–3; Gautam Buddha Nagar, India. doi:10.1109/UPCON59197.2023.10434423.
94. Rashkovska A, Depolli M, Tomašić I, Avbelj V, Trobec R. Medical-grade ECG sensor for long-term monitoring. Sensors. 2020;20(6):1695. doi:10.3390/s20061695.
95. Ma Y, Liu Q, Yang L. Exploring seafarers' workload recognition model with EEG, ECG and task scenarios' complexity: a bridge simulation study. J Mar Sci Eng. 2022;10(10):1438. doi:10.3390/jmse10101438.
96. Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: a review. Comput Methods Programs Biomed. 2018;161:1–13. doi:10.1016/j.cmpb.2018.04.005.
97. Li Y, Xu Z, Zhang Y, Cao Z, Chen H. Automatic sleep stage classification based on a two-channel electrooculogram and one-channel electromyogram. Physiol Meas. 2022;43(7):07NT02. doi:10.1088/1361-6579/ac6bdb.
98. Lin CT, Jiang WL, Chen SF, Huang KC, Liao LD. Design of a wearable eye-movement detection system based on electrooculography signals and its experimental validation. Biosensors. 2021;11(9):343. doi:10.3390/bios11090343.
99. Lu Y, Zhang C, Zhou BY, Gao XP, Lv Z. A dual model approach to EOG-based human activity recognition. Biomed Signal Process Control. 2018;45:50–7. doi:10.1016/j.bspc.2018.05.011.
100. Alian AA, Shelley KH. Photoplethysmography. Best Pract Res Clin Anaesthesiol. 2014;28(4):395–406. doi:10.1016/j.bpa.2014.08.006.
101. Alian AA, Shelley KH. Photoplethysmography: analysis of the pulse oximeter waveform. In: Monitoring technologies in acute care environments. Berlin/Heidelberg, Germany: Springer; 2013. p. 165–78. doi:10.1007/978-1-4614-8557-5_19.
102. Biagetti G, Crippa P, Falaschetti L, Orcioni S, Turchetti C. Human activity recognition using accelerometer and photoplethysmographic signals. In: Proceedings of the 9th KES International Conference on Intelligent Decision Technologies (KES-IDT 2017)—Part II; 2017 Jun 21–23; Vilamoura, Algarve, Portugal. doi:10.1007/978-3-319-59424-8_6.
103. Hnoohom N, Mekruksavanich S, Jitpattanakul A. Physical activity recognition based on deep learning using photoplethysmography and wearable inertial sensors. Electronics. 2023;12(3):693. doi:10.3390/electronics12030693.
104. Shokrollahi A, Persson JA, Malekian R, Sarkheyli-Hägele A, Karlsson F. Passive infrared sensor-based occupancy monitoring in smart buildings: a review of methodologies and machine learning approaches. Sensors. 2024;24(5):1533. doi:10.3390/s24051533.
105. Ramos RG, Domingo JD, Zalama E, Gómez-García-Bermejo J, López J. SDHAR-HOME: a sensor dataset for human activity recognition at home. Sensors. 2022;22(21):8109. doi:10.3390/s22218109.
106. Uddin MZ, Khaksar W, Torresen J. Ambient sensors for elderly care and independent living: a survey. Sensors. 2018;18(7):2027. doi:10.3390/s18072027.

107. Lombardi M, Vezzani R, Cucchiara R. Detection of human movements with pressure floor sensors. In: Image Analysis and Processing—ICIAP 2015; 2015 Sep 7–11; Genoa, Italy. doi:10.1007/978-3-319-23234-8_57.
108. van Kasteren T, Noulas A, Englebienne G, Kröse B. Accurate activity recognition in a home setting. In: Proceedings of the 10th International Conference on Ubiquitous Computing; 2008 Sep 21–24; Seoul, Republic of Korea. doi:10.1145/1409635.1409637.
109. Stiefmeier T, Roggen D, Ogris G, Lukowicz P, Tröster G. Wearable activity tracking in car manufacturing. IEEE Pervasive Comput. 2008;7(2):42–50. doi:10.1109/MPRV.2008.40.
110. Cook DJ, Schmitter-Edgecombe M. Assessing the quality of activities in a smart environment. Methods Inf Med. 2009;48(5):480–5. doi:10.3414/ME0592.
111. Roggen D, Calatroni A, Rossi M, Holleczek T, Förster K, Tröster G, et al. Collecting complex activity datasets in highly rich networked sensor environments. In: 2010 Seventh International Conference on Networked Sensing Systems (INSS); 2010 Jun 15–18; Kassel, German. doi:10.1109/INSS.2010.5573462.
112. Singla G, Cook DJ, Schmitter-Edgecombe M. Recognizing independent and joint activities among multiple residents in smart environments. J Ambient Intell Humaniz Comput. 2010;1(1):57–63. doi:10.1007/s12652-009-0007-1.
113. Kawaguchi N, Ogawa N, Iwasaki Y, Kaji K, Terada T, Murao K, et al. HASC Challenge: gathering large scale human activity corpus for the real-world activity understandings. In: Proceedings of the 2nd Augmented Human International Conference; 2011 Mar 13; Tokyo, Japan. doi:10.1145/1959826.1959853.
114. van Kasteren TLM, Englebienne G, Kröse BJA. Human activity recognition from wireless sensor network data: benchmark and software. In: Activity recognition in pervasive intelligent environments. Paris, France: Atlantis Press; 2011. p. 165–86. doi:10.2991/978-94-91216-05-3_8.
115. Reiss A, Stricker D. Introducing a new benchmarked dataset for activity monitoring. In: 2012 16th International Symposium on Wearable Computers; 2012 Jun 18–22; Newcastle, UK. doi:10.1109/ISWC.2012.13.
116. Cook DJ. Learning setting-generalized activity models for smart spaces. IEEE Intell Syst. 2010;2010(99):1. doi:10.1109/MIS.2010.112.
117. Zhang M, Sawchuk AA. USC-HAD: a daily activity dataset for ubiquitous activity recognition using wearable sensors. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing; 2012 Sep 5–8; Pittsburgh, PA, USA. doi:10.1145/2370216.2370438.
118. Baños O, Damas M, Pomares H, Rojas I, Tóth MA, Amft O. A benchmark dataset to evaluate sensor displacement in activity recognition. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing; 2012 Sep 5–8; Pittsburgh, PA, USA. doi:10.1145/2370216.2370437.
119. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL. A public domain dataset for human activity recognition using smartphones. In: European Symposium on Artificial Neural Networks; 2013 Apr 24–26; Bruges, Belgium.
120. Alemdar H, Durmaz Incel O, Ertan H, Ersoy C. ARAS human activity datasets in multiple homes with multiple residents. In: Proceedings of the ICTs for Improving Patients Rehabilitation Research Techniques; 2013 May 5; Venice, Italy. doi:10.4108/pervasivehealth.2013.252120.
121. Bruno B, Mastrogiovanni F, Sgorbissa A, Vernazza T, Zaccaria R. Analysis of human behavior recognition algorithms based on acceleration data. In: 2013 IEEE International Conference on Robotics and Automation; 2013 May 6–10; Karlsruhe, Germany. doi:10.1109/ICRA.2013.6630784.
122. Barshan B, Yüksek MC. Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. Comput J. 2014;57(11):1649–67. doi:10.1093/comjnl/bxt075.
123. Banos O, Villalonga C, Garcia R, Saez A, Damas M, Holgado-Terriza JA, et al. Design, implementation and validation of a novel open framework for agile development of mobile health applications. Biomed Eng Online. 2015;14(Suppl 2):S6. doi:10.1186/1475-925X-14-S2-S6.
124. Stisen A, Blunck H, Bhattacharya S, Prentow TS, Kjærgaard MB, Dey A, et al. Smart devices are different: assessing and MitigatingMobile sensing heterogeneities for activity recognition. In: Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems; 2015 Nov 1–4; Seoul, Republic of Korea. doi:10.1145/2809695.2809718.

125. Chen C, Jafari R, Kehtarnavaz N. UTD-MHAD: a multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In: 2015 IEEE International Conference on Image Processing (ICIP); 2015 Sep 27–30; Quebec City, QC, Canada. doi:10.1109/ICIP.2015.7350781.
126. Shoaib M, Bosch S, Incel OD, Scholten H, Havinga PJM. Complex human activity recognition using smartphone and wrist-worn motion sensors. Sensors. 2016;16(4):426. doi:10.3390/s16040426.
127. Sztyler T, Stuckenschmidt H. On-body localization of wearable devices: an investigation of position-aware activity recognition. In: 2016 IEEE International Conference on Pervasive Computing and Communications (PerCom); 2016 Mar 14–19; Sydney, NSW, Australia. doi:10.1109/PERCOM.2016.7456521.
128. Palumbo F, Gallicchio C, Pucci R, Micheli A. Human activity recognition using multisensor data fusion based on Reservoir Computing. J Ambient Intell Smart Environ. 2016;8(2):87–107. doi:10.3233/ais-160372.
129. Vavoulas G, Chatzaki C, Malliotakis T, Pediaditis M, Tsiknakis M. The MobiAct dataset: recognition of activities of daily living using smartphones. In: Proceedings of the International Conference on Information and Communication Technologies for Ageing Well and E-Health; 2016 Apr 21–22; Rome, Italy. doi:10.5220/0005792401430151.
130. Chatzaki C, Pediaditis M, Vavoulas G, Tsiknakis M. Human daily activity and fall recognition using a smartphone's acceleration sensor. In: Information and Communication Technologies for Ageing Well and e-Health: Second International Conference; 2016 Apr 21–22; Rome, Italy. doi:10.1007/978-3-319-62704-5_7.
131. Micucci D, Mobilio M, Napoletano P. UniMiB SHAR: a dataset for human activity recognition using acceleration data from smartphones. Appl Sci. 2017;7(10):1101. doi:10.3390/app7101101.
132. Vaizman Y, Ellis K, Lanckriet G. Recognizing detailed human context in the wild from smartphones and smartwatches. IEEE Pervasive Comput. 2017;16(4):62–74. doi:10.1109/MPRV.2017.3971131.
133. Lago P, Lang F, Roncancio C, Jiménez-Guarín C, Mateescu R, Bonnefond N. The ContextAct@A4H real-life dataset of daily-living activities. In: Modeling and Using Context: 10th International and Interdisciplinary Conference; 2017 Jun 20–23; Paris, France. doi:10.1007/978-3-319-57837-8_14.
134. Alshammari T, Alshammari N, Sedky M, Howard C. SIMADL: simulated activities of daily living dataset. Data. 2018;3(2):11. doi:10.3390/data3020011.
135. Saha SS, Rahman S, Rasna MJ, Mahfuzul Islam AKM, Rahman Ahad MA. DU-MD: an open-source human action dataset for ubiquitous wearable sensors. In: 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR); 2018 Jun 25–29; Kitakyushu, Japan. doi:10.1109/ICIEV.2018.8641051.
136. Chereshnev R, Kertész-Farkas A. HuGaDB: human gait database for activity recognition from wearable inertial sensor networks. In: Analysis of Images, Social Networks and Texts: 6th International Conference, AIST 2017; 2017 Jul 27–29; Moscow, Russia. doi:10.1007/978-3-319-73013-4_12.
137. Malekzadeh M, Clegg RG, Cavallaro A, Haddadi H. Protecting sensory data against sensitive inferences. In: Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems; 2018 Apr 23–26; Porto, Portugal. doi:10.1145/3195258.3195260.
138. Weiss GM, Yoneda K, Hayajneh T. Smartphone and smartwatch-based biometrics using activities of daily living. IEEE Access. 2019;7:133190–202. doi:10.1109/access.2019.2940729.
139. Cruciani F, Sun C, Zhang S, Nugent C, Li C, Song S, et al. A public domain dataset for human activity recognition in free-living conditions. In: 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI); 2019 Aug 19–23; Leicester, UK. doi:10.1109/smartworld-uic-atc-scalcom-iop-sci.2019.00071.
140. Bhat G, Tran N, Shill H, Ogras UY. W-HAR: an activity recognition dataset and framework using low-power wearable devices. Sensors. 2020;20(18):5356. doi:10.3390/s20185356.
141. Arrotta L, Bettini C, Civitarese G. The MARBLE dataset: multi-inhabitant activities of daily living combining wearable and environmental sensors data. In: Mobile and ubiquitous systems: computing, networking and services. Berlin/Heidelberg, Germany: Springer; 2022. p. 451–68. doi:10.1007/978-3-030-94822-1_25.
142. Liu H, Hartmann Y, Schultz T. CSL-SHARE: a multimodal wearable sensor-based human activity dataset. Front Comput Sci. 2021;3:759136. doi:10.3389/fcomp.2021.759136.

143. De-La-Hoz-Franco E, Bernal Monroy E, Ariza-Colpas P, Mendoza-Palechor F, Espinilla M. UJA Human Activity Recognition multi-occupancy dataset. In: Proceedings of the Annual Hawaii International Conference on System Sciences 2021; 2021 Jan 5–8; Maui, HI, USA. doi:10.24251/hicss.2021.236.
144. Logacjov A, Bach K, Kongsvold A, Bårdstu HB, Mork PJ. HARTH: a human activity recognition dataset for machine learning. Sensors. 2021;21(23):7853. doi:10.3390/s21237853.
145. Sikder N, Nahid AA. KU-HAR: an open dataset for heterogeneous human activity recognition. Pattern Recognit Lett. 2021;146:46–54. doi:10.1016/j.patrec.2021.02.024.
146. Ustad A, Logacjov A, Trollebø SØ, Thingstad P, Vereijken B, Bach K, et al. Validation of an activity type recognition model classifying daily physical behavior in older adults: the HAR70+ model. Sensors. 2023;23(5):2368. doi:10.3390/s23052368.
147. Mekruksavanich S, Jitpattanakul A. Deep convolutional neural network with RNNs for complex activity recognition using wrist-worn wearable sensor data. Electronics. 2021;10(14):1685. doi:10.3390/electronics10141685.
148. Li S, Tao Y, Tang E, Xie T, Chen R. A survey of field programmable gate array (FPGA)-based graph convolutional neural network accelerators: challenges and opportunities. PeerJ Comput Sci. 2022;8:e1166. doi:10.7717/peerj-cs.1166.
149. Shiri FM, Perumal T, Mustapha N, Mohamed R. A comprehensive overview and comparative analysis on deep learning models. J Artif Intell. 2024;6(1):301–60. doi:10.32604/jai.2024.054314.
150. Hasan MN, Ahmed T, Ashik M, Hasan MJ, Azmin T, Uddin J. An analysis of COVID-19 pandemic outbreak on economy using neural network and random forest. J Inf Syst Telecommun. 2023;11(42):163–75. doi:10.52547/jist.34246.11.42.163.
151. Ke KC, Huang MS. Quality prediction for injection molding by using a multilayer perceptron neural network. Polymers. 2020;12(8):1812. doi:10.3390/polym12081812.
152. Hara K, Saito D, Shouno H. Analysis of function of rectified linear unit used in deep learning. In: 2015 International Joint Conference on Neural Networks (IJCNN); 2015 Jul 12–17; Killarney, Ireland. doi:10.1109/IJCNN.2015.7280578.
153. Kılıçarslan S, Adem K, Çelik M. An overview of the activation functions used in deep learning algorithms. J New Results Sci. 2021;10(3):75–88. doi:10.54187/jnrs.1011739.
154. Nwankpa C, Ijomah W, Gachagan A, Marshall S. Activation functions: comparison of trends in practice and research for deep learning. arXiv:1811.03378. 2018.
155. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv:1412.6980. 2014.
156. Bottou L. Stochastic gradient descent tricks. In: Neural networks: tricks of the trade. 2nd ed. Berlin, Heidelberg: Springer; 2012. p. 421–36. doi:10.1007/978-3-642-35289-8_25.
157. Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. J Mach Learn Res. 2011;12(7):2121–59.
158. Dozat T. Incorporating nesterov momentum into Adam. In: ICLR 2016 Workshop Track; 2016 May 2–4; Caribe Hilton, San Juan, Puerto Rico.
159. Sarker IH. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci. 2021;2(6):420. doi:10.1007/s42979-021-00815-1.
160. Putra DS, Yulita IN. Multilayer perceptron for activity recognition using a batteryless wearable sensor. IOP Conf Ser Earth Environ Sci. 2019;248:012039. doi:10.1088/1755-1315/248/1/012039.
161. Bejani MM, Ghatee M. A systematic review on overfitting control in shallow and deep neural networks. Artif Intell Rev. 2021;54(8):6391–438. doi:10.1007/s10462-021-09975-1.
162. Nadia A, Lyazid S, Okba K, Abdelghani C. A CNN-MLP deep model for sensor-based human activity recognition. In: 2023 15th International Conference on Innovations in Information Technology (IIT); 2023 Nov 14–15; Al Ain, United Arab Emirates. doi:10.1109/IIT59782.2023.10366481.
163. Rustam F, Ahmad Reshi A, Ashraf I, Mehmood A, Ullah S, Khan DM, et al. Sensor-based human activity recognition using deep stacked multilayered perceptron model. IEEE Access. 2020;8:218898–910. doi:10.1109/access.2020.3041822.
164. Majidzadeh Gorjani O, Byrtus R, Dohnal J, Bilik P, Koziorek J, Martinek R. Human activity classification using multilayer perceptron. Sensors. 2021;21(18):6207. doi:10.3390/s21186207.

165. STMicroelectronics NV. Discovery kit for IoT node, multi-channel communication with STM32L4. Geneva, Switzerland: STMicroelectronics NV; 2019.
166. Natani A, Sharma A, Perumal T. Sequential neural networks for multi-resident activity recognition in ambient sensing smart homes. Appl Intell. 2021;51(8):6014–28. doi:10.1007/s10489-020-02134-z.
167. Shi S, Wang Y, Dong H, Gui G, Ohtsuki T. Smartphone-aided human activity recognition method using residual multi-layer perceptron. In: IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS); 2022 May 2–5; New York, NY, USA. doi:10.1109/INFOCOMWKSHPS54753.2022.9798274.
168. Hendrycks D, Gimpel K. Gaussian error linear units (GELUs). arXiv:1606.08415. 2016.
169. Mao Y, Yan L, Guo H, Hong Y, Huang X, Yuan Y. A hybrid human activity recognition method using an MLP neural network and Euler angle extraction based on IMU sensors. Appl Sci. 2023;13(18):10529. doi:10.3390/app131810529.
170. Wang S, Zhang L, Wang X, Huang W, Wu H, Song A. PatchHAR: a MLP-like architecture for efficient activity recognition using wearables. IEEE Trans Biom Behav Identity Sci. 2024;6(2):169–81. doi:10.1109/TBIOM.2024.3354261.
171. Miyoshi T, Koshino M, Nambo H. Applying MLP-mixer and gMLP to human activity recognition. Sensors. 2025;25(2):311. doi:10.3390/s25020311.
172. Tolstikhin IO, Houlsby N, Kolesnikov A, Beyer L, Zhai X, Unterthiner T, et al. MLP-mixer: an all-MLP architecture for vision. Adv Neural Inf Process Syst. 2021;34:24261–72.
173. Liu H, Dai Z, So D, Le QV. Pay attention to MLPs. Adv Neural Inf Process Syst. 2021;34:9204–15.
174. Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE Trans Neural Netw Learn Syst. 2022;33(12):6999–7019. doi:10.1109/TNNLS.2021.3084827.
175. Chen L, Li S, Bai Q, Yang J, Jiang S, Miao Y. Review of image classification algorithms based on convolutional neural networks. Remote Sens. 2021;13(22):4712. doi:10.3390/rs13224712.
176. Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE. A survey of deep neural network architectures and their applications. Neurocomputing. 2017;234:11–26. doi:10.1016/j.neucom.2016.12.038.
177. Zafar A, Aamir M, Mohd Nawi N, Arshad A, Riaz S, Alruban A, et al. A comparison of pooling methods for convolutional neural networks. Appl Sci. 2022;12(17):8643. doi:10.3390/app12178643.
178. Yu D, Wang H, Chen P, Wei Z. Mixed pooling for convolutional neural networks. In: Rough Sets and Knowledge Technology: 9th International Conference, RSKT 2014; 2014 Oct 24–26; Shanghai, China. doi:10.1007/978-3-319-11740-9_34.
179. Zeiler MD, Fergus R. Stochastic pooling for regularization of deep convolutional neural networks. arXiv:1301.3557. 2013.
180. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In: Computer Vision–ECCV 2014; 2014 Sep 6–12; Zurich, Switzerland. doi:10.1007/978-3-319-10578-9_23.
181. Gong Y, Wang L, Guo R, Lazebnik S. Multi-scale orderless pooling of deep convolutional activation features. In: Computer Vision–ECCV 2014; 2014 Sep 6–12; Zurich, Switzerland. doi:10.1007/978-3-319-10584-0_26.
182. Krichen M. Convolutional neural networks: a survey. Computers. 2023;12(8):151. doi:10.3390/computers12080151.
183. Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. In: Proceedings of the 30th International Conference on Machine Learning; 2013 Jun 16–21; Atlanta, GA, USA.
184. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: 2015 IEEE International Conference on Computer Vision (ICCV); 2015 Dec 7–13; Santiago, Chile. doi:10.1109/ICCV.2015.123.
185. Xu B, Wang N, Chen T, Li M. Empirical evaluation of rectified activations in convolutional network. arXiv:1505.00853. 2015.
186. Jin X, Xu C, Feng J, Wei Y, Xiong J, Yan S. Deep learning with S-shaped rectified linear activation units. Proc AAAI Conf Artif Intell. 2016;30(1):1737–43. doi:10.1609/aaai.v30i1.10287.
187. Clevert DA, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs). arXiv:1511.07289. 2015.

188. Dubey SR, Singh SK, Chaudhuri BB. Activation functions in deep learning: a comprehensive survey and benchmark. Neurocomputing. 2022;503:92–108. doi:10.1016/j.neucom.2022.06.111.
189. Mohmed G, Lotfi A, Pourabdollah A. Employing a deep convolutional neural network for human activity recognition based on binary ambient sensor data. In: Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments; 2020 Jun 30–Jul 3; Corfu, Greece. doi:10.1145/3389189.3397991.
190. Raj R, Kos A. An improved human activity recognition technique based on convolutional neural network. Sci Rep. 2023;13(1):22581. doi:10.1038/s41598-023-49739-1.
191. Ismail WN, Alsalamah HA, Hassan MM, Mohamed E. AUTO-HAR: an adaptive human activity recognition framework using an automated CNN architecture design. Heliyon. 2023;9(2):e13636. doi:10.1016/j.heliyon.2023.e13636.
192. Ali Imran H, Latif U. HHARNet: taking inspiration from inception and dense networks for human activity recognition using inertial sensors. In: 2020 IEEE 17th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET); 2020 Dec 14–16; Charlotte, NC, USA. doi:10.1109/honet50430.2020.9322655.
193. Arabzadeh S, Almasganj F, Ahmadi MM. CNN autoencoders for hierarchical feature extraction and fusion in multi-sensor human activity recognition. arXiv:2502.04489. 2025.
194. Cheng D, Zhang L, Qin L, Wang S, Wu H, Song A. MaskCAE: masked convolutional AutoEncoder via sensor data reconstruction for self-supervised human activity recognition. IEEE J Biomed Health Inform. 2024;28(5):2687–98. doi:10.1109/JBHI.2024.3373019.
195. Feng H, Shen Q, Song R, Shi L, Xu H. ATFA: adversarial Time-Frequency Attention network for sensor-based multimodal human activity recognition. Expert Syst Appl. 2024;236:121296. doi:10.1016/j.eswa.2023.121296.
196. Wang Y, Xu H, Zheng L, Zhao G, Liu Z, Zhou S, et al. A multidimensional parallel convolutional connected network based on multisource and multimodal sensor data for human activity recognition. IEEE Internet Things J. 2023;10(16):14873–85. doi:10.1109/JIOT.2023.3265937.
197. Qi W, Su H, Yang C, Ferrigno G, De Momi E, Aliverti A. A fast and robust deep convolutional neural networks for complex human activity recognition using smartphone. Sensors. 2019;19(17):3731. doi:10.3390/s19173731.
198. Zhang W, Zhu T, Yang C, Xiao J, Ning H. Sensors-based human activity recognition with convolutional neural network and attention mechanism. In: 2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS); 2020 Oct 16–18; Beijing, China. doi:10.1109/ICSESS49938.2020.9237720.
199. Cao J, Guo F, Lai X, Zhou Q, Dai J. A tree-structure convolutional neural network for temporal features exaction on sensor-based multi-resident activity recognition. In: Neural computing for advanced applications. Amsterdam, The Netherlands: Elsevier; 2020. p. 513–25. doi:10.1007/978-981-15-7670-6_43.
200. Bouchabou D, Nguyen SM, Lohr C, LeDuc B, Kanellos I. Fully convolutional network bootstrapped by word encoding and embedding for activity recognition in smart homes. In: Deep Learning for Human Activity Recognition: Second International Workshop; 2021 Jan 8; Kyoto, Japan. doi:10.1007/978-981-16-0575-8_9.
201. Ata MM, Francies ML, Mohamed MA. A robust optimized convolutional neural network model for human activity recognition using sensing devices. Concurr Comput. 2022;34(17):e6964. doi:10.1002/cpe.6964.
202. Ataseven B, Madani A, Semiz B, Gursoy ME. Physical activity recognition using deep transfer learning with convolutional neural networks. In: 2022 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech); 2022 Sep 12–15; Falerna, Italy. doi:10.1109/DASC/PiCom/CBDCom/Cy55231.2022.9928021.
203. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015 Jun 7–12; Boston, MA, USA. doi:10.1109/CVPR.2015.7298594.

204. Dahou A, Al-Qaness MAA, Elaziz MA, Helmi AM. MLCNNwav: multilevel convolutional neural network with wavelet transformations for sensor-based human activity recognition. IEEE Internet Things J. 2024;11(1):820–8. doi:10.1109/JIOT.2023.3286378.
205. Kobayashi S, Hasegawa T, Miyoshi T, Koshino M. MarNASNets: toward CNN model architectures specific to sensor-based human activity recognition. IEEE Sens J. 2023;23(16):18708–17. doi:10.1109/JSEN.2023.3292380.
206. Zoph B, Le QV. Neural architecture search with reinforcement learning. arXiv:1611.01578. 2016.
207. Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–23; Salt Lake City, UT, USA. doi:10.1109/CVPR.2018.00907.
208. Lafontaine V, Bouchard K, Maître J, Gaboury S. Denoising UWB radar data for human activity recognition using convolutional autoencoders. IEEE Access. 2023;11:81298–309. doi:10.1109/access.2023.3300224.
209. Cavagnaro M, Pisa S, Pittella E. Safety aspects of people exposed to ultra wideband radar fields. Int J Anntenas Propag. 2013;2013:291064. doi:10.1155/2013/291064.
210. Li Y, Yang G, Su Z, Li S, Wang Y. Human activity recognition based on multienvironment sensor data. Inf Fusion. 2023;91:47–63. doi:10.1016/j.inffus.2022.10.015.
211. Najeh H, Lohr C, Leduc B. Convolutional neural network bootstrapped by dynamic segmentation and stigmergy-based encoding for real-time human activity recognition in smart homes. Sensors. 2023;23(4):1969. doi:10.3390/s23041969.
212. Tan TH, Chang YL, Wu JR, Chen YF, Alkhaleefah M. Convolutional neural network with multihead attention for human activity recognition. IEEE Internet Things J. 2023;11(2):3032–43. doi:10.1109/JIOT.2023.3294421.
213. Lai YC, Kan YC, Hsu KC, Lin HC. Multiple inputs modeling of hybrid convolutional neural networks for human activity recognition. Biomed Signal Process Control. 2024;92:106034. doi:10.1016/j.bspc.2024.106034.
214. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–23; Salt Lake City, UT, USA. doi:10.1109/CVPR.2018.00745.
215. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27–30; Las Vegas, NV, USA. doi:10.1109/CVPR.2016.90.
216. Sezavar A, Atta R, Ghanbari M. DCapsNet: deep capsule network for human activity and gait recognition with smartphone sensors. Pattern Recognit. 2024;147:110054. doi:10.1016/j.patcog.2023.110054.
217. Hinton GE, Krizhevsky A, Wang SD. Transforming auto-encoders. In: Artificial Neural Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks; 2011 Jun 14–17; Espoo, Finland.
218. Kwabena Patrick M, Felix Adekoya A, Abra Mighty A, Edward BY. Capsule networks-a survey. J King Saud Univ Comput Inf Sci. 2022;34(1):1295–310. doi:10.1016/j.jksuci.2019.09.014.
219. Ryu S, Yun S, Lee S, Jeong IC. Exploring the possibility of photoplethysmography-based human activity recognition using convolutional neural networks. Sensors. 2024;24(5):1610. doi:10.3390/s24051610.
220. Yu X, Al-qaness MAA. ASK-HAR: attention-based multi-core selective kernel convolution network for human activity recognition. Measurement. 2025;242:115981. doi:10.1016/j.measurement.2024.115981.
221. Woo S, Park J, Lee JY, Kweon IS. CBAM: convolutional block attention module. In: Computer Vision–ECCV 2018; 2018 Sep 8–14; Munich, Germany. doi:10.1007/978-3-030-01234-2_1.
222. Abbaspour S, Fotouhi F, Sedaghatbaf A, Fotouhi H, Vahabi M, Linden M. A comparative analysis of hybrid deep learning models for human activity recognition. Sensors. 2020;20(19):5707. doi:10.3390/s20195707.
223. Kratzert F, Klotz D, Brenner C, Schulz K, Herrnegger M. Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrol Earth Syst Sci. 2018;22(11):6005–22. doi:10.5194/hess-22-6005-2018.
224. Khan IU, Afzal S, Lee JW. Human activity recognition via hybrid deep learning based model. Sensors. 2022;22(1):323. doi:10.3390/s22010323.
225. Liang JM, Chung PL, Ye YJ, Mishra S. Applying machine learning technologies based on historical activity features for multi-resident activity recognition. Sensors. 2021;21(7):2520. doi:10.3390/s21072520.
226. Apaydin H, Feizi H, Sattari MT, Colak MS, Shamshirband S, Chau KW. Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting. Water. 2020;12(5):1500. doi:10.3390/w12051500.

227. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. doi:10.1162/neco.1997.9.8.1735.
228. Graves A. Generating sequences with recurrent neural networks. arXiv:1308.0850. 2013.
229. Barot V, Kapadia V. Long short term memory neural network-based model construction and fne-tuning for air quality parameters prediction. Cybern Inf Technol. 2022;22(1):171–89. doi:10.2478/cait-2022-0011.
230. Minaee S, Azimi E, Abdolrashidi A. Deep-sentiment: sentiment analysis using ensemble of CNN and bi-LSTM models. arXiv:1904.04206. 2019.
231. Shiri FM, Ahmadi E, Rezaee M, Perumal T. Detection of student engagement in E-learning environments using EfficientnetV2-L together with RNN-based models. J Artif Intell. 2024;6(1):85–103. doi:10.32604/jai.2024.048911.
232. Fang W, Chen Y, Xue Q. Survey on research of RNN-based spatio-temporal sequence prediction algorithms. J Big Data. 2021;3(3):97–110. doi:10.32604/jbd.2021.016993.
233. Zhao F, Feng J, Zhao J, Yang W, Yan S. Robust LSTM-autoencoders for face de-occlusion in the wild. IEEE Trans Image Process. 2018;27(2):778–90. doi:10.1109/TIP.2017.2771408.
234. Li Y, Wang L. Human activity recognition based on residual network and BiLSTM. Sensors. 2022;22(2):635. doi:10.3390/s22020635.
235. Graves A, Liwicki M, Fernández S, Bertolami R, Bunke H, Schmidhuber J. A novel connectionist system for unconstrained handwriting recognition. IEEE Trans Pattern Anal Mach Intell. 2009;31(5):855–68. doi:10.1109/TPAMI.2008.137.
236. Aldhyani THH, Alkahtani H. A bidirectional long short-term memory model algorithm for predicting COVID-19 in gulf countries. Life. 2021;11(11):1118. doi:10.3390/life11111118.
237. Du Y, Lim Y, Tan Y. Activity prediction using LSTM in smart home. In: 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE); 2019 Oct 15–18; Osaka, Japan. doi:10.1109/gcce46687.2019.9015492.
238. Barut O, Zhou L, Luo Y. Multitask LSTM model for human activity recognition and intensity estimation using wearable sensor data. IEEE Internet Things J. 2020;7(9):8760–8. doi:10.1109/JIOT.2020.2996578.
239. Kandpal M, Sharma B, Barik RK, Chowdhury S, Patra SS, Ben Dhaou I. Human activity recognition in smart cities from smart watch data using LSTM recurrent neural networks. In: 2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC); 2023 Jan 23–25; Jeddah, Saudi Arabia. doi:10.1109/ICAISC56366.2023.10085688.
240. Alawneh L, Mohsen B, Al-Zinati M, Shatnawi A, Al-Ayyoub M. A comparison of unidirectional and bidirectional LSTM networks for human activity recognition. In: 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops); 2020 Mar 23–27; Austin, TX, USA. doi:10.1109/percomworkshops48775.2020.9156264.
241. Modukuri SV, Mogaparthi N, Burri S, Kalangi RK, Venkatrama PKS. Bi-LSTM based real-time human activity recognition from smartphone sensor data. In: 2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit); 2024 Sep 4–6; Greater Noida, India. doi:10.1109/GlobalAISummit62156.2024.10947907.
242. Mekruksavanich S, Jantawong P, Jitpattanakul A. LSTM-XGB: a new deep learning model for human activity recognition based on LSTM and XGBoost. In: 2022 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON); 2022 Jan 26–28; Chiang Rai, Thailand. doi:10.1109/ECTIDAMTNCON53731.2022.9720409.
243. Budiarso Z. Optimizing LSTM with grid search and regularization techniques to enhance accuracy in human activity recognition. J Appl Data Sci. 2024;5(4):2002–14. doi:10.47738/jads.v5i4.433.
244. Bijrothiya S, Soni V. An architecture for human activity recognition using TCN-Bi-LSTM HAR based on wearable sensor. Procedia Comput Sci. 2025;260:805–13. doi:10.1016/j.procs.2025.03.261.
245. Hu Y, Zhang XQ, Xu L, Feng XH, Tian Z, She W, et al. Harmonic loss function for sensor-based human activity recognition based on LSTM recurrent neural networks. IEEE Access. 2020;8:135617–27. doi:10.1109/access.2020.3003162.

246. Hu SG, Liu Y, Chen TP, Liu Z, Yu Q, Deng LJ, et al. Emulating the Ebbinghaus forgetting curve of the human brain with a NiO-based memristor. Appl Phys Lett. 2013;103(13):133701. doi:10.1063/1.4822124.
247. Liciotti D, Bernardini M, Romeo L, Frontoni E. A sequential deep learning application for recognising human activities in smart homes. Neurocomputing. 2020;396:501–13. doi:10.1016/j.neucom.2018.10.104.
248. Gajjala KS, Chakraborty B. Human activity recognition based on LSTM neural network optimized by PSO algorithm. In: 2021 IEEE 4th International Conference on Knowledge Innovation and Invention (ICKII); 2021 Jul 23–25; Taichung, Taiwan. doi:10.1109/ICKII51822.2021.9574788.
249. Wang D, Tan D, Liu L. Particle swarm optimization algorithm: an overview. Soft Comput. 2018;22(2):387–408. doi:10.1007/s00500-016-2474-6.
250. Ramos RG, Domingo JD, Zalama E, Gómez-García-Bermejo J. Daily human activity recognition using non-intrusive sensors. Sensors. 2021;21(16):5270. doi:10.3390/s21165270.
251. Thapa K, Md Abdhulla AI Z, Sung-Hyun Y. Adapted long short-term memory (LSTM) for concurrent\\ human activity recognition. Comput Mater Contin. 2021;69(2):1653–70. doi:10.32604/cmc.2021.015660.
252. Benhaili Z, Kabbaj I, Balouki Y, Moumoun L. Human activity recognition using stacked LSTM. In: Advances in information, communication and cybersecurity. Berlin/Heidelberg, Germany: Springer; 2022. p. 33–42. doi:10.1007/978-3-030-91738-8_4.
253. Cao J, Wang Y, Tao H, Guo X. Sensor-based human activity recognition using graph LSTM and multi-task classification model. ACM Trans Multimedia Comput Commun Appl. 2022;18(3s):1–19. doi:10.1145/3561387.
254. Mekruksavanich S, Jantawong P, Hnoohom N, Jitpattanakul A. Refined LSTM network for sensor-based human activity recognition in real world scenario. In: 2022 IEEE 13th International Conference on Software Engineering and Service Science (ICSESS); 2022 Oct 21–23; Beijing, China. doi:10.1109/ICSESS54813.2022.9930218.
255. Yang SH, Baek DG, Thapa K. Semi-supervised adversarial learning using LSTM for human activity recognition. Sensors. 2022;22(13):4755. doi:10.3390/s22134755.
256. Tehrani A, Yadollahzadeh-Tabari M, Zehtab-Salmasi A, Enayatifar R. Wearable sensor-based human activity recognition system employing Bi-LSTM algorithm. Comput J. 2024;67(3):961–75. doi:10.1093/comjnl/bxad035.
257. El Ghazi M, Aknin N. Optimizing deep LSTM model through hyperparameter tuning for sensor-based human activity recognition in smart home. Informatica. 2023;47(10):109–22. doi:10.31449/inf.v47i10.5268.
258. Cho K, van Merrienboer B, Bahdanau D, Bengio Y. On the properties of neural machine translation: encoder-decoder approaches. arXiv:1409.1259. 2014.
259. Shiri FM, Perumal T, Mustapha N, Mohamed R, Bin Ahmadon MA, Yamaguchi S. Recognition of student engagement and affective states using ConvNeXtlarge and ensemble GRU in E-learning. In: 2024 12th International Conference on Information and Education Technology (ICIET); 2024 Mar 18–20; Yamaguchi, Japan. doi:10.1109/ICIET60671.2024.10542707.
260. Dutta A, Kumar S, Basu M. A gated recurrent unit approach to Bitcoin price prediction. J Risk Financ Manag. 2020;13(2):23. doi:10.3390/jrfm13020023.
261. Sansano E, Montoliu R, Belmonte Fernández Ó. A study of deep neural networks for human activity recognition. Comput Intell. 2020;36(3):1113–39. doi:10.1111/coin.12318.
262. Ravanelli M, Brakel P, Omologo M, Bengio Y. Light gated recurrent units for speech recognition. IEEE Trans Emerg Top Comput Intell. 2018;2(2):92–102. doi:10.1109/TETCI.2017.2762739.
263. Chai C, Ren C, Yin C, Xu H, Meng Q, Teng J, et al. A multifeature fusion short-term traffic flow prediction model based on deep learnings. J Adv Transp. 2022;2022:1702766. doi:10.1155/2022/1702766.
264. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv:1409.0473. 2014.
265. Han Y, Liu Y, Huang Q, Zhang Y. SOC estimation for lithium-ion batteries based on BiGRU with SE attention and Savitzky-Golay filter. J Energy Storage. 2024;90:111930. doi:10.1016/j.est.2024.111930.
266. Kolkar R, Singh Tomar RP, Vasantha G. IoT-based human activity recognition models based on CNN, LSTM and GRU. In: 2022 IEEE Silchar Subsection Conference (SILCON); 2022 Nov 4–6; Silchar, India. doi:10.1109/SILCON55242.2022.10028803.

267. Zhao S, Wei H, Zhang K. Deep Bidirectional GRU network for human activity recognition using Wearable inertial sensors. In: 2022 3rd International Conference on Electronic Communication and Artificial Intelligence (IWECAI); 2022 Jan 14–16; Zhuhai, China. doi:10.1109/IWECAI55315.2022.00054.
268. Zhou J, Sun C, Jang K, Yang S, Kim Y. Human activity recognition based on continuous-wave radar and bidirectional gate recurrent unit. Electronics. 2023;12(19):4060. doi:10.3390/electronics12194060.
269. Kang H, Kim D, Toh KA. Human activity recognition through augmented WiFi CSI signals by lightweight attention-GRU. Sensors. 2025;25(5):1547. doi:10.3390/s25051547.
270. Haque MN, Tanjid Hasan Tonmoy M, Mahmud S, Ali AA, Asif Hossain Khan M, Shoyaib M. GRU-based attention mechanism for human activity recognition. In: 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT); 2019 May 3–5; Dhaka, Bangladesh. doi:10.1109/icasert.2019.8934659.
271. Natani A, Sharma A, Peruma T, Sukhavasi S. Deep learning for multi-resident activity recognition in ambient sensing smart homes. In: 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE); 2019 Oct 15–18; Osaka, Japan. doi:10.1109/gcce46687.2019.9015212.
272. Jian Q, Guo S, Chen P, Wu P, Cui G. A robust real-time human activity recognition method based on attention-augmented GRU. In: 2021 IEEE Radar Conference (RadarConf21); 2021 May 7–14; Atlanta, GA, USA. doi:10.1109/RadarConf2147009.2021.9455322.
273. Mekruksavanich S, Jitpattanakul A. RNN-based deep learning for physical activity recognition using smartwatch sensors: a case study of simple and complex activity recognition. Math Biosci Eng. 2022;19(6):5671–98. doi:10.3934/mbe.2022265.
274. Pan J, Hu Z, Yin S, Li M. GRU with dual attentions for sensor-based human activity recognition. Electronics. 2022;11(11):1797. doi:10.3390/electronics11111797.
275. Sun X, Xu H, Dong Z, Shi L, Liu Q, Li J, et al. CapsGaNet: deep neural network based on capsule and GRU for human activity recognition. IEEE Syst J. 2022;16(4):5845–55. doi:10.1109/JSYST.2022.3153503.
276. Tong L, Ma H, Lin Q, He J, Peng L. A novel deep learning Bi-GRU-I model for real-time human activity recognition using inertial sensors. IEEE Sens J. 2022;22(6):6164–74. doi:10.1109/JSEN.2022.3148431.
277. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27–30; Las Vegas, NV, USA. doi:10.1109/CVPR.2016.308.
278. Mohsen S. Recognition of human activity using GRU deep learning algorithm. Multimed Tools Appl. 2023;82(30):47733–49. doi:10.1007/s11042-023-15571-y.
279. Abraham S, James RK. Significance of handcrafted features in human activity recognition with attention-based RNN models. Int J Electr Comput Eng Syst. 2023;14(10):1151–63. doi:10.32985/ijeces.14.10.8.
280. Mim TR, Amatullah M, Afreen S, Abu Yousuf M, Uddin S, Alyami SA, et al. GRU-INC: an inception-attention based approach using GRU for human activity recognition. Expert Syst Appl. 2023;216:119419. doi:10.1016/j.eswa.2022.119419.
281. Fährmann D, Boutros F, Kubon P, Kirchbuchner F, Kuijper A, Damer N. Ubiquitous multi-occupant detection in smart environments. Neural Comput Appl. 2024;36(6):2941–60. doi:10.1007/s00521-023-09162-z.
282. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Advances in Neural Information Processing Systems 30 (NIPS 2017); 2017 Dec 4–9; Long Beach, CA, USA.
283. Ba JL, Kiros JR, Hinton GE. Layer normalization. arXiv:1607.06450. 2016.
284. Soydaner D. Attention mechanism in neural networks: where it comes and where it goes. Neural Comput Appl. 2022;34(16):13371–85. doi:10.1007/s00521-022-07366-3.
285. Lesani FS, Fatahi R. A review on transformer-based methods for human activity recognition. Int J Web Res. 2024;7(4):81–100.
286. Liu Y, Wu L. Intrusion detection model based on improved transformer. Appl Sci. 2023;13(10):6251. doi:10.3390/app13106251.
287. Chen Z, Liu R, Huang S, Guo Y, Ren Y. A survey of large-scale deep learning models in medicine and healthcare. Comput Model Eng Sci. 2025;144(1):37–81. doi:10.32604/cmes.2025.067809.

288. Chen D, Yongchareon S, Lai EMK, Yu J, Sheng QZ, Li Y. Transformer with bidirectional GRU for nonintrusive, sensor-based activity recognition in a multiresident environment. IEEE Internet Things J. 2022;9(23):23716–27. doi:10.1109/JIOT.2022.3190307.
289. Lee TH, Kim H, Lee D. Transformer based early classification for real-time human activity recognition in smart homes. In: Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing; 2023 Mar 27–31; Tallinn, Estonia. doi:10.1145/3555776.3577693.
290. Dirgová Luptáková I, Kubovčík M, Pospíchal J. Wearable sensor-based human activity recognition with transformer model. Sensors. 2022;22(5):1911. doi:10.3390/s22051911.
291. Li J, Yao L, Li B, Wang X, Sammut C. Multi-agent transformer networks for multimodal human activity recognition. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management; 2022 Oct 17–21; Atlanta, GA, USA. doi:10.1145/3511808.3557402.
292. Samoon S, Laghari G, Malkani YA, Ali Shah SA. HAR-AttenNet: multi-head transformer for precise human activity recognition using wearable devices. VAWKUM Trans Comput Sci. 2025;13(2):36–49. doi:10.21015/vtcs.v13i2.2179.
293. Xiao S, Wang S, Huang Z, Wang Y, Jiang H. Two-stream transformer network for sensor-based human activity recognition. Neurocomputing. 2022;512:253–68. doi:10.1016/j.neucom.2022.09.099.
294. Huang X, Zhang S. Human activity recognition based on transformer in smart home. In: Proceedings of the 2023 2nd Asia Conference on Algorithms, Computing and Machine Learning; 2023 Mar 17–19; Shanghai, China. doi:10.1145/3590003.3590100.
295. Saidani O, Alsafyani M, Alroobaea R, Alturki N, Jahangir R, Jamel L. An efficient human activity recognition using hybrid features and transformer model. IEEE Access. 2023;11:101373–86. doi:10.1109/access.2023.3314492.
296. Pramanik R, Sikdar R, Sarkar R. Transformer-based deep reverse attention network for multi-sensory human activity recognition. Eng Appl Artif Intell. 2023;122:106150. doi:10.1016/j.engappai.2023.106150.
297. Suh S, Rey VF, Lukowicz P. TASKED: transformer-based Adversarial learning for human activity recognition using wearable sensors via Self-KnowledgE Distillation. Knowl Based Syst. 2023;260:110143. doi:10.1016/j.knosys.2022.110143.
298. Luo F, Li A, Khan S, Wu K, Wang L. Bi-DeepViT: binarized transformer for efficient sensor-based human activity recognition. IEEE Trans Mob Comput. 2025;24(5):4419–33. doi:10.1109/TMC.2025.3526166.
299. Zhou D, Kang B, Jin X, Yang L, Lian X, Jiang Z, et al. Deepvit: towards deeper vision transformer. arXiv:2103.11886. 2021.
300. Qin H, Gong R, Liu X, Shen M, Wei Z, Yu F, et al. Forward and backward information retention for accurate binary neural networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA. doi:10.1109/CVPR42600.2020.00232.
301. Pham P, Abraham JA, Chung J. Training multi-bit quantized and binarized networks with a learnable symmetric quantizer. IEEE Access. 2021;9:47194–203. doi:10.1109/access.2021.3067889.
302. Muniasamy A, Asiri F. Advanced human activity recognition on wearables with multi-scale sparse attention. Multimed Tools Appl. 2025;84(30):36623–52. doi:10.1007/s11042-024-20475-6.
303. Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. Proc Mach Learn Res. 2019;97:6105–14.
304. Alqahtani N, Alam S, Aqeel I, Shuaib M, Mohsen Khormi I, Khan SB, et al. Deep belief networks (DBN) with IoT-based Alzheimer's disease detection and classification. Appl Sci. 2023;13(13):7833. doi:10.3390/app13137833.
305. Kale AP, Wahul RM, Patange AD, Soman R, Ostachowicz W. Development of deep belief network for tool faults recognition. Sensors. 2023;23(4):1872. doi:10.3390/s23041872.
306. Oukrich N, Cherraqi EB, Maach A, Elghanami D. Multi-resident activity recognition method based in deep belief network. J Artif Intell. 2018;11(2):71–8. doi:10.3923/jai.2018.71.78.
307. Pisner DA, Schnyer DM. Support vector machine. In: Machine learning. Amsterdam, The Netherlands: Elsevier; 2020. p. 101–21. doi:10.1016/b978-0-12-815739-8.00006-7.
308. Cilimkovic M. Neural networks and back propagation algorithm [master's thesis]. Dublin, Ireland: Institute of Technology; 2015.

309. Huang YP, Basanta H, Kuo HC, Chiao HT. Sensor-based detection of abnormal events for elderly people using deep belief networks. Int J Ad Hoc Ubiquitous Comput. 2020;33(1):36. doi:10.1504/ijahuc.2020.104714.
310. Kumar LMA, Murugan S. Design of cuckoo search optimization with deep belief network for human activity recognition and classification. Multimed Tools Appl. 2023;82(19):29823–41. doi:10.1007/s11042-023-14977-y.
311. Gandomi AH, Yang XS, Alavi AH. Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems. Eng Comput. 2013;29(1):17–35. doi:10.1007/s00366-011-0241-y.
312. Choudhury NA, Soni B. An efficient CNN-LSTM approach for smartphone sensor-based human activity recognition system. In: 2022 5th International Conference on Computational Intelligence and Networks (CINE); 2022 Dec 1–3; Bhubaneswar, India. doi:10.1109/CINE56307.2022.10037495.
313. Khalid HUR, Gorji A, Bourdoux A, Pollin S, Sahli H. Multi-view CNN-LSTM architecture for radar-based human activity recognition. IEEE Access. 2022;10:24509–19. doi:10.1109/access.2022.3150838.
314. Mohd Noor MH, Tan SY, Ab Wahab MN. Deep temporal conv-LSTM for activity recognition. Neural Process Lett. 2022;54(5):4027–49. doi:10.1007/s11063-022-10799-5.
315. Al Farshi Oman S, Jamil MN, Raju SMTU. BCL: a branched CNN-LSTM architecture for human activity recognition using smartphone sensors. In: 2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM); 2023 Jun 16–17; Gazipur, Bangladesh. doi:10.1109/NCIM59001.2023.10212972.
316. Nafea O, Abdul W, Muhammad G, Alsulaiman M. Sensor-based human activity recognition with spatio-temporal deep learning. Sensors. 2021;21(6):2141. doi:10.3390/s21062141.
317. Chandramouli NA, Natarajan S, Alharbi AH, Kannan S, Khafaga DS, Raju SK, et al. Enhanced human activity recognition in medical emergencies using a hybrid deep CNN and bi-directional LSTM model with wearable sensors. Sci Rep. 2024;14(1):30979. doi:10.1038/s41598-024-82045-y.
318. Mekruksavanich S, Tancharoen D, Jitpattanakul A. A hybrid deep neural network with attention mechanism for human activity recognition based on smartphone sensors. In: 2023 7th International Conference on Information Technology (InCIT); 2023 Nov 16–17; Chiang Rai, Thailand. doi:10.1109/InCIT60207.2023.10413113.
319. Dahal A, Moulik S. Multistream CNN-BiLSTM framework for enhanced human activity recognition leveraging physiological signal. IEEE Sens Lett. 2025;9(2):6002304. doi:10.1109/LSENS.2025.3526446.
320. Lu L, Zhang C, Cao K, Deng T, Yang Q. A multichannel CNN-GRU model for human activity recognition. IEEE Access. 2022;10:66797–810. doi:10.1109/access.2022.3185112.
321. Wang Y. Human activity recognition algorithm based on ACNN-GRU. In: 2023 Asia Conference on Advanced Robotics, Automation, and Control Engineering (ARACE); 2023 Aug 18–20; Chengdu, China. doi:10.1109/ARACE60380.2023.00016.
322. Anguita-Molina MÁ, Cardoso PJS, Rodrigues JMF, Medina-Quero J, Polo-Rodríguez A. Multioccupancy activity recognition based on deep learning models fusing UWB localization heatmaps and nearby-sensor interaction. IEEE Internet Things J. 2025;12(11):16037–52. doi:10.1109/jiot.2025.3531316.
323. Tao S, Zhao Z, Qin J, Ji C, Wang Z. Attention-based convolutional neural network and bidirectional gated recurrent unit for human activity recognition. In: 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE); 2020 Dec 25–27; Harbin, China. doi:10.1109/ICMCCE51767.2020.00249.
324. Xia K, Huang J, Wang H. LSTM-CNN architecture for human activity recognition. IEEE Access. 2020;8:56855–66. doi:10.1109/access.2020.2982225.
325. Dua N, Singh SN, Semwal VB. Multi-input CNN-GRU based human activity recognition using wearable sensors. Computing. 2021;103(7):1461–78. doi:10.1007/s00607-021-00928-8.
326. Kim YW, Joa KL, Jeong HY, Lee S. Wearable IMU-based human activity recognition algorithm for clinical balance assessment using 1D-CNN and GRU ensemble model. Sensors. 2021;21(22):7628. doi:10.3390/s21227628.
327. Lima CA, Ricci NA, Nogueira EC, Perracini MR. The Berg Balance Scale as a clinical screening tool to predict fall risk in older adults: a systematic review. Physiotherapy. 2018;104(4):383–94. doi:10.1016/j.physio.2018.02.002.
328. Agarwal M, Flach P. Activity recognition using conditional random field. In: Proceedings of the 2nd International Workshop on Sensor-Based Activity Recognition and Interaction; 2015 Jun 25–26; Rostock, Germany. doi:10.1145/2790044.2790045.

329. Challa SK, Kumar A, Semwal VB. A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data. Vis Comput. 2022;38(12):4095–109. doi:10.1007/s00371-021-02283-3.
330. Khatun MA, Yousuf MA, Ahmed S, Uddin MZ, Alyami SA, Al-Ashhab S, et al. Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor. IEEE J Transl Eng Health Med. 2022;10:2700316. doi:10.1109/JTEHM.2022.3177710.
331. Kim YW, Cho WH, Kim KS, Lee S. Oversampling technique-based data augmentation and 1D-CNN and bidirectional gru ensemble model for human activity recognition. J Mech Med Biol. 2022;22(9):2240048. doi:10.1142/s0219519422400486.
332. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321–57. doi:10.1613/jair.953.
333. Han H, Wang WY, Mao BH. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: Advances in intelligent computing. Berlin/Heidelberg, Germany: Springer; 2005. p. 878–87. doi:10.1007/11538059_91.
334. Menardi G, Torelli N. Training and assessing classification rules with imbalanced data. Data Min Knowl Discov. 2014;28(1):92–122. doi:10.1007/s10618-012-0295-5.
335. Perumal T, Ramanujam E, Suman S, Sharma A, Singhal H. Internet of Things centric-based multiactivity recognition in smart home environment. IEEE Internet Things J. 2023;10(2):1724–32. doi:10.1109/JIOT.2022.3209970.
336. Agac S, Durmaz Incel O. On the use of a convolutional block attention module in deep learning-based human activity recognition with motion sensors. Diagnostics. 2023;13(11):1861. doi:10.3390/diagnostics13111861.
337. Khodabandelou G, Moon H, Amirat Y, Mohammed S. A fuzzy convolutional attention-based GRU network for human activity recognition. Eng Appl Artif Intell. 2023;118:105702. doi:10.1016/j.engappai.2022.105702.
338. Huang X, Yuan Y, Chang C, Gao Y, Zheng C, Yan L. Human activity recognition method based on edge computing-assisted and GRU deep learning network. Appl Sci. 2023;13(16):9059. doi:10.3390/app13169059.
339. Choudhury NA, Soni B. An adaptive batch size-based-CNN-LSTM framework for human activity recognition in uncontrolled environment. IEEE Trans Ind Inform. 2023;19(10):10379–87. doi:10.1109/TII.2022.3229522.
340. Ali Imran H, Riaz Q, Hussain M, Tahir H, Arshad R. Smart-wearable sensors and CNN-BiGRU model: a powerful combination for human activity recognition. IEEE Sens J. 2024;24(2):1963–74. doi:10.1109/JSEN.2023.3338264.
341. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-ResNet and the impact of residual connections on learning. Proc AAAI Conf Artif Intell. 2017;31(1):4278–84. doi:10.1609/aaai.v31i1.11231.
342. Ramanujam E, Perumal T. MLMO-HSM: multi-label Multi-output Hybrid Sequential Model for multi-resident smart home activity recognition. J Ambient Intell Humaniz Comput. 2023;14(3):2313–25. doi:10.1007/s12652-022-04487-4.
343. Zhao Y, Wang X, Luo Y, Aslam MS. Research on human activity recognition algorithm based on LSTM-1DCNN. Comput Mater Contin. 2023;77(3):3325–47. doi:10.32604/cmc.2023.040528.
344. Praba RA, Suganthi L. HARNet: automatic recognition of human activity from mobile health data using CNN and transfer learning of LSTM with SVM. Automatika. 2024;65(1):167–78. doi:10.1080/00051144.2023.2290736.
345. Lalwani P, Ramasamy G. Human activity recognition using a multi-branched CNN-BiLSTM-BiGRU model. Appl Soft Comput. 2024;154:111344. doi:10.1016/j.asoc.2024.111344.
346. Thakur D, Dangi S, Lalwani P. A novel hybrid deep learning approach with GWO-WOA optimization technique for human activity recognition. Biomed Signal Process Control. 2025;99:106870. doi:10.1016/j.bspc.2024.106870.
347. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Adv Eng Softw. 2014;69:46–61. doi:10.1016/j.advengsoft.2013.12.007.
348. Mirjalili S, Lewis A. The whale optimization algorithm. Adv Eng Softw. 2016;95:51–67. doi:10.1016/j.advengsoft.2016.01.008.
349. Garcia-Gonzalez D, Rivero D, Fernandez-Blanco E, Luaces MR. A public domain dataset for real-life human activity recognition using smartphone sensors. Sensors. 2020;20(8):2200. doi:10.3390/s20082200.
350. McMahan B, Moore E, Ramage D, Hampson S, Arcas BA. Communication-efficient learning of deep networks from decentralized data. Proc Mach Learn Res. 2017;54:1273–82.

351. Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning. ACM Trans Intell Syst Technol. 2019;10(2):1–19. doi:10.1145/3298981.
352. Wen J, Zhang Z, Lan Y, Cui Z, Cai J, Zhang W. A survey on federated learning: challenges and applications. Int J Mach Learn Cybern. 2023;14(2):513–35. doi:10.1007/s13042-022-01647-y.
353. Nguyen DC, Ding M, Pathirana PN, Seneviratne A, Li J, Vincent Poor H. Federated learning for Internet of Things: a comprehensive survey. IEEE Commun Surv Tutor. 2021;23(3):1622–58. doi:10.1109/comst.2021.3075439.
354. Nguyen DC, Ding M, Pham QV, Pathirana PN, Le LB, Seneviratne A, et al. Federated learning meets blockchain in edge computing: opportunities and challenges. IEEE Internet Things J. 2021;8(16):12806–25. doi:10.1109/JIOT.2021.3072611.
355. Zhang K, Song X, Zhang C, Yu S. Challenges and future directions of secure federated learning: a survey. Front Comput Sci. 2021;16(5):165817. doi:10.1007/s11704-021-0598-z.
356. Mammen PM. Federated learning: opportunities and challenges. arXiv:2101.05428. 2021.
357. Sharma S, Xing C, Liu Y, Kang Y. Secure and efficient federated transfer learning. In: 2019 IEEE International Conference on Big Data (Big Data); 2019 Dec 9–12; Los Angeles, CA, USA. doi:10.1109/bigdata47090.2019.9006280.
358. Li Q, Wen Z, Wu Z, Hu S, Wang N, Li Y, et al. A survey on federated learning systems: vision, hype and reality for data privacy and protection. IEEE Trans Knowl Data Eng. 2023;35(4):3347–66. doi:10.1109/TKDE.2021.3124599.
359. Xie C, Koyejo S, Gupta I. Asynchronous federated optimization. arXiv:1903.03934. 2019.
360. Aouedi O, Sacco A, Khan LU, Nguyen DC, Guizani M. Federated learning for human activity recognition: overview, advances, and challenges. IEEE Open J Commun Soc. 2024;5:7341–67. doi:10.1109/ojcoms.2024.3484228.
361. Aouedi O, Piamrat K, Sûdholt M. HFedSNN: efficient hierarchical federated learning using spiking neural networks. In: Proceedings of the Int'l ACM Symposium on Mobility Management and Wireless Access; 2023 Oct 30–Nov 3; Montreal, QC, Canada. doi:10.1145/3616390.3618288.
362. Liu L, Zhang J, Song SH, Letaief KB. Client-edge-cloud hierarchical federated learning. In: ICC 2020—2020 IEEE International Conference on Communications (ICC); 2020 Jun 7–11; Dublin, Ireland. doi:10.1109/ICC40277.2020.9148862.
363. Kou WB, Lin Q, Tang M, Ye R, Wang S, Zhu G, et al. Fast-convergent and communication-alleviated heterogeneous hierarchical federated learning in autonomous driving. IEEE Trans Intell Transp Syst. 2025;26(7):10496–511. doi:10.1109/TITS.2025.3543235.
364. Li Y, Wang X, An L. Hierarchical clustering-based personalized federated learning for robust and fair human activity recognition. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2023;7(1):1–38. doi:10.1145/3580795.
365. Pei J, Liu W, Li J, Wang L, Liu C. A review of federated learning methods in heterogeneous scenarios. IEEE Trans Consum Electron. 2024;70(3):5983–99. doi:10.1109/TCE.2024.3385440.
366. Kairouz P, McMahan HB, Avent B, Bellet A, Bennis M, Nitin Bhagoji A, et al. Advances and open problems in federated learning. FNT Mach Learn. 2021;14(1–2):1–210. doi:10.1561/2200000083.
367. Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated optimization in heterogeneous networks. Proc Mach Learn Syst. 2020;2:429–50.
368. Wang H, Yurochkin M, Sun Y, Papailiopoulos D, Khazaeni Y. Federated learning with matched averaging. arXiv:2002.06440. 2020.
369. Karimireddy SP, Kale S, Mohri M, Reddi S, Stich S, Suresh AT. Scaffold: stochastic controlled averaging for federated learning. Proc Mach Learn Res. 2020;119:5132–43.
370. Quintana GI, Vancamberg L, Jugnon V, Mougeot M, Desolneux A. BN-SCAFFOLD: controlling the drift of batch normalization statistics in federated learning. arXiv:2410.03281. 2024.
371. Sozinov K, Vlassov V, Girdzijauskas S. Human activity recognition using federated learning. In: 2018 IEEE International Conference on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom); 2018 Dec 11–13; Melbourne, VIC, Australia. doi:10.1109/BDCloud.2018.00164.
372. Gosselin R, Vieu L, Loukil F, Benoit A. Privacy and security in federated learning: a survey. Appl Sci. 2022;12(19):9901. doi:10.3390/app12199901.

373. Wei K, Li J, Ding M, Ma C, Yang HH, Farokhi F, et al. Federated learning with differential privacy: algorithms and performance analysis. IEEE Trans Inf Forensics Secur. 2020;15:3454–69. doi:10.1109/TIFS.2020.2988575.
374. Mugunthan V, Polychroniadou A, Byrd D, Balch TH. SMPAI: secure multi-party computation for federated learning. In: Proceedings of the NeurIPS 2019 Workshop on Robust AI in Financial Services; 2019 Dec 13; Vancouver, BC, Canada.
375. Aouedi O, Piamrat K. F-BIDS: federated-blending based intrusion detection system. Pervasive Mob Comput. 2023;89:101750. doi:10.1016/j.pmcj.2023.101750.
376. Agrawal S, Sarkar S, Aouedi O, Yenduri G, Piamrat K, Alazab M, et al. Federated learning for intrusion detection system: concepts, challenges and future directions. Comput Commun. 2022;195:346–61. doi:10.1016/j.comcom.2022.09.012.
377. Brik B, Ksentini A, Bouaziz M. Federated learning for UAVs-enabled wireless networks: use cases, challenges, and open problems. IEEE Access. 2020;8:53841–9. doi:10.1109/access.2020.2981430.
378. Le DD, Tran AK, Pham TB, Huynh TN. A survey of model compression and its feedback mechanism in federated learning. In: Proceedings of the Fifth Workshop on Intelligent Cross-Data Analysis and Retrieval; 2024 Jun 10–14; Phuket, Thailand. doi:10.1145/3643488.3660293.
379. Wu W, He L, Lin W, Mao R, Maple C, Jarvis S. SAFA: a semi-asynchronous protocol for fast federated learning with low overhead. IEEE Trans Comput. 2020;70(5):655–68. doi:10.1109/TC.2020.2994391.
380. Cheng D, Zhang L, Bu C, Wang X, Wu H, Song A. ProtoHAR: prototype guided personalized federated learning for human activity recognition. IEEE J Biomed Health Inform. 2023;27(8):3900–11. doi:10.1109/JBHI.2023.3275438.
381. Ek S, Portet F, Lalanda P, Vega G. Evaluation of federated learning aggregation algorithms: application to human activity recognition. In: Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers; 2020 Sep 12–17; Virtual. doi:10.1145/3410530.3414321.
382. Arivazhagan MG, Aggarwal V, Singh AK, Choudhary S. Federated learning with personalization layers. arXiv:1912.00818. 2019.
383. Kabir MR, Borshon RH, Khan R. Federated learning for human activity recognition: balancing privacy, efficiency, and accuracy through innovative aggregation techniques. Array. 2025;27:100462. doi:10.1016/j.array.2025.100462.
384. Ye X, Sakurai K, Nair NC, Wang KI. Machine learning techniques for sensor-based human activity recognition with data heterogeneity—a review. Sensors. 2024;24(24):7975. doi:10.3390/s24247975.
385. Anicai C, Shakir MZ. Federated learning and genetic mutation for multi-resident activity recognition. In: Proceedings of the 2023 IEEE 19th International Conference on e-Science (e-Science); 2023 Oct 9–13; Limassol, Cyprus. doi:10.1109/e-Science58273.2023.10254878.
386. Tu L, Ouyang X, Zhou J, He Y, Xing G. FedDL: federated learning via dynamic layer sharing for human activity recognition. In: Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems; 2021 Nov 15–17; Coimbra, Portugal. doi:10.1145/3485730.3485946.
387. Yu H, Chen Z, Zhang X, Chen X, Zhuang F, Xiong H, et al. FedHAR: semi-supervised online learning for personalized federated human activity recognition. IEEE Trans Mob Comput. 2023;22(6):3318–32. doi:10.1109/TMC.2021.3136853.
388. Wang P, Ouyang T, Wu Q, Huang Q, Gong J, Chen X. Hydra: hybrid-model federated learning for human activity recognition on heterogeneous devices. J Syst Archit. 2024;147:103052. doi:10.1016/j.sysarc.2023.103052.
389. Teerapittayanon S, McDanel B, Kung HT. BranchyNet: fast inference via early exiting from deep neural networks. In: 2016 23rd International Conference on Pattern Recognition (ICPR); 2016 Dec 4–8; Cancun, Mexico. doi:10.1109/ICPR.2016.7900006.
390. Khan AR, Manzoor HU, Ayaz F, Imran MA, Zoha A. A privacy and energy-aware federated framework for human activity recognition. Sensors. 2023;23(23):9339. doi:10.3390/s23239339.
391. Shaik T, Tao X, Higgins N, Gururajan R, Li Y, Zhou X, et al. FedStack: personalized activity monitoring using stacked federated learning. Knowl Based Syst. 2022;257:109929. doi:10.1016/j.knosys.2022.109929.

392. Presotto R, Civitarese G, Bettini C. Semi-supervised and personalized federated activity recognition based on active learning and label propagation. Pers Ubiquitous Comput. 2022;26(5):1281–98. doi:10.1007/s00779-022-01688-8.
393. Presotto R, Civitarese G, Bettini C. FedCLAR: federated clustering for personalized sensor-based human activity recognition. In: 2022 IEEE International Conference on Pervasive Computing and Communications (PerCom); 2022 Mar 21–25; Pisa, Italy. doi:10.1109/PerCom53586.2022.9762352.
394. Iwan I, Yahya BN, Lee SL. Federated model with contrastive learning and adaptive control variates for human activity recognition. Front Inform Technol Electron Eng. 2025;26(6):896–911. doi:10.1631/fitee.2400797.
395. Ouyang X, Xie Z, Zhou J, Huang J, Xing G. Clusterfl: a similarity-aware federated learning system for human activity recognition. In: Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services; 2021 Jun 24–Jul 2; Virtual.
396. Shen Q, Feng H, Song R, Teso S, Giunchiglia F, Xu H. Federated multi-task attention for cross-individual human activity recognition. In: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence; 2022 Jul 23–29; Vienna, Austria. doi:10.24963/ijcai.2022/475.
397. Zhang C, Zhu T, Wu H, Ning H. PerMl-Fed: enabling personalized multi-level federated learning within heterogenous IoT environments for activity recognition. Clust Comput. 2024;27(5):6425–40. doi:10.1007/s10586-024-04289-7.
398. Albogamy FR. Federated learning for IoMT-enhanced human activity recognition with hybrid LSTM-GRU networks. Sensors. 2025;25(3):907. doi:10.3390/s25030907.
399. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020;58:82–115. doi:10.1016/j.inffus.2019.12.012.
400. Dwivedi R, Dave D, Naik H, Singhal S, Omer R, Patel P, et al. Explainable AI (XAI): core ideas, techniques, and solutions. ACM Comput Surv. 2023;55(9):1–33. doi:10.1145/3561048.
401. Gohel P, Singh P, Mohanty M. Explainable AI: current status and future directions. arXiv:2107.07045. 2021.
402. Ortigossa ES, Gonçalves T, Nonato LG. EXplainable artificial intelligence (XAI)—from theory to methods and applications. IEEE Access. 2024;12:80799–846.
403. Kalasampath K, Spoorthi KN, Sajeev S, Kuppa SS, Ajay K, Maruthamuthu A. A literature review on applications of explainable artificial intelligence (XAI). IEEE Access. 2025;13:41111–40. doi:10.1109/access.2025.3546681.
404. Chaddad A, Peng J, Xu J, Bouridane A. Survey of explainable AI techniques in healthcare. Sensors. 2023;23(2):634. doi:10.3390/s23020634.
405. Ribeiro M, Singh S, Guestrin C. Why should I trust you?: explaining the predictions of any classifier. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations; 2016 Jun 12–17; San Diego, CA, USA. doi:10.18653/v1/n16-3020.
406. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the Neural Information Processing Systems 30 (NIPS 2017); 2017 Dec 4–9; Long Beach, CA, USA.
407. Ribeiro MT, Singh S, Guestrin C. Anchors: high-precision model-agnostic explanations. Proc AAAI Conf Artif Intell. 2018;32(1):1527–35. doi:10.1609/aaai.v32i1.11491.
408. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. Proc Mach Learn Res. 2017;70:3319–28.
409. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV); 2017 Oct 22–29; Venice, Italy. doi:10.1109/ICCV.2017.74.
410. Salih AM, Raisi-Estabragh Z, Galazzo IB, Radeva P, Petersen SE, Lekadir K, et al. A perspective on explainable artificial intelligence methods: shap and LIME. Adv Intell Syst. 2025;7:2400304. doi:10.1002/aisy.202400304.
411. Mortezapour Shiri F, Yamaguchi S, Ahmadon MAB. A deep learning model based on bidirectional temporal convolutional network (Bi-TCN) for predicting employee attrition. Appl Sci. 2025;15(6):2984. doi:10.3390/app15062984.
412. Hassija V, Chamola V, Mahapatra A, Singal A, Goel D, Huang K, et al. Interpreting black-box models: a review on explainable artificial intelligence. Cogn Comput. 2024;16(1):45–74. doi:10.1007/s12559-023-10179-8.

413. Albahri AS, Duhaim AM, Fadhel MA, Alnoor A, Baqer NS, Alzubaidi L, et al. A systematic review of trustworthy and explainable artificial intelligence in healthcare: assessment of quality, bias risk, and data fusion. Inf Fusion. 2023;96:156–91. doi:10.1016/j.inffus.2023.03.008.
414. Ali S, Abuhmed T, El-Sappagh S, Muhammad K, Alonso-Moral JM, Confalonieri R, et al. Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. Inf Fusion. 2023;99:101805. doi:10.1016/j.inffus.2023.101805.
415. Arrotta L, Civitarese G, Bettini C. Dexar: deep explainable sensor-based activity recognition in smart-home environments. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2022;6(1):1–30. doi:10.1145/3517224.
416. Li O, Liu H, Chen C, Rudin C. Deep learning for case-based reasoning through prototypes: a neural network that explains its predictions. Proc AAAI Conf Artif Intell. 2018;32(1):3530–7. doi:10.1609/aaai.v32i1.11771.
417. Arrotta L, Civitarese G, Fiori M, Bettini C. Explaining human activities instances using deep learning classifiers. In: 2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA); 2022 Oct 13–16; Shenzhen, China. doi:10.1109/DSAA54385.2022.10032345.
418. Aquino G, Costa MGF, Costa Filho CFF. Explaining one-dimensional convolutional models in human activity recognition and biometric identification tasks. Sensors. 2022;22(15):5644. doi:10.3390/s22155644.
419. Das D, Nishimura Y, Vivek RP, Takeda N, Fish ST, Plötz T, et al. Explainable activity recognition for smart home systems. ACM Trans Interact Intell Syst. 2023;13(2):1–39. doi:10.1145/3561533.
420. Liu C, Perumal T, Cheng J, Xie Y. Enhanced human activity recognition framework for wearable devices based on explainable AI. In: 2024 IEEE International Symposium on Consumer Technology (ISCT); 2024 Aug 13–16; Kuta, Bali, Indonesia. doi:10.1109/ISCT62336.2024.10791196.
421. Jeyakumar JV, Sarker A, Garcia LA, Srivastava M. X-CHAR: a concept-based explainable complex human activity recognition model. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2023;7(1):17. doi:10.1145/3580804.
422. Lago P, Alia SS, Takeda S, Mairittha T, Mairittha N, Faiz F, et al. Nurse care activity recognition challenge: summary and results. In: Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers; 2019 Sep 9–13; London, UK. doi:10.1145/3341162.3345577.
423. Iwana BK, Uchida S. An empirical survey of data augmentation for time series classification with neural networks. PLoS One. 2021;16(7):e0254841. doi:10.1371/journal.pone.0254841.
424. Jeong CY, Shin HC, Kim M. Sensor-data augmentation for human activity recognition with time-warping and data masking. Multimed Tools Appl. 2021;80(14):20991–1009. doi:10.1007/s11042-021-10600-0.
425. Wang J, Zhu T, Gan J, Chen LL, Ning H, Wan Y. Sensor data augmentation by resampling in contrastive learning for human activity recognition. IEEE Sens J. 2022;22(23):22994–3008. doi:10.1109/JSEN.2022.3214198.
426. Alzantot M, Chakraborty S, Srivastava M. SenseGen: a deep learning architecture for synthetic sensor data generation. In: 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops); 2017 Mar 13–17; Kona, HI, USA. doi:10.1109/PERCOMW.2017.7917555.
427. Wang J, Chen Y, Gu Y, Xiao Y, Pan H. SensoryGANs: an effective generative adversarial framework for sensor-based human activity recognition. In: 2018 International Joint Conference on Neural Networks (IJCNN); 2018 Jul 8–13; Rio de Janeiro, Brazil. doi:10.1109/IJCNN.2018.8489106.
428. Alharbi F, Ouarbya L, Ward JA. Synthetic sensor data for human activity recognition. In: 2020 International Joint Conference on Neural Networks (IJCNN); 2020 Jul 19–24; Glasgow, UK. doi:10.1109/IJCNN48605.2020.9206624.
429. Jimale AO, Mohd Noor MH. Fully connected generative adversarial network for human activity recognition. IEEE Access. 2022;10:100257–66. doi:10.1109/access.2022.3206952.
430. Hu Y. BSDGAN: balancing sensor data generative adversarial networks for human activity recognition. In: 2023 International Joint Conference on Neural Networks (IJCNN); 2023 Jun 18–23; Gold Coast, Australia. doi:10.1109/IJCNN54540.2023.10191928.
431. Chen L, Hu R, Wu M, Zhou X. HMGAN: a hierarchical multi-modal generative adversarial network model for wearable human activity recognition. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2023;7(3):1–27. doi:10.1145/3610909.

432. Bota P, Silva J, Folgado D, Gamboa H. A semi-automatic annotation approach for human activity recognition. Sensors. 2019;19(3):501. doi:10.3390/s19030501.
433. Zeng M, Yu T, Wang X, Nguyen LT, Mengshoel OJ, Lane I. Semi-supervised convolutional neural networks for human activity recognition. In: 2017 IEEE International Conference on Big Data (Big Data); 2017 Dec 11–14; Boston, MA, USA. doi:10.1109/BigData.2017.8257967.
434. Qu Y, Tang Y, Yang X, Wen Y, Zhang W. Context-aware mutual learning for semi-supervised human activity recognition using wearable sensors. Expert Syst Appl. 2023;219:119679. doi:10.1016/j.eswa.2023.119679.
435. Yun T, Wang Z. MF-match: a semi-supervised model for human action recognition. Sensors. 2024;24(15):4940. doi:10.3390/s24154940.
436. Ige AO, Mohd Noor MH. A survey on unsupervised learning for wearable sensor-based activity recognition. Appl Soft Comput. 2022;127:109363. doi:10.1016/j.asoc.2022.109363.
437. Sheng T, Huber M. Unsupervised embedding learning for human activity recognition using wearable sensor data. In: Proceedings of the Thirty-Third International Flairs Conference; 2020 May 17–20; North Miami Beach, FL, USA.
438. Takenaka K, Hasegawa T. Unsupervised representation learning method in sensor based human activity recognition. In: 2022 International Conference on Machine Learning and Cybernetics (ICMLC); 2022 Sep 9–11; Toyama, Japan. doi:10.1109/ICMLC56445.2022.9941334.
439. Bi H, Perello-Nieto M, Santos-Rodriguez R, Flach P, Craddock I. An active semi-supervised deep learning model for human activity recognition. J Ambient Intell Humaniz Comput. 2023;14(10):13049–65. doi:10.1007/s12652-022-03768-2.
440. Spelmen VS, Porkodi R. A review on handling imbalanced data. In: 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT); 2018 Mar 1–3; Coimbatore, India. doi:10.1109/ICCTCT.2018.8551020.
441. Tarawneh AS, Hassanat AB, Altarawneh GA, Almuhaimeed A. Stop oversampling for class imbalance learning: a review. IEEE Access. 2022;10:47643–60. doi:10.1109/access.2022.3169512.
442. Chen K, Yao L, Zhang D, Wang X, Chang X, Nie F. A semisupervised recurrent convolutional attention model for human activity recognition. IEEE Trans Neural Netw Learn Syst. 2020;31(5):1747–56. doi:10.1109/TNNLS.2019.2927224.
443. Alani AA, Cosma G, Taherkhani A. Classifying imbalanced multi-modal sensor data for human activity recognition in a smart home using deep learning. In: 2020 International Joint Conference on Neural Networks (IJCNN); 2020 Jul 19–24; Glasgow, UK. doi:10.1109/IJCNN48605.2020.9207697.
444. Narasimman G, Lu K, Raja A, Foo CS, Aly MS, Lin J, et al. A* HAR: a new benchmark towards semi-supervised learning for class-imbalanced human activity recognition. arXiv:2101.04859. 2021.
445. Mohamed R, Azizan NH, Perumal T, Manaf SA, Marlisah E, Hardhienata MKD. Discovering and recognizing of imbalance human activity in healthcare monitoring using data resampling technique and decision tree model. J Adv Res Appl Sci Eng Technol. 2023;33(2):340–50. doi:10.37934/araset.33.2.340350.
446. Guo S, Liu Y, Chen R, Sun X, Wang X. Improved SMOTE algorithm to deal with imbalanced activity classes in smart homes. Neural Process Lett. 2019;50(2):1503–26. doi:10.1007/s11063-018-9940-3.
447. Singh D, Merdivan E, Kropf J, Holzinger A. Class imbalance in multi-resident activity recognition: an evaluative study on explainability of deep learning approaches. Univers Access Inf Soc. 2025;24(2):1173–91. doi:10.1007/s10209-024-01123-0.
448. Sfar H, Bouzeghoub A. DataSeg: dynamic streaming sensor data segmentation for activity recognition. In: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing; 2019 Apr 8–12; Limassol, Cyprus. doi:10.1145/3297280.3297332.
449. Guo J, Li Y, Hou M, Han S, Ren J. Recognition of daily activities of two residents in a smart home based on time clustering. Sensors. 2020;20(5):1457. doi:10.3390/s20051457.
450. Geetika S, Cook DDJ. Interleaved activity recognition for smart home residents. In: Intelligent environments 2009. Amsterdam, The Netherlands: IOS Press; 2009. doi:10.3233/978-1-60750-034-6-145.

451. Aminikhanghahi S, Cook DJ. Enhancing activity recognition using CPD-based activity segmentation. Pervasive Mob Comput. 2019;53:75–89. doi:10.1016/j.pmcj.2019.01.004.
452. Wan J, O'Grady MJ, O'Hare GMP. Dynamic sensor event segmentation for real-time activity recognition in a smart home context. Pers Ubiquitous Comput. 2015;19(2):287–301. doi:10.1007/s00779-014-0824-x.
453. Chen D, Yongchareon S, Lai EMK, Yu J, Sheng QZ. Hybrid fuzzy C-means CPD-based segmentation for improving sensor-based multiresident activity recognition. IEEE Internet Things J. 2021;8(14):11193–207. doi:10.1109/JIOT.2021.3051574.
454. Kim K, Kim H, Lee D. A correlation-based real-time segmentation scheme for multi-user collaborative activities. In: 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC); 2022 Jun 27–Jul 1; Los Alamitos, CA, USA. doi:10.1109/COMPSAC54236.2022.00150.
455. Najeh H, Lohr C, Leduc B. Dynamic segmentation of sensor events for real-time human activity recognition in a smart home context. Sensors. 2022;22(14):5458. doi:10.3390/s22145458.
456. Hong Z, Li Z, Zhong S, Lyu W, Wang H, Ding Y, et al. CrossHAR: generalizing cross-dataset human activity recognition via hierarchical self-supervised pretraining. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2024;8(2):1–26. doi:10.1145/3659597.
457. Qian H, Pan SJ, Miao C. Latent independent excitation for generalizable sensor-based cross-person activity recognition. Proc AAAI Conf Artif Intell. 2021;35(13):11921–9. doi:10.1609/aaai.v35i13.17416.
458. Lu W, Wang J, Chen Y, Pan SJ, Hu C, Qin X. Semantic-discriminative mixup for generalizable sensor-based cross-domain activity recognition. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2022;6(2):1–19. doi:10.1145/3534589.
459. Qin X, Wang J, Chen Y, Lu W, Jiang X. Domain generalization for activity recognition via adaptive feature fusion. ACM Trans Intell Syst Technol. 2023;14(1):1–21. doi:10.1145/3552434.
460. Hu R, Chen L, Miao S, Tang X. SWL-adapt: an unsupervised domain adaptation model with sample weight learning for cross-user wearable human activity recognition. Proc AAAI Conf Artif Intell. 2023;37(5):6012–20. doi:10.1609/aaai.v37i5.25743.
461. Qin X, Wang J, Ma S, Lu W, Zhu Y, Xie X, et al. Generalizable low-resource activity recognition with diverse and discriminative representation learning. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; 2023 Aug 6–10; Long Beach, CA, USA. doi:10.1145/3580305.3599360.
462. Qin X, Chen Y, Wang J, Yu C. Cross-dataset activity recognition via adaptive spatial-temporal transfer learning. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2019;3(4):1–25. doi:10.1145/3369818.
463. Presotto R, Ek S, Civitarese G, Portet F, Lalanda P, Bettini C. Combining public human activity recognition datasets to mitigate labeled data scarcity. In: 2023 IEEE International Conference on Smart Computing (SMARTCOMP); 2023 Jun 26–30; Nashville, TN, USA. doi:10.1109/SMARTCOMP58114.2023.00022.