Journal of Intelligent Systems and Internet of Things
Vol. 16, No. 02, PP. 286-307, 2025
DOI: https://doi.org/10.54216/JISIoT.160221

# Sensor-Based Spatio-Temporal Human Activity Recognition: A Systematic Review of Advancements, Challenges, and Future Directions

**Asmaa Badran[1], Ahmad Salah[2,3] , A. A. Soliman[4], Dina A. Elmanakhly[5], Ahmed Fathalla[5,*]**

[1]Department of computer science, Faculty of Computers and Information, Arish University, Arish, Egypt
[2]College of Computing and Information Sciences, University of Technology and Applied Sciences, Ibri, Sultanate of Oman
[3]Department of Computer Science, Faculty of Computers and Informatics, Zagazig University, Zagazig, Sharkeya, Egypt
[4]Department of Mathematics, Faculty of Science, Arish University, Arish, 45111, Egypt
[5]Department of Mathematics, Faculty of Science,Suez Canal University, Ismailia, Egypt

Emails: asmaa.ahmedbadran@gmail.com; ahmad.salah@utas.edu.om; asoliman_99@yahoo.com; dina_almnakhly@science.suez.edu.eg; fathalla_sci@science.suez.edu.eg

**Abstract**

Spatio-temporal human activity recognition (HAR) is an emerging field that uses spatial and temporal data to identify and classify human activities accurately. It has been effectively applied in areas like healthcare for monitoring daily activities, detecting anomalies, and aiding rehabilitation with real-time feedback. However, there is a gap in research specifically focusing on integrating spatio-temporal data with advanced machine and deep learning techniques for HAR based on sensor data. Existing reviews do not comprehensively cover spatio-temporal HAR based on sensor data, resulting in a lack of summaries on recent models, datasets, sensor technologies, applications, and machine/deep learning techniques used in this field. This systematic review provides a comprehensive overview of spatio-temporal HAR based on sensor data, tracing its development from the origin of sensor-based spatio-temporal HAR field to the present. It highlights the main challenges in spatio-temporal HAR. The review also examines model trends over the years, including the distribution of models used in HAR and the identification of those frequently combined to form hybrid models. Additionally, it analyzes accuracy trends of the commonly used datasets and identifies the datasets that are widely used in spatio-temporal HAR research. Furthermore, various application domains and sensor technologies used in spatio-temporal HAR are identified.

**Received: January 17, 2025 Revised: February 12, 2025 Accepted: March 10, 2025**

**Keywords:** Deep learning; Human activity recognition; Machine learning; Systematic review; Spatio-temporal; Wearable sensors

## 1 Introduction

The space and time dimensions are combined to form spatio-temporal data. This combined data makes them widely and commonly used in various fields. However, the presence of both space and time dimensions in spatio-temporal data makes analyzing them a complex process that requires more attention. As a result, research in this field is growing and expanding rapidly. Moreover, data with only one dimension, such as space or time, have an uncomplicated analysis process compared to two-dimensional data. This complexity in analyzing spatiotemporal data makes it an interesting research area.[1] Spatio-temporal HAR is a rising field that integrates the two dimensional data to accurately recognize and classify human activities.[2–4] This approach plays an important role in environments that require continuous monitoring and real-time analysis, such as healthcare,[5] smart homes,[6] smart living,[7] and industrial applications.[8]

Spatio-temporal HAR has been effectively utilized in multiple domains.[9,10] In healthcare, these systems play an interesting role in monitoring daily activities, detecting anomalies, and supporting rehabilitation programs through real-time feedback.[2,3,5] Real-time activity recognition is especially important in healthcare applications, as prompt intervention can help prevent accidents or other health complications. These systems are specifically designed for accurate, continuous monitoring and instant feedback, tracking activity patterns over time. This capability is critical for identifying irregular movements early and facilitating rapid medical or assistive responses.[11,12]

The capabilities of spatiotemporal HAR systems extend beyond the healthcare sector to the industrial sector, where they ensure that tasks are carried out efficiently and safely through monitoring of worker activities. For example, in the logistics and manufacturing field, understanding the sequence and duration of tasks can lead to better productivity, smoother workflows, and fewer errors. These systems also contribute to a safer work environment by identifying risky movements, spotting inefficiencies, and analyzing activities to prevent potential accidents or delays.[9] From the industrial sector to the sports sector, these systems improve the performance of sports analytics by tracking and analyzing athlete movements.[10]

Spatio-temporal HAR systems play a remarkable role in advancing smart homes and intelligent environments.[6,7,9] For instance, these systems can automatically adjust lighting, temperature, and security settings based on detected occupant activities.[3,10] This not only improves daily life comfort, but also makes daily living more efficient and tailored to individual preferences.[9]

Moreover, more intuitive and responsive systems are introduced through the integration of spatio-temporal HAR and human-computer interaction,[2] where understanding the user's movements and activities in real time enhances the system's ability to adapt to the user's needs. This integration benefits virtual and augmented reality environments by enhancing user experience through accurate movement tracking, creating immersive and interactive experiences.[3]

Cameras or sensors (e.g., mobile phones, single sensors, and Inertial measurement unit (IMUs)) form the basic underlying architecture of spatio-temporal HAR systems. The focus of this study is on sensor data, such as wearable sensors, embedded environmental sensors, and IMUs. A large volume of time-series data is generated by employing these sensors. The main challenge lies in processing this large raw data to extract valuable features that represent the underlying activities.[3] This results in challenges for spatio-temporal analysis, including key issues such as data structure complexity, noise and missing data, feature extraction, large data volume, and synchronization.[13]

The crux of spatio-temporal HAR lies in effectively extracting features that capture both spatial and temporal aspects of human movements.[2] Spatial features may include the orientation and position of body parts, while temporal features capture the sequence and duration of movements.[3] Techniques such as convolutional neural networks (CNNs) are often used to extract spatial features by identifying patterns from sensor data, while recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, are employed to model temporal dependencies.[14]

Several challenges still exist in spatio-temporal HAR systems despite their progress. These challenges include handling variability in human movements and the high dimensionality of sensor data. Additionally, there is a requirement for robust models that can be generalized across various environments.[2] Moreover, data privacy, sensor placement, and energy efficiency are additional challenges to the integration of these systems in real-world settings.[3,14] Current review studies primarily focus on sensor-based HAR,[15,16] smartphone-based HAR,[17] and sensor-based datasets for HAR.[18]

The use of smartphones for HAR is systematically reviewed in,[17] and detailed information on smartphones is extracted (e.g., body location, sensors, and physical activity types), as well as data transformation and classification techniques are studied. The study focuses on health research in free-living settings based on HAR, with the objective of investigating various methods used for data acquisition, data preprocessing, feature extraction, and activity classification. Moreover, challenges that need to be addressed in utilizing smartphones for HAR in public health research are identified. Sensor-based datasets in HAR for activity recognition system evaluation are systematically reviewed in.[18]The study proposes some recommendations to researchers on the most appropriate dataset based on the type of research.

A systematic review of sensor-based Physical Activity Recognition and Monitoring (PARM) studies from an IoT layer-based perspective is proposed in.[19] It summarizes state-of-the-art traditional PARM methodologies in healthcare, including sensory, feature extraction, and recognition techniques. The paper also identifies emerging research trends and challenges in IoT environments and discusses key techniques to address them. Current advancements in wearable technologies and IoT applications for supporting independent living are highlighted in.[20] It also examines the barriers and challenges of using these solutions for older adults. The work in[21] proposes a systematic review on the use of wearable sensors for monitoring physical activity (PA) for various purposes, including assessing gait and balance, preventing or detecting falls, recognizing different PAs, conducting and evaluating rehabilitation exercises, and monitoring the progression of neurological diseases.

Table 1: Related systematic reviews on HAR based on sensor data.

| Reference | Sensor Technologies Discussion | Machine / Deep Learning Discussion | Dataset Discussion | Spatio-temporal Discussion | Applications |
|---|---|---|---|---|---|
| [21] | ✓ | Mention only the model names. | ✓ | ✗ | Healthcare |
| [20] | Overview of wearable technologies for older adults | Mention only the model names | ✗ | ✗ | Monitoring older adults |
| [19] | ✓ | Machine Learning only | ✗ | ✗ | Healthcare |
| [18] | List of sensor types in each dataset. | Mention only the model names. | ✓ | ✗ | No Application Discussion |
| [17] | Smartphone only | ✓ | ✓ | ✗ | Healthcare |
| Proposed work | ✓ | ✓ | ✓ | ✓ | ✓ |

Table 1 summarizes various systematic reviews focused on sensor-based HAR. A common limitation in these studies is the narrow discussion of sensor technologies, limited coverage of deep learning, insufficient focus on datasets, and the neglect of spatio-temporal analysis. The systematic reviews[21] utilizes sensor technologies but lacks coverage of both machine/deep learning and spatio-temporal analysis. Similarly,[20] neither incorporates sensor technologies nor employs machine/deep learning or datasets, focusing only on monitoring applications for older adults. While[19] addresses healthcare applications, machine learning, and sensor technologies, it does not cover datasets or spatio-temporal analysis. Another review study,,[18] focuses on machine learning and datasets but does not focus on sensor technologies, spatio-temporal analysis, or various applications.

Although[17] considers smartphones, datasets, healthcare applications, and machine/deep learning models, it still does not explore spatio-temporal analysis. The gaps across these studies highlight the need for more comprehensive approaches that discuss sensor technologies, machine/deep learning models, various applications, and spatio-temporal dimensions in HAR based on sensor data.

This comprehensive approach suggests that while existing studies provide valuable insights, there is a gap in the literature regarding the integration of spatio-temporal data with advanced machine/deep learning techniques for HAR. None of these reviews specifically address spatio-temporal HAR based on sensor data. As a result there is a lack of systematic review studies on spatio-temporal HAR based on sensor data that summarize recent models, datasets, current sensor technologies, applications, and the machine/deep learning techniques employed in this field.

In this paper, we conduct a systematic review of the literature on spatio-temporal HAR research based on sensor data, with the goal of exploring the challenges and leveraging the advancements in this field. The following research objectives (RO) are proposed in order to achieve this study's goal:

- **RO1:** Identify the challenges, performance metrics, and future research directions in the most recent spatio-temporal HAR models on sensor data.
- **RO2:** Explore commonly used datasets in spatio-temporal HAR research based on sensor data and evaluate their contributions to the development of effective models.

- **RO3:** Recognize the application domains and current sensor technologies employed in spatio-temporal HAR based on sensor data.

The remainder of this paper is organized as follows: Section 2 gives Research Methodology. Section 3 provides the research results. Discussions are presented in Section 4. Section 5 concluded the proposed work.

## 2 Research Methodology

The methodology used in this study is a Systematic Literature Review (SLR), as proposed by Kitchenham and Charters.[22] Planning, conducting, and reporting are the phases of their methodology. Moreover, the first phase of this study consists of a four-stage structure . In the first stage, the proposed review objectives were used to identify the research questions. During the second stage, we focused on specifying the research strategy for retrieving relevant research papers. This involved identifying the most suitable search terms and establishing the criteria for selecting papers. In the third stage, we set forth precise study selection measures, which encompassed the formulation of inclusion and exclusion rules. In the last stage, we formulated a data extraction approach specifically tailored to answer the research questions posed in the study. The review protocol employed in this paper is further clarified in the following subsections.

### 2.1 Stage 1: Research Questions

This review aims to systematically examine studies that employ sensor data for spatio-temporal HAR. To achieve this study's objectives, the following research questions were formulated. Table 2 lists the link between this study objectives and research questions.

- RQ1: What are the current challenges and future research directions in spatio-temporal HAR?
- RQ2: What datasets have been commonly used in research on spatio-temporal HAR on sensor data?
- RQ3: What are the latest models and their corresponding performance indicators in spatio-temporal HAR?
- RQ4: What are the application domains and current sensor technologies employed in spatio-temporal HAR on sensor data?

Table 2: Mapping between objectives and corresponding question number.

| Objectives | Question Number |
|---|---|
| Identify the recent models' challenges, performance issues, and solutions | Q1 & Q3 |
| Explore datasets that have been commonly used in recent studies | Q2 |
| Recognize the application domains and current sensor technologies employed | Q4 |

### 2.2 Stage 2: Search Strategy

We identified relevant studies by searching reliable databases (from the beginning of sensor-based spatio-temporal HAR field to April 21, 2024) using a comprehensive search strategy. A detailed description of the search methodology utilized in this review is presented below:

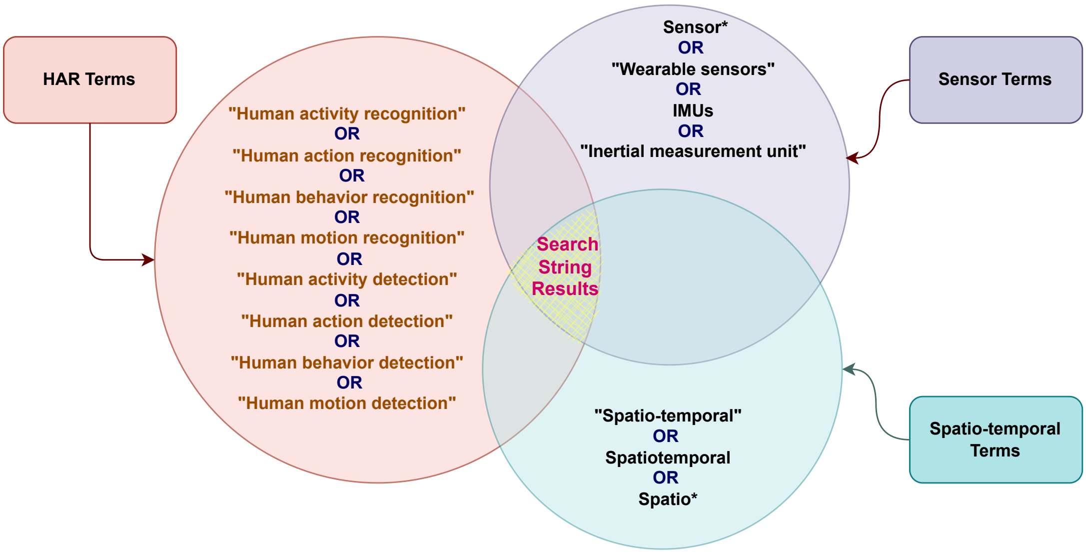


Figure 1: Query string text based on the search terms.

### 2.2.1 Search Terms

The following criteria were used to identify search terms:

- Research questions guided the identification of main search terms.
- Additional terms were derived from published literature.
- Boolean operators (AND, OR) were employed to narrow the search results. Figure 1 depicts the methodology for preparing the utilized query string where the highlighted intersection area between the three terms represents the query results. The following search terms were utilized:

  ( (”Human activity recognition” ) OR (”Human action recognition” ) OR (”Human behavior recognition” ) OR (”Human motion recognition” ) OR (”Human activity detection” ) OR (”Human action detection” ) OR (”Human behavior detection” ) OR (”Human motion detection” ) )

  AND ( (Sensor*) OR (”Wearable sensors”) or (IMUs) or (Inertial measurement unit) )

  AND ( (”Spatio-temporal”) OR (Spatiotemporal) or (Spatio*) )

### 2.2.2 Survey Resources

A comprehensive search was performed in two major academic databases: Scopus and Web of Science. These databases were chosen for their extensive coverage of peer-reviewed journals, conference proceedings, and other scholarly literature. Table 3 shows the databases and search strings used.

### 2.2.3 Search Phases

The predetermined search terms were employed to identify relevant research within the specified databases. The following section outlines the inclusion and exclusion criteria. Based on the applied inclusion and exclusion criteria, 37 publications were selected for this review.

Table 3: Databases and search strings for literature review.

| Name of the database | Search string |
|---|---|
| Scopus | TITLE-ABS-KEY ( ("Human activity recognition" ) OR ("Human action recognition" ) OR ("Human behavior recognition" ) OR ("Human motion recognition" ) OR ("Human activity detection" ) OR ("Human action detection" ) OR ("Human behavior detection" ) OR ("Human motion detection" ) ) AND ( (Sensor*) OR ("Wearable sensors") or (IMUs) or (Inertial measurement unit) ) AND ( ("Spatio-temporal") OR (Spatiotemporal) or (Spatio*) ) |
| Web Of Science | ((TI=(( ("Human activity recognition" ) OR ("Human action recognition" ) OR ("Human behavior recognition" ) OR ("Human motion recognition" ) OR ("Human activity detection" ) OR ("Human action detection" ) OR ("Human behavior detection" ) OR ("Human motion detection" ) ) AND ( (Sensor*) OR ("Wearable sensors") or (IMUs) or (Inertial measurement unit) ) AND ( ("Spatio-temporal") OR (Spatiotemporal) or (Spatio*) ))) OR AB=(( ("Human activity recognition" ) OR ("Human action recognition" ) OR ("Human behavior recognition" ) OR ("Human motion recognition" ) OR ("Human activity detection" ) OR ("Human action detection" ) OR ("Human behavior detection" ) OR ("Human motion detection" ) ) AND ( (Sensor*) OR ("Wearable sensors") or (IMUs) or (Inertial measurement unit) ) AND ( ("Spatio-temporal") OR (Spatiotemporal) or (Spatio*) ))) OR AK=(( ("Human activity recognition" ) OR ("Human action recognition" ) OR ("Human behavior recognition" ) OR ("Human motion recognition" ) OR ("Human activity detection" ) OR ("Human action detection" ) OR ("Human behavior detection" ) OR ("Human motion detection" ) ) AND ( (Sensor*) OR ("Wearable sensors") or (IMUs) or (Inertial measurement unit) ) AND ( ("Spatio-temporal") OR (Spatiotemporal) or (Spatio*) )) |

### 2.3 Stage 3: Study Selection

The initial search, using the specified keywords, identified 256 papers. A precise selection process was employed to identify relevant studies, and regular author meetings were held to discuss results. The following steps were used to select and filter papers:

- Step 1: Removing duplicate research papers across databases.
- Step 2: Remove Conference details papers from the list of papers.
- Step 3: Ensure only relevant papers are included by employing inclusion/exclusion criteria.

The used inclusion/exclusion criteria in this review paper are defined below:
Inclusion criteria:

- Include papers that focus on spatio-temporal HAR utilizing sensor data.
- Published as full-length studies written in English.
- Include studies that provide quantitative data or measurable outcomes related to human activity recognition.
- Include studies that provide machine learning or deep learning related to the interest area.

Exclusion criteria:

- Exclude papers that focus on spatio-temporal HAR utilizing other than sensor data.
- Exclude papers that are related to utilizing sensor data but do not focus on spatio-temporal HAR.
- Exclude studies that do not provide machine learning or deep learning related to the interest area.
- Exclude retracted papers.
- Non-English language studies.

Table 4: List of Challenges Identified in Spatio-Temporal HAR Based on Sensor Data

| Challenges | Ref. |
|---|---|
| The generalization of the model to different datasets or subjects is limited. | ,14,23,24,4,25,2,26,27,28,29,30,31,32,8,33,34,35,36,1237 |
| Limited robustness to noise and variability. | ,7,6,24,38,4,39,2,31,32,8,33,34,4037 |
| Complexity in Real-World Applications. | ,14,7,6,41,38,4,42,43,39,2,34,44,3212 |
| Misclassification of Similar Activities. | ,23,442,45,25,46,47,31,32,8,33,34,3537 |
| Generalization Across Different Users and Environments. | ,7,6,38,41,48,43,11,44,2936 |
| Dependency on Sensor Placement and Accuracy. | ,14,24,42,47,3,28,1235 |
| Handling of Complex and Overlapping Activities. | ,7,6,38,43,2812 |
| Complexity in Model Architecture. | ,25,27,29,32,33,3537 |
| Computational Complexity. | ,49,41,26,11,318 |
| The generalization of the model to a larger number of action categories is limited. | ,14,3040 |
| Scalability to Different Sensors. | ,2,336 |
| Limited Flexibility Across Different Data Modalities. | ,493 |
| Sensitivity to Unbalanced Datasets. | ,463 |
| Complexity in Multimodal Data Fusion. | ,4823 |
| Overfitting with Increased Layers. | ,4727 |
| Misclassification of transitions. | ,4544 |
| Difficulty in Detecting Long-Duration Activities. | 44 |

### 2.4 Stage 4: Data Extraction Strategy

In this stage, the selected papers were systematically analyzed to address the research questions. Key information extracted from each paper included: paper ID, paper title, publication year, publication type, RQ1, RQ2, RQ3, and RQ4. For RQ1, the analysis focused on identifying the challenges and future work. For RQ2, specific details were extracted, such as the dataset name, types of activities, number of subjects, device placement, number of samples, number of sensors, overlap, time window, lab versus real environment, and whether the data was raw or pre-processed. For RQ3, detailed information was collected, including the model name, model type, and performance metrics. For RQ4, information on the type of sensor and its application was collected. It is important to acknowledge that not every paper fully answered the research questions.

## 3 Results

This section will address the four research questions to accomplish the three objectives. Table 2 illustrates the link between these objectives and their corresponding questions.

### 3.1 Research Question 1

To achieve the objective of identifying the challenges, performance metrics, and future research directions in the most recent spatio-temporal HAR models, we will address two key questions (RQ1 and RQ3). This section will cover the first question: What are the current challenges and future research directions in spatio-temporal HAR? The second question will be discussed in the section 3.3.

Table 4 highlights the primary challenges faced in Spatio-temporal HAR by organizing them according to the frequency of references in the literature. This approach allows for a clear understanding of which issues are most prevalent and thus most critical to address in advancing HAR research and applications. The challenge of generalizing HAR models across different datasets, populations, users, and environments is considered one of the most frequently cited issues.

This highlights a significant constraint in model development for various applications, which is crucial for creating robust HAR systems. The inability to generalize HAR models effectively forms a barrier to the deployment and scalability of HAR systems, making it the most important focus for the research community.

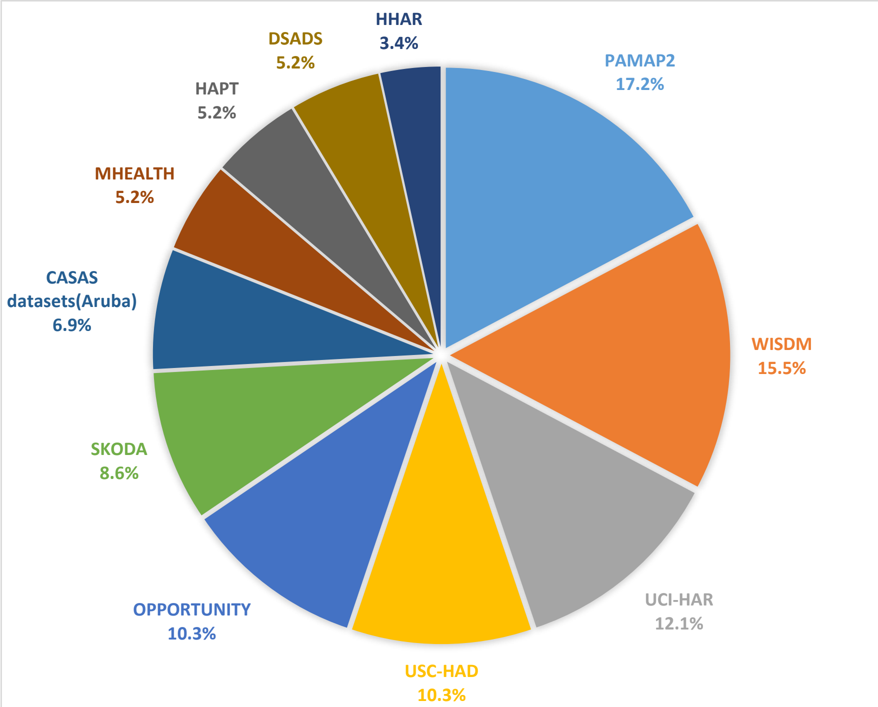


Figure 2: Percentage of Various Dataset Usage in Spatio-Temporal HAR.

The limited robustness of HAR systems against noise and variability in data is another important threat in these systems. The accuracy and reliability of these systems can be impacted by flaws in the data. As a result, the development of techniques that enhance the resilience of HAR systems against such noise and variability is vital for their effective use in practical applications.

The deployment of HAR systems in real-world environments is a complex operation and presents another important challenge. The transition from simulated environments to real-world environments introduces many variables that can affect model performance. This indicates that there is a gap between theoretical research and practical applications, which suggesting the need for scalable and adaptable techniques.

The misclassification of similar activities is also an important issue. This issue indicates the difficulty in distinguishing between activities, which can lead to errors in recognition and classification. Enhancing the accuracy and precision of HAR models in such cases is necessary to enhance their overall effectiveness.

Finally, while other challenges such as computational complexity, sensitivity to unbalanced datasets, and scalability across different sensors are referenced less frequently, they remain important considerations. These issues, though perhaps more context-specific, still represent significant hurdles that need to be addressed to further refine and optimize HAR systems.

### 3.2 Research Question 2

In order to identify the datasets that have been commonly used in research on spatio-temporal HAR based on sensor data and to achieve the second objective, Figure 2 shows the repetition counts for each dataset in the research, considering only datasets with more than one repetition.

The contribution of each dataset in terms of its repetition count relative to the total is represented in Figure 2. PAMAP2 has the largest percentage, accounting for approximately 17.2% of the total repetitions. This indicates that it is the most frequently used dataset, contributing significantly to research in spatio-temporal human activity recognition. The second most frequently used dataset is WIDSM and makes up around 15.5%

of the total repetitions. UCI-HAR contributes about 12.1% to the total repetitions, indicating its widespread use and importance in the research community.

USC-HAD accounts for approximately 10.3% of the repetitions, showing its moderate but significant presence in the research. OPPORTUNITY adds about 10.3%, much like USC-HAD, showing its importance in specific research areas. SKODA covers about 8.6% of the total uses. Its use is moderate, showing it has some importance in select research contexts. CASAS datasets (Aruba) provide around 6.9%, suggesting it is not often used but has some importance in specific research areas. MHEALTH adds about 5.2%, showing it has a smaller yet still important role in research. DSADS also offers about 5.2%, indicating it is used similarly to MHEALTH in the research domain. HAPT accounts for 5.2% of the repetitions, indicating its role is on par with DSADS and MHEALTH. HHAR has the smallest part, making up about 3.4% of the total repeats. This means that the dataset is used the least compared to those shown in the chart.

As a high usage in research PAMAP2, WIDSM, and UCI-HAR are the most contributing datasets. PAMAP2 dataset is well-suited for HAR tasks, due to the diversity of activities, the high-quality data collected at 100 Hz, and its balanced class distribution.[46] This makes the dataset a useful benchmark to compare the models,[12] where also advanced deep learning approaches[48] can be efficiently utilized for developing the precise and robust HAR models. The WISDM dataset is useful for HAR research as it describes a lot of diverse activities.[34,37] It enables high accuracy in recognition tasks[34,50] as well as rich temporal sequences.[2]

Moreover, its public availability and wide utilization encourage standardization,[31] thus serving as a great resource for HAR models' benchmarking and testing among the academic community. UCI-HAR is a popular benchmark dataset in HAR research and it is known for its diversity of daily activities and standardized method for data collection.[26,34] It is providing rich sensor information that enhances model accuracy.[26,30] Its structure is well-suited for deep learning models, enabling effective training and evaluation.[30,34] Table5 provides summary of commonly datasets used for activity recognition based sensors.

Raw data is the initial, unaltered data captured by sensors, while preprocessed data has been cleaned and transformed to facilitate its use in spatio-temporal HAR systems. In Table 5, we can see that 82% of the considered datasets (with more than one repetition) consist of raw data. This is because raw data provides the most detailed and comprehensive information, but it requires significant processing to be useful. It is essential for researchers who want to develop or test new pre-processing methods. On the other hand, pre-processed data is ready for immediate use in machine learning models, reducing the time and effort required to develop HAR systems. It is crucial for standardizing datasets and ensuring consistency in model training and evaluation. The choice between using raw or pre-processed data depends on the specific goals of the research or application.

Laboratory datasets are ideal for developing and testing specific models where consistency and control are crucial. They are useful in the initial research stages, where the goal is to understand the basic mechanics of human activities. In contrast, real datasets are more applicable when the goal is to deploy HAR systems in real-world applications, where the variability and complexity of everyday environments must be accounted for.

Real datasets are essential for developing models that can operate effectively in the chaotic and unpredictable conditions of the real world. From Table5, we can see that 55% of the considered datasets (with more than one repetition) are real datasets, while 45% are laboratory datasets. The nearly even split between real and laboratory datasets suggests a balanced approach in HAR research. While real datasets are crucial for practical deployment, laboratory datasets still play a vital role in the early stages of model development and in understanding fundamental aspects of human activity. This balance allows researchers to leverage the strengths of both controlled experiments and real-world observations. Both laboratory and real datasets play crucial roles in HAR research, but they serve different purposes. Laboratory datasets provide a controlled foundation for understanding activities, while real datasets are critical for ensuring that HAR systems are robust and effective in everyday scenarios.

### 3.3 Research Question 3

This section will tackle the question: What are the latest models and their corresponding performance indicators in spatio-temporal HAR? This question corresponds to the second step in achieving the first objective.

Table 5: Summary of activity recognition datasets.

| Dataset | Activities | # of Subjects | Place of Device | # of Samples | # of Sensors | Overlap | Time Window | Lab / Real | Raw / Pre-processed |
|---|---|---|---|---|---|---|---|---|---|
| OPPORTUNITY | 18 activities: Open Door 1, Open Door 2, Close Door 1, Close Door 2, Open Fridge, Close Fridge, Open Dishwasher, Close Dishwasher, Open Drawer 1, Close Drawer 1, Open Drawer 2, Close Drawer 2, Open Drawer 3, Close Drawer 3, Clean Table, Drink from Cup, Toggle Switch, Other | 4 | Wrist, Chest, Limb, Shoulder, Foot | 49,609 | 19 | 50% | 1.5 - 5 s | Lab | Pre-processed |
| PAMAP2 | 12 activities: Lying, Sitting, Standing, Walking, Running, Cycling, Nordic walking, Ascending stairs, Descending stairs, Vacuum cleaning, Ironing, Rope jumping | 9 | wrist of the dominant arm, chest, ankle | 118,604 | 3 | 50% | 5 s | Real | Raw data |
| SKODA | 10 activities: Write on notepad, Open hood, Close hood, Check gaps on the front door, Open left front door, Close left front door, Close both left doors, Check trunk gaps, Open trunk, Close trunk | 1 | left and right arms | 235,303 | 20 | 50% | 1.5 s | Lab | Raw data |
| USC-HAD | 12 activities: Walking Forward, Walking Left, Walking Right, Walking Upstairs, Walking Downstairs, Running Forward, Jumping Up, Sitting, Standing, Sleeping, Elevators Up, Elevators Down | 14 | Right Hip | 59,941 | 2 | 50% | 1 s | Lab | Raw data |
| HAPT | 12 activities (6 basic activities: Standing, Sitting, Laying, Walking, Walking upstairs, Walking downstairs; 6 postural transitions: Stand-to-sit, Sit-to-stand, Sit-to-lie, Lie-to-sit, Stand-to-lie, Lie-to-stand) | 30 | waist | 11,883 | 2 | 50% | 2.56 s | Lab | Raw data |
| MHEALTH | 12 activities: Climbing stairs, Cycling, Front elevation of arms, Jogging, Jump front & back, Knees bending, Lying down, Running, Sitting & relaxing, Standing still, Waist bends forward, Walking | 10 | Right wrist, Left ankle, Chest | 34,097 | 3 | 50% | 2.56 s | Real | Raw data |
| UCIDSADS | 19 activities: Sitting, Standing, Lying on back, Lying on right side, Ascending stairs, Descending stairs, Standing in an elevator still, Moving around in an elevator, walking in a parking lot, Walking on a treadmill at 4 km/h in flat position, Walking on a treadmill at 4 km/h in 15° inclined position, Running on a treadmill at 8 km/h, Exercising on a stepper, Exercising on a cross trainer, Cycling on an exercise bike in horizontal position, Cycling on an exercise bike in vertical position, Rowing, Jumping, Playing basketball | 8 | torso, arms, and legs | 113,848 | 5 | 50% - NON | 0.8 s - 5 s | Real | Raw data |
| WISDM | 6 activities: Walking, Jogging, Walking Upstairs, Walking Downstairs, Sitting, Standing | 36 | front pocket of the participant's pants | 109,8207 | 1 | 50% | 4 s | Lab | Pre-processed |
| UCI-HAR | 6 activities: Walking, Laying, Walking Upstairs, Walking Downstairs, Sitting, Standing | 30 | Waist | 10,299 | 2 | 50% | 2.56 s | Real | Raw data |
| Aruba dataset | 10 activities: Meal_Preparation, Relax, Eating, Work, Sleeping, Wash_Dishes, Bed_to_Toilet, Enter_Home, Leave_Home, Housekeeping | 1 | distributed across various rooms | 538,731 | 34 | NON | 5 s - 20 min | Real | Raw data |
| HHAR | 6 activities: Biking, Sitting, Standing, Walking, Stair Up, Stair Down | 9 | pocket, wrist | 176,237 | 2 | 50% | 2.56 s | Real | Raw data |

As shown in Figure 3, the models used during the early years (2008-2017) are simpler and more traditional, including decision trees, K-Nearest Neighbors (KNN), and Naive Bayes. These are foundational machine learning techniques that were widely adopted for their simplicity and effectiveness in various domains.

Emergence of Deep Learning a significant shift occurs in 2020, where deep learning models like CNNs and LSTMs begin to dominate. This marks the start of a trend where models increasingly focus on sequence data (e.g., LSTM, Bi-LSTM) and spatial data (e.g., 2D-CNN). The introduction of attention mechanisms also becomes noticeable, indicating an interest in improving model performance on more complex tasks.

Between 2021 to 2022, The progression for more complex architectures stays. Models start by including many techniques, such as combining CNNs with LSTMs or GRUs. There's also a rise in custom models and GANs, showing a shift towards more innovative and tailored solutions for specific tasks.

In 2023, there has been an an explosion of model diversity, including advanced neural networks (Capsule Networks, Transformer models), ensemble methods, and newer architectures like SpikeCNN and SpikeDNN. The focus on model compositions remains continued, with a combination of CNN with KNN or BiGRU, underlining the requirements for hybrid approaches for the complex problem solution. In 2024, the models further evolve, with a strong emphasis on graph-based models (Graph Neural Network (GNN), Graph Convolutional Network (GCN) ) and attention mechanisms, showing a trend towards handling more structured data and complex dependencies. The use of Random Forest and other ensemble methods also persists, indicating that traditional methods are still valuable, especially when combined with deep learning techniques.

To identify the most commonly combined models to form hybrid models, Figure 4 presents the frequency of these models. As shown in Figure 4, LSTM and CNN are the two most commonly used models, with LSTM accounting for 30% and CNN accounting for 28%. Due to the capability of LSTM, make it have a strong preference for modeling sequences in time-series data, which is important for understanding the temporal

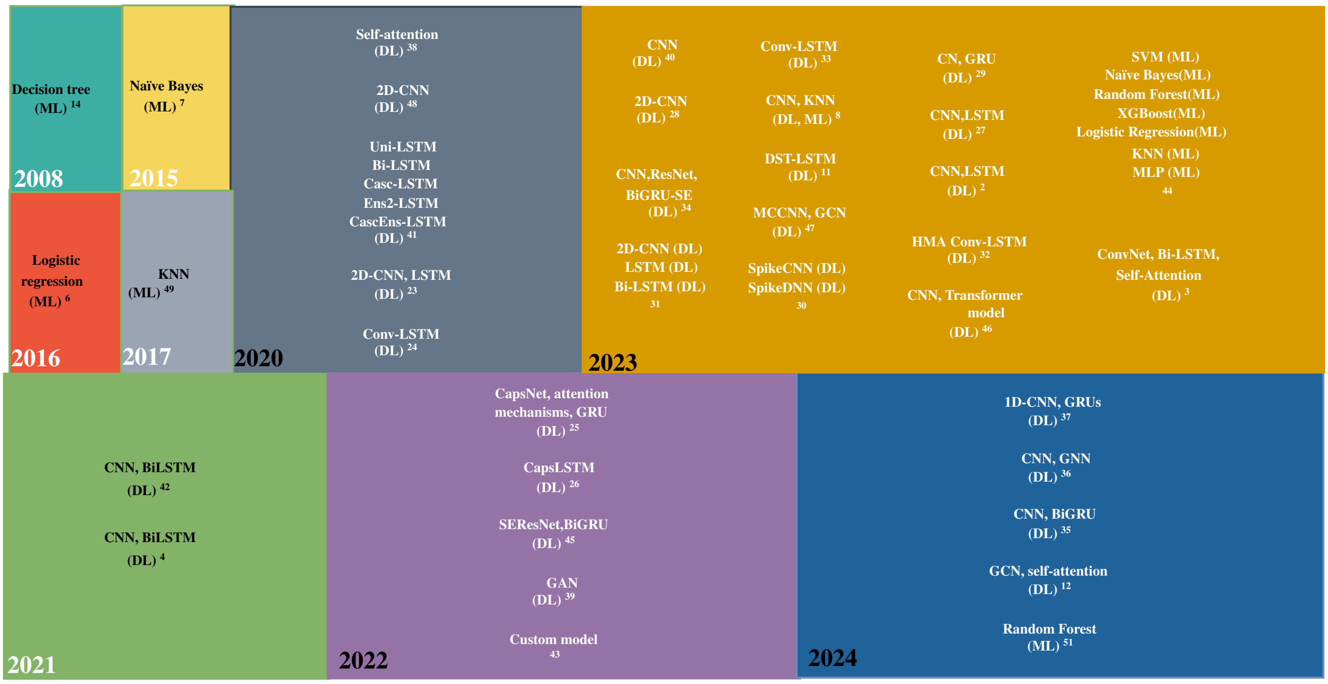


Figure 3: Distribution of models employed over the years in HAR

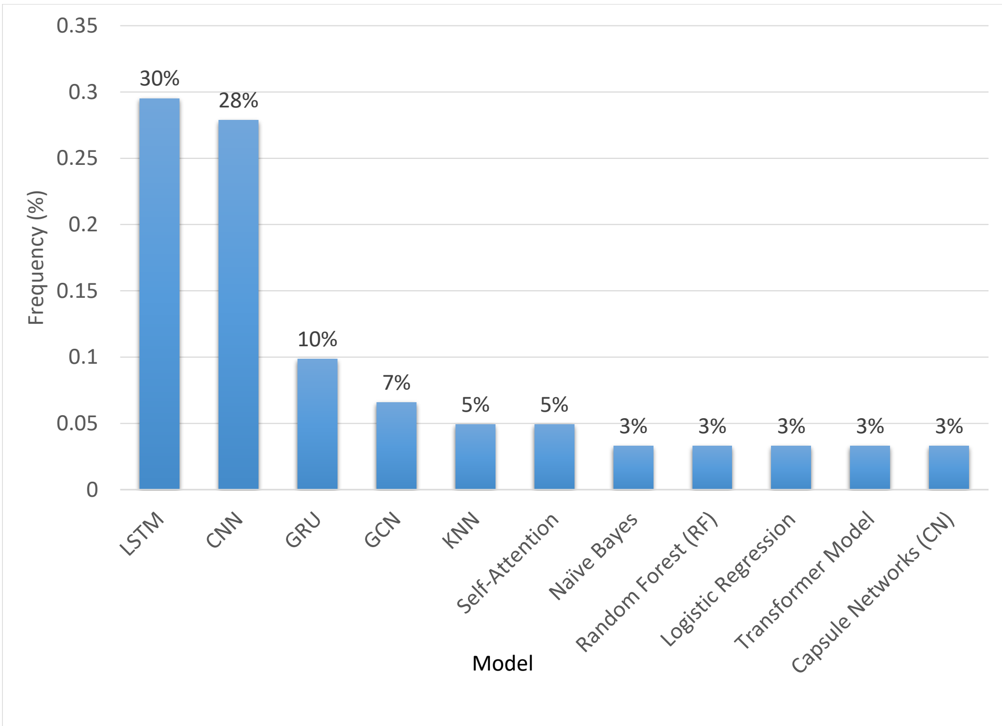


Figure 4: Frequency distribution of various models used to form hybrid models.

Table 6: Count of Usage of Various Models in Spatio-Temporal HAR.

| **Model** | 2D-CNN | CNN, BiLSTM | CNN, LSTM | Conv-LSTM | Bi-LSTM | KNN | Naïve Bayes | Random Forest (RF) | Other |
|---|---|---|---|---|---|---|---|---|---|
| **Frequency** | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 34 |

dynamics of human activities. CNN's near-equal frequency suggests its vital role in extracting spatial features from data, such as those found in video frames or sensor data, which are essential for recognizing the spatial aspects of human movements.

Gated Recurrent Unit (GRU) also plays a significant role in Spatio-temporal HAR, with a usage frequency of 10%. GRU's simpler structure compared to LSTM makes it an attractive option when computational efficiency is a priority, yet it still provides the benefits of handling temporal sequences effectively. The presence of GRU in Figure 4 suggests that GRU is still related to the scenarios that required a balance between performance and efficiency.

Models such as GCN, with a frequency of 7%, and KNN and Self-Attention, both at 5%, appear less frequently. These lower frequencies suggest that these models are used more selectively in spatio-temporal HAR. GCNs are specifically designed to work with graph structured data. Moreover, GCNs are often applied in scenarios where activities can be represented as nodes and edges. On the other hand, the simplicity of KNN may limit its application to less complex scenarios where deep-learning models like LSTM or CNN are not required.

The less frequently used models in spatio-temporal HAR include Naive Bayes, Random Forest (RF), Logistic Regression, the Transformer model, and Capsule Networks (CN), where each with a frequency of 3%. This lower frequency might be due to their limited capability to effectively handle both spatial and temporal dependencies at the same time, which are essential for correctly recognizing human activities.

Table 6 presents various models employed in spatio-temporal HAR and highlights the importance of certain models over others. As indicated by Table 6, the 2D-CNN model is the most frequently used, with a frequency of 3. This suggests that 2D-CNN has acuired significant attention within the research community, this could due to its effectiveness in handling spatial features in HAR tasks.

On the other hand, other models with a frequency of 2 are CNN combined with BiLSTM[4,42] and LSTM,[2,27] Conv-LSTM,[32,33] Bi-LSTM,[3,31] K-Nearest Neighbors (KNN),[44,49] Naïve Bayes,[7,44] and Random Forest (RF).[44,51] While the other models, each with a frequency of one, have a total count of 34. The consistent frequency among these models indicates a balanced exploration and application in HAR research. The usage of CNN in conjunction with BiLSTM and LSTM reflects a common approach to leveraging both spatial and temporal features in human activity data. Conv-LSTM's frequency suggests its relevance in tasks that require both convolutional layers for spatial feature extraction and LSTM layers for capturing temporal dependencies.

Bi-LSTM, with its ability to process sequences in both forward and backward directions, is also seen as a valuable tool in HAR. The inclusion of KNN, Naïve Bayes, and Random Forest models, which are more traditional machine learning approaches, alongside deep learning models, highlights the diversity of strategies employed by researchers in this field. This diversity may be driven by the varying nature of HAR datasets and the specific needs of different applications, which sometimes benefit from the simplicity and interpretability of traditional models. The models presented in Table 6 have a frequency greater than one and account for 31% of the total models in the research papers. From Figure 3, we can identify all the models proposed by the research papers, including the remaining 69% of models with one frequency.

The accuracy of the three datasets—PAMAP2, UCI-HAR, and WISDM—from 2020 to 2024 are given in Figure 5. The overall direction demonstrates significant improvements in model accuracy over time, with a noticeable peak in 2023, followed by a slight decline in 2024.

Starting with the PAMAP2 dataset, the accuracy in 2020 was already high at 95%, reflecting a solid baseline performance. By 2023, the accuracy had jumped to 99%, indicating significant advancements in model performance, potentially due to better training techniques or enhancements in data processing. However, in 2024, there is a minor decline in accuracy to 98.18%. While this decrease is slight, it suggests that some challenges may have arisen, such as more complex data or slight overfitting in previous models that didn't generalize as well on newer data.

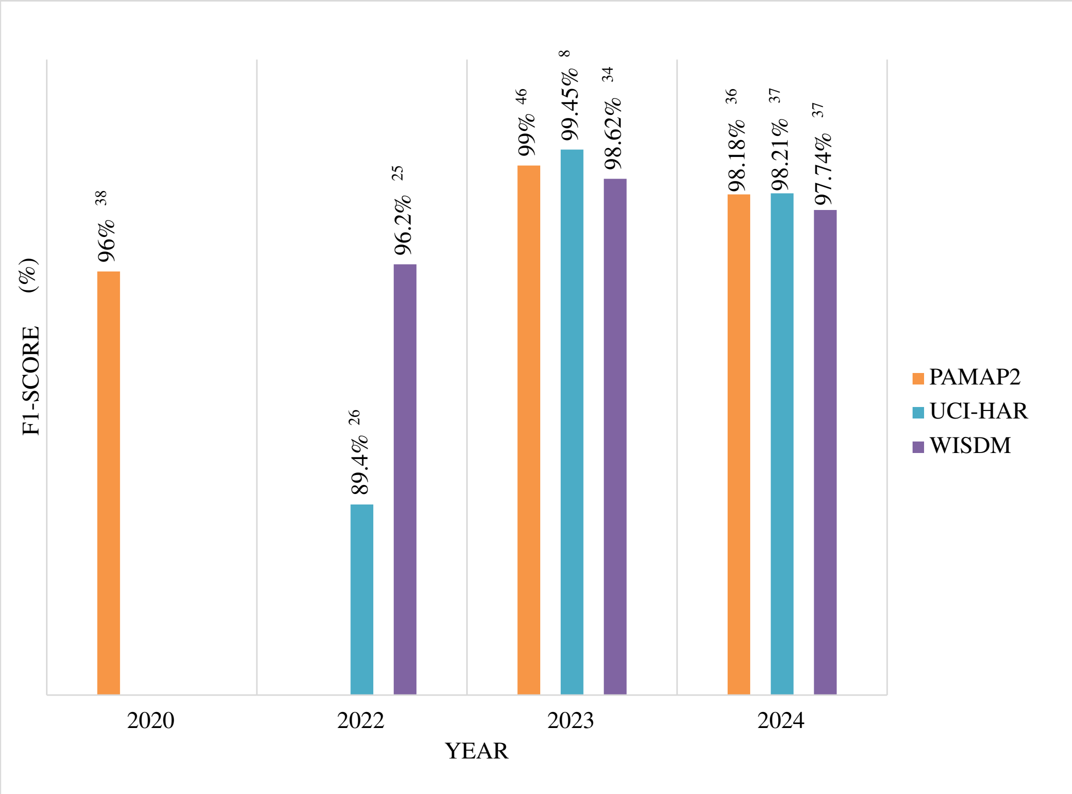


Figure 5: Accuracy trends of PAMAP2, UCI-HAR, and WISDM datasets from 2020 to 2024.

The UCI-HAR dataset shows a different trajectory. In 2022, the accuracy was relatively low at 89.4%, possibly indicating more challenging data or less effective models at that time. However, the accuracy surged dramatically to 99.45% in 2023, the highest of all the datasets across all years. This sharp increase could be attributed to substantial improvements in model architectures, feature extraction techniques, or data handling processes. Despite this impressive gain, the accuracy slightly declined to 98.21% in 2024, similar to PAMAP2, suggesting that while the model remains robust, maintaining such high accuracy may be challenging as the complexity of the data or the evaluation criteria evolve.

For the WISDM dataset, the trend is somewhat consistent, starting at 96.2% accuracy in 2022 and increasing to 98.62% in 2023. This steady improvement highlights the model's effective handling of the data. However, like the other datasets, WISDM experienced a slight dip in accuracy to 97.74% in 2024. This consistent pattern of a minor drop across all datasets in 2024 could indicate a common underlying issue, such as the introduction of more challenging data scenarios or the limitations of the current modeling approaches.

### 3.4 Research Question 4

The final objective is to identify the application domains and the current sensor technologies used in spatio-temporal HAR based on sensor data. In this section, we will address the question: What are the application domains and current sensor technologies utilized in spatio-temporal HAR based on sensor data? This will help us achieve this objective.

Table 7 summarizes various application domains where spatio-temporal HAR is employed, alongside the specific types of sensors used in these applications. Different domains or contexts where spatio-temporal HAR is applied. Include the following applications:

- Smart homes: systems that monitor and automate activities within home environments.
- Healthcare: monitoring and assessing health-related activities, often in clinical or home settings.

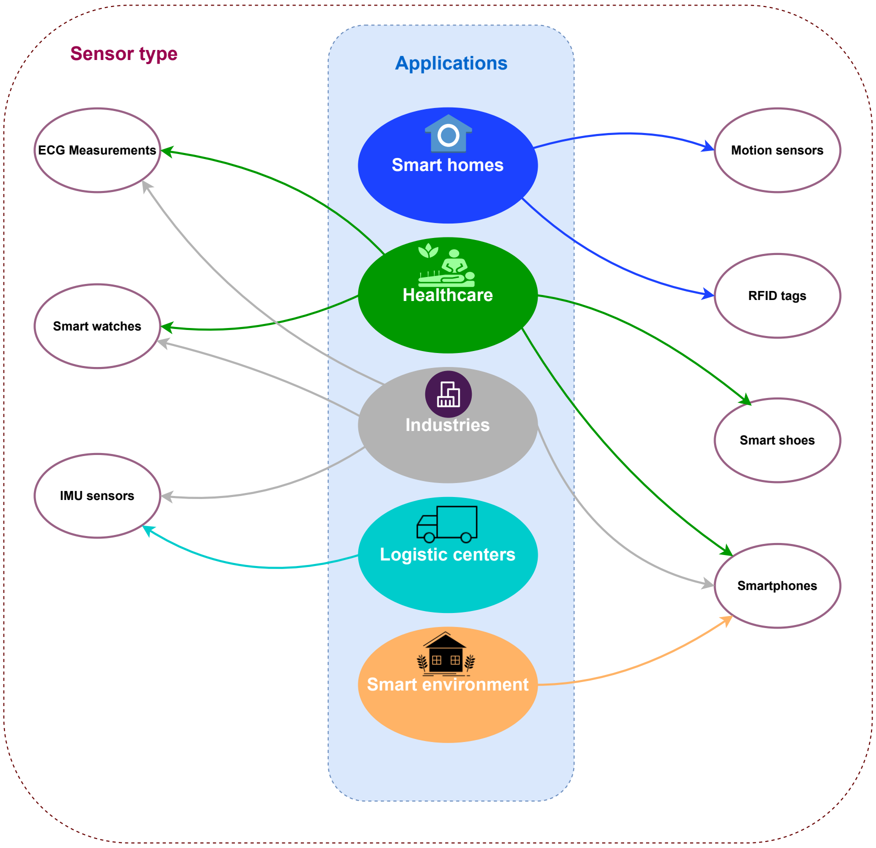


Figure 6: Applications and sensors types.

- Industries: industrial applications where HAR can improve safety, efficiency, and automation.
- Logistic centres: environments like warehouses where tracking human activity can optimize operations.
- Smart environment: broader environments, including smart cities and offices, where human activities are monitored for improved interaction and management.

In each application, specific types of sensors are employed to capture spatio-temporal data. These types of sensors include Motion sensors, RFID tags, Smart shoes, Smartphones, Smart watches, ECG Measurements, and IMU sensors. Table 8 lists various sensor types and their usage in spatio-temporal HAR applications.

It is observed from Table 7 that motion sensors and RFID tags are widely employed to monitor activities in smart homes, while in healthcare applications, smart shoes, smartphones, smartwatches, and ECG measurements are widely used for health-related activities. Industries similar to healthcare, IMU sensors, smartphones, smartwatches, and ECG measurements are used for activity monitoring in industrial settings. Logistic Centres primarily use IMU sensors for tracking and optimizing human activities within warehouses. Smartphones are the primary sensors used to monitor activities within broader smart environments, such as cities, offices, and prisons. Figure 6 categorizes various sensor types and their potential applications across different domains. From Figure 6, smartphones are commonly used as sensor platforms due to: Smartphones are equipped with a variety of sensors, such as accelerometers, gyroscopes, GPS, cameras, microphones, light sensors, and sometimes even heart rate monitors. This makes them versatile tools for a wide range of applications, from navigation and fitness tracking to environmental monitoring. Furthermore, the availability of smartphones allows users to Innovate applications that leverage the device's sensors for specific purposes, ranging from fitness tracking to environmental monitoring. It is noted from Table 7 that only 47% of the studies provide specific application domains for spatio-temporal HAR.

Figure 7 represents various sensor types along with their occurrence counts, representing how frequently each sensor type was mentioned or utilized.

Table 7: List of Applications and Sensors Types Identified in Spatio-Temporal HAR.

| Applications | Sensors type | Ref. |
|---|---|---|
| Smart homes | Motion sensors | [14], [6], [41], [43], [28][44] |
| | RFID tags | |
| Healthcare | Smart shoes | [48], [26], [39], [11], [27], [30][12] |
| | Smartphones | |
| | Smart watches | |
| | ECG Measurements | |
| Industries | IMU sensors | [8] |
| | Smartphones | |
| | Smart watches | |
| | ECG Measurements | |
| Logistic centers | IMU sensors | [3] |
| Smart environment | Smartphones | [7][25] |

Table 8: List of Sensor Types and Their Usage in Spatio-Temporal HAR

| **Sensor Type** | **Sensor Usage in Spatio-Temporal HAR Applications.** |
|---|---|
| Motion sensors | Detect movement and occupancy |
| RFID tags | Tracking the presence and movement of objects or people. |
| Smart shoes | Monitor walking patterns and other lower-body activities, commonly used in healthcare. |
| Smartphones | Used across various applications for mobility tracking. |
| Smart watches | Monitor physical activity. |
| ECG Measurements | Monitor heart activity, used in healthcare and some industrial applications. |
| IMU sensors | Track motion in industries and logistics. |

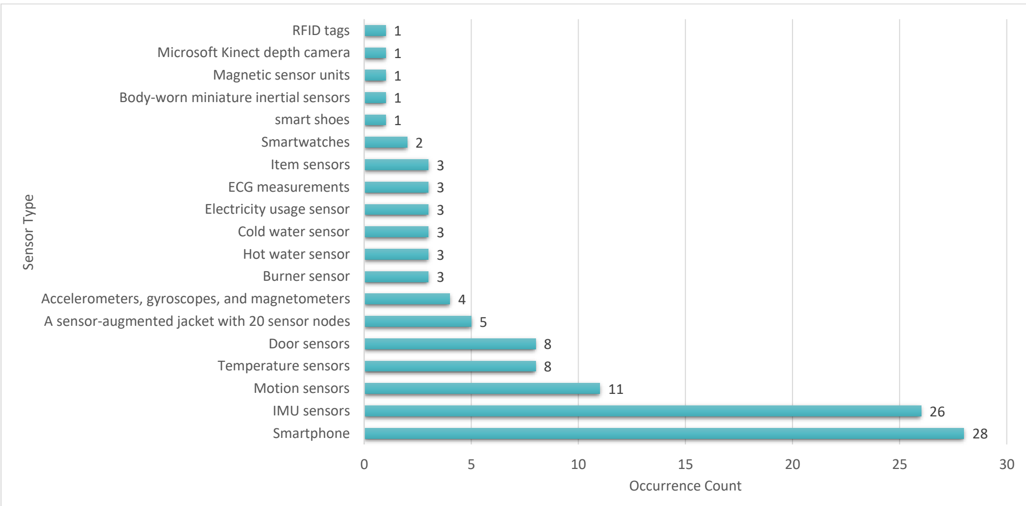


Figure 7: Distribution of sensor types highlights the diverse range of technologies employed in HAR.

Highly Used Sensor Types:

- Smartphone: the most frequently mentioned sensor type with 28 occurrences. This highlights the widespread use of smartphones in HAR due to their built-in sensors (e.g., accelerometers, gyroscopes, GPS) and their accessibility.
- IMU Sensors: occurring 26 times, these sensors are critical in capturing motion-related data, often used in wearables and other portable devices for precise movement tracking.

Moderately Used Sensor Types:

- Motion Sensors: with 11 occurrences, these sensors are commonly used in smart environments to detect human presence and movement.
- Temperature Sensors: mentioned 8 times, these sensors are typically used in environmental monitoring or to assess conditions in smart homes or industrial settings.
- Door Sensors/Closure Sensors: both types have 8 occurrences, commonly used in smart home systems for detecting entry and exit.

Less Used but Noteworthy Sensor Types:

- Sensor-Augmented Jacket with 20 Sensor Nodes: mentioned 5 times, indicating the use of specialized wearable technology in certain applications.
- Accelerometers, Gyroscopes, and Magnetometers: this combination of sensors appears 4 times, reflecting its importance in providing comprehensive motion data.
- Burner Sensor, Hot Water Sensor, Cold Water Sensor, ECG Measurements, Item sensors and Electricity Usage Sensor: each of these sensor types appears 3 times, likely indicating their use in smart home or industrial monitoring systems.

Rarely used Sensor Types: Sensor types with a single or two occurrence include Smart Watches, Body-Worn Miniature Inertial, Magnetic Sensor Units, Microsoft Kinect Depth Camera, RFID Tags, and Single MotionNode. These may be used in niche applications or less frequently in HAR scenarios.

## 4 Discussions

As the field of spatio-temporal HAR continues to evolve, significant advancements have been made in developing models and methods that harness sensor data to accurately identify and classify human activities. Despite these significant advancements, many challenges still exist that hinder the applicability and integration of HAR systems in the real world in an efficient way. As a result, addressing these challenges is vital for the continuity and growth of HAR systems.

In this section, key challenges discovered during the exploration of research in spatio-temporal HAR based on sensors are outlined, along with the challenges researchers face in the development and deployment of HAR systems. Insights and future research directions are proposed, aimed at improving system performance and ensuring that HAR technologies can be effectively integrated into diverse environments. By addressing these challenges and opportunities, the field can move closer to realizing the full potential of HAR systems across various applications.

### 4.1 Discussions on RO1

The gradual development of HAR models has progressed from traditional machine learning techniques to deep learning techniques, and more recently, to hybrid and graph-based models.

The growth in the development of HAR models indicates the need for more innovative methods to effectively address the increasing complexity of HAR tasks.

To address the challenges raised by the complexity of HAR tasks,the following areas could be considered: employing hybrid models, focusing on the usage of graph-based models, integrating models to adapt to various environments, and ensuring robustness and scalability. The strengths of machine learning, deep learning, and domain-specific techniques can be combined to create hybrid models and to achieve enhanced performance. Hybrid models can present improved performance by investing the complementary strengths of different techniques. For example, by combining feature extraction capabilities of machine learning with the predictive capability of deep learning. Graph-based models propose a promising area for future development. The ability of these models to capture and analyze relationships between entities makes them highly useful in various domains (e.g., social networks, recommendation systems, chemical analysis, and healthcare data analysis). The applications of graph based models in HAR should be explored to better understand and interpret complex HAR tasks.

The challenge of generalizing models across different environments is another future research direction that should be addressed. This due to HAR applications expanded into various environments and domains. This could include creating more robust training datasets, or designing adaptable models.

Additionally, scalability and robustness against noise and variability across different sensors and data modalities should be considered in future models. This includes developing models that are aware of noisy and incomplete data and can scale well across different types of sensors and data sources, managing a large number of sensors without affected by specific hardware constraints.

Future research could explore the integration of advanced noise reduction techniques and the use of more robust algorithms capable of handling noisy inputs effectively, and develop models that are less sensitive to variability in sensor data and environmental conditions.

The deployment of HAR systems in real world environments can be simplified by reducing model complexity and improving scalability. This includes improving model architectures and verifying that models can be integrated into existing systems. As a research direction, there should be a focus on developing modular HAR systems that can be deployed with various simplified configurations. Furthermore, making HAR systems more practical for deployment on resource constrained devices, such as mobile phones or embedded systems is essential. This can be achieved by reducing computational complexity of these systems while maintaining high accuracy.

Another research direction could involve enhancing classification techniques that can effectively recognize and differentiate similar activities by integrating various multimodal data sources. Moreover, it is essential to increase the capability of HAR systems to detect and classify complex and overlapping activities, as well as transitions between activities.

### 4.2 Discussions on RO2

The remarkable diversity in activities, sensor placements, number of subjects, and data collection methods gives rise to a range of research needs, from specific task recognition (e.g., opening doors, operating car parts) to general physical activity recognition (e.g., walking, running, sitting). To further enrich the HAR field, further research should concentrate on the following areas:

- Diverse Dataset Development: Create datasets that covering a large number and wide scope of activities and scenarios, including more accurate and complex tasks. This will improve the generalization of HAR models to a wider range of real-world activities.

- Sensor Placement and Configuration Enhancement: Explorer new sensor placements and settings to capture more comprehensive and contextually relevant data. This could involve exploring new sensor types or optimizing existing configurations for various environments.
- Subject Diversity: Increase the variety of subjects in datasets to better represent various populations and physical conditions , which will benefit in developing generalized HAR models.
- Improvement of Data Collection Techniques: Develop data collection methods to improve data quality and reliability. This may include developing new techniques for collecting data in challenging environments or enhancing the precision of data recording tools.
- Standardization and Benchmarking: Encourage the standardization of datasets and benchmarking practices to enable more consistent comparisons across different research studies. Establishing common evaluation metrics and benchmarks will facilitate progress and innovation in HAR techniques.
- Multimodal Data Integration: Incorporate data from multiple sensors and modalities to create more robust and comprehensive datasets. Multimodal data integration can provide richer information and improve the performance of activity recognition techniques.

The majority of datasets (about 55%) are real datasets, indicating a growing focus on ensuring that HAR systems are applicable and robust in real-world scenarios. This reflects the recognition that models trained and tested only on controlled laboratory data may not perform as well in the unpredictable conditions of everyday life. While real datasets are slightly more prevalent, the close percentage with laboratory datasets (about 45%) suggests that collecting and processing real-world data remains challenging. Real-world data is often noisy, complex, and less structured, requiring advanced processing techniques and more robust models to handle its variability.

The relatively close percentages may also highlight the need for hybrid datasets that combine elements of both laboratory and real-world data. Such datasets could offer the best of both of them, providing the controlled conditions needed for initial model development while also incorporating the variability of real-world scenarios to ensure broader applicability. Researchers should consider this balance when designing new studies or developing HAR systems.

### 4.3 Discussions on RO3

As we examine the research literature to identify the current sensor technologies utilized in spatio-temporal HAR based on sensor data, smartphones and IMU sensors are the most frequently mentioned types. The numerous built-in sensors in smartphones enable continuous measurement of daily activities, making them especially well-suited for HAR research. This allows researchers to translate smartphone data into various types of physical activity and enhance HAR systems.[17] The ubiquitous nature of smartphones allows for widespread data collection, providing a rich source of real-time information for HAR systems.

On the other hand, IMU sensors are highly effective in capturing detailed motion data, such as linear acceleration and angular velocity, and are used to track body activity and gestures while maintaining privacy. Besides, IMU sensors allow users to perform daily activities without being restricted by the scope of the activity.[52] These sensors can monitor physical activities by integrating them into wearable devices or smartphones, which making them essential for accurate HAR.

Smartphones and IMU sensors provide spatio-temporal data that enables HAR systems to precisely recognize and analyze human activities in both spatial and temporal dimensions. This leads to enhancing the system's overall effectiveness and applicability in real-world scenarios.

Various new applications can be supported by the vast amount of data collected from smartphones and IMU sensors. HAR systems use this data to gain insights into various people's activity patterns and behaviors. Planners can use this data to design more efficient and accessible public spaces, improve traffic flow, and enhance overall infrastructure, which leads to user friendly environments that better meet the needs of the population.

Less frequently used sensor types are Burner Sensors, Hot Water Sensors, Cold Water Sensors, ECG Measurements, Item Sensors, and Electricity Usage Sensors. These sensor types can be integrated to provide new applications. Spatio-temporal HAR based on sensor data involves operating with multiple types of sensors to capture an overall overview of human activities over time and space. By integrating data from these sensors, spatio-temporal HAR systems can achieve a more detailed and accurate understanding of various activities, which is valuable for valuable for different applications.

Spatio-temporal HAR based on sensor data has various real and potential applications that require further inspection. For example, spatio-temporal HAR based sensor data systems can be integrated with cognitive smart cities to support the sustainability of smart cities. Besides, these systems can be integrated with different Sport Analysis Systems to provide more comprehensive insights into sports activities. Moreover, integrating these systems into education systems could enhance student learning, well-being, and safety. As education continues to evolve, especially with the growing importance of remote and hybrid learning models, the integration of HAR systems will likely play a critical role in shaping the future of education.

## 5 Conclusion

The various applications of sensor-based spatio-temporal HAR in different sectors of life have made it a critical focus of research. In this systematic review, 37 articles covering spatio-temporal HAR based on sensors were identified through searches conducted on Scopus and Web of Science databases. These articles were selected from 256 papers after applying inclusion and exclusion criteria. Key information was extracted from each of the 37 articles to answer RQ1, RQ2, RQ3, and RQ4 and achieve our objectives.

Results show that LSTM and CNN are the most frequently employed models in spatio-temporal HAR based on sensors. Moreover, the 2D-CNN model is the most frequently used, suggesting its significant traction within the research community due to its effectiveness in capturing spatial features. Graph-based models, such as GNN, GCN, and self-attention mechanisms, have evolved in 2024 due to their ability to naturally represent spatial and temporal relationships.

The accuracy of the three datasets, PAMAP2, UCI-HAR, and WISDM, across the years 2020, 2022, 2023, and 2024 highlight their significance as the most used datasets in research. The review also highlighted that Smartphones are the most frequently mentioned sensor platform due to built-in sensors such as accelerometers, gyroscopes, and GPS, which provide accessible and versatile data for HAR. Followed by Inertial Measurement Unit (IMU) sensors, emphasizing their critical role in capturing precise motion data, especially in wearables and portable devices.

It was found that 82% of the datasets (with more than one repetition) consist of raw data, providing the most detailed and comprehensive information, but requiring significant processing. On the other hand, preprocessed data is ready for immediate use in machine learning models, reducing the time and effort required to develop HAR systems. Furthermore, 55% of the datasets are real-world datasets, while 45% are laboratory datasets, reflecting a balanced approach in HAR research.

By highlighting challenges and future directions, this review encourages further research on advanced models and innovative sensor applications. Overall, it supports standardization efforts and offers a comprehensive resource for researchers aiming to advance the field of spatio-temporal HAR.

## References

[1] C.-M. Forke and M. Tropmann-Frick, “Feature engineering techniques and spatio-temporal data processing,” *Datenbank-Spektrum*, vol. 21, pp. 237–244, 2021.

[2] S. Diwakar, D. Dwivedi, S. P. Singh, and M. Sharma, “Self-attention-based human activity detection using wearable sensors,” in *International Conference on Signals, Machines, and Automation*. Springer, 2022, pp. 629–636.

[3] S. M. Wagh, "Precise human activity recognition for the openpack challenge 2022," in *2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, 2023, pp. 259–261.

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

[5] E. Ferrara, "Large language models for wearable sensor-based human activity recognition, health monitoring, and behavioral modeling: A survey of early trends, datasets, and challenges," *Sensors*, vol. 24, no. 15, p. 5045, 2024.

[6] F. Al Machot and H. C. Mayr, "Improving human activity recognition by smart windowing and spatio-temporal feature analysis," in *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, 2016, pp. 1–7.

[7] R. Chen, T. Chu, K. Liu, J. Liu, and Y. Chen, "Inferring human activity in mobile devices by computing multiple contexts," *Sensors*, vol. 15, no. 9, pp. 21 219–21 238, 2015.

[8] A. Sarkar, S. S. Hossain, and R. Sarkar, "Human activity recognition from sensor data using spatial attention-aided cnn with genetic algorithm," *Neural Computing and Applications*, vol. 35, no. 7, pp. 5165–5191, 2023.

[9] D. Khan, A. Alshahrani, A. Almjally, N. Al Mudawi, A. Algarni, K. Al Nowaiser, and A. Jalal, "Advanced iot-based human activity recognition and localization using deep polynomial neural network," *IEEE Access*, 2024.

[10] Ł. Czekaj, M. Kowalewski, J. Domaszewicz, R. Kitłowski, M. Szwoch, and W. Duch, "Real-time sensor-based human activity recognition for efitness and ehealth platforms," *Sensors*, vol. 24, no. 12, p. 3891, 2024.

[11] H. Wu, Z. Zhang, X. Li, K. Shang, Y. Han, Z. Geng, and T. Pan, "A novel pedal musculoskeletal response based on differential spatio-temporal lstm for human activity recognition," *Knowledge-Based Systems*, vol. 261, p. 110187, 2023.

[12] A. Bolatov, A. Yessenbayeva, and A. Yazici, "Glula: Linear attention-based model for efficient human activity recognition from wearable sensors," *Wearable Technologies*, vol. 5, p. e10, 2024.

[13] A. Hamdi, K. Shaban, A. Erradi, A. Mohamed, S. K. Rumi, and F. D. Salim, "Spatiotemporal data mining: a survey on challenges and open problems," *Artificial Intelligence Review*, pp. 1–48, 2022.

[14] Y. Isoda, S. Kurakake, and K. Imai, "Ubiquitous sensor-based human behaviour recognition using the spatio-temporal representation of user states," *International Journal of Wireless and Mobile Computing*, vol. 3, no. 1-2, pp. 46–55, 2008.

[15] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–40, 2021.

[16] V. Bijalwan, V. B. Semwal, and V. Gupta, "Wearable sensor-based pattern mining for human activity recognition: Deep learning approach," *Industrial Robot: the international journal of robotics research and application*, vol. 49, no. 1, pp. 21–33, 2022.

[17] M. Straczkiewicz, P. James, and J.-P. Onnela, "A systematic review of smartphone-based human activity recognition methods for health research," *NPJ Digital Medicine*, vol. 4, no. 1, p. 148, 2021.

[18] E. De-La-Hoz-Franco, P. Ariza-Colpas, J. M. Quero, and M. Espinilla, "Sensor-based datasets for human activity recognition – a systematic review of literature," *IEEE Access*, vol. 6, pp. 59 192–59 210, 2018.

[19] J. Qi, P. Yang, A. Waraich, Z. Deng, Y. Zhao, and Y. Yang, "Examining sensor-based physical activity recognition and monitoring for healthcare using internet of things: A systematic review," *Journal of Biomedical Informatics*, vol. 87, pp. 138–153, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S153204641830176X

[20] M. M. Baig, S. Afifi, H. GholamHosseini, and F. Mirza, "A systematic review of wearable sensors and iot-based monitoring applications for older adults–a focus on ageing population and independent living," *Journal of medical systems*, vol. 43, pp. 1–11, 2019.

[21] A. Kristoffersson and M. Lindén, "A systematic review of wearable sensors for monitoring physical activity," *Sensors*, vol. 22, no. 2, p. 573, 2022.

[22] B. Kitchenham, S. Charters *et al.*, "Guidelines for performing systematic literature reviews in software engineering version 2.3," *Engineering*, vol. 45, no. 4ve, p. 1051, 2007.

[23] M. M. Islam and T. Iqbal, "Hamlet: A hierarchical multimodal attention-based human activity recognition algorithm," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 10 285–10 292.

[24] O. Konak, P. Wegner, and B. Arnrich, "Imu-based movement trajectory heatmaps for human activity recognition," *Sensors*, vol. 20, no. 24, p. 7179, 2020.

[25] X. Sun, H. Xu, Z. Dong, L. Shi, Q. Liu, J. Li, T. Li, S. Fan, and Y. Wang, "Capsganet: deep neural network based on capsule and gru for human activity recognition. ieee syst. j.(2022)."

[26] P. Khan, Y. Kumar, and S. Kumar, "Capslstm-based human activity recognition for smart healthcare with scarce labeled data," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 1, pp. 707–716, 2022.

[27] M. Sethi, M. Yadav, M. Singh, and P. G. Shambharkar, "Attnhar: Human activity recognition using data collected from wearable sensors," in *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*. IEEE, 2023, pp. 1–6.

[28] H. Najeh, C. Lohr, and B. Leduc, "Convolutional neural network bootstrapped by dynamic segmentation and stigmergy-based encoding for real-time human activity recognition in smart homes," *Sensors*, vol. 23, no. 4, p. 1969, 2023.

[29] S. Arokiaraj and N. Viswanathan, "Ecg-nets–a novel integration of capsule networks and extreme gated recurrent neural network for iot based human activity recognition," *Journal of Intelligent & Fuzzy Systems*, vol. 44, no. 5, pp. 8219–8229, 2023.

[30] Y. Li, R. Yin, Y. Kim, and P. Panda, "Efficient human activity recognition with spatio-temporal spiking neural networks," *Frontiers in Neuroscience*, vol. 17, p. 1233037, 2023.

[31] K. Monica and R. Parvathi, "Efficient gait analysis using deep learning techniques." *Computers, Materials & Continua*, vol. 74, no. 3, 2023.

[32] W. Xie, Q. Huang, C. Li, Y. Wang, and Y. Liu, "Hierarchical multi-scale adaptive conv-lstm network for human action recognition based on wearable sensors," in *Proceedings of the 5th ACM International Conference on Multimedia in Asia*, 2023, pp. 1–8.

[33] Q. Huang, W. Xie, C. Li, Y. Wang, and Y. Liu, "Human action recognition based on hierarchical multi-scale adaptive conv-long short-term memory network," *Applied Sciences*, vol. 13, no. 19, p. 10560, 2023.

[34] S. Mekruksavanich and A. Jitpattanakul, "Hybrid convolution neural network with channel attention mechanism for sensor-based human activity recognition," *Scientific Reports*, vol. 13, no. 1, p. 12067, 2023.

[35] H. Zhang and L. Xu, "Multi-stmt: multi-level network for human activity recognition based on wearable sensors," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[36] H. Zou, Z. Chen, C. Zhang, A. Yuan, B. Wang, L. Wang, J. Li, and Y. Pan, "Stfnet: Enhanced and lightweight spatiotemporal fusion network for wearable human activity recognition," *IEEE Sensors Journal*, 2024.

[37] F. Zeng, M. Guo, L. Tan, F. Guo, and X. Liu, "Wearable sensor-based residual multifeature fusion shrinkage networks for human activity recognition," *Sensors*, vol. 24, no. 3, p. 758, 2024.

[38] S. Mahmud, M. Tanjid Hasan Tonmoy, K. Kumar Bhaumik, A. Mahbubur Rahman, M. Ashraful Amin, M. Shoyaib, M. Asif Hossain Khan, and A. Ahsan Ali, "Human activity recognition from wearable sensor data using self-attention," in *ECAI 2020*. IOS Press, 2020, pp. 1332–1339.

[39] J. Wang, Y. Chen, and Y. Gu, "A wearable-har oriented sensory data generation method based on spatio-temporal reinforced conditional gans," *Neurocomputing*, vol. 493, pp. 548–567, 2022.

[40] S. Deng, Z. Guo, D. Teng, B. Lin, D. Chen, T. Jia, and H. Wang, "Self-relation attention networks for weakly supervised few-shot activity recognition," *Knowledge-Based Systems*, vol. 276, p. 110720, 2023.

[41] D. Liciotti, M. Bernardini, L. Romeo, and E. Frontoni, "A sequential deep learning application for recognising human activities in smart homes," *Neurocomputing*, vol. 396, pp. 501–513, 2020.

[42] N. T. H. Thu and D. S. Han, "Hihar: A hierarchical hybrid deep learning architecture for wearable sensor-based human activity recognition," *IEEE Access*, vol. 9, pp. 145 271–145 281, 2021.

[43] H. Najeh, C. Lohr, and B. Leduc, "Dynamic segmentation of sensor events for real-time human activity recognition in a smart home context," *Sensors*, vol. 22, no. 14, p. 5458, 2022.

[44] A. E. Nieto-Vallejo, C. A. Parra-Rodriguez, and O. Ramirez-Perez, "Classification of activities of daily living for older adults using machine learning and fixed time windowing technique," *IEEE Sensors Journal*, 2023.

[45] S. Mekruksavanich, N. Hnoohom, and A. Jitpattanakul, "A hybrid deep residual network for efficient transitional activity recognition based on wearable sensors," *Applied Sciences*, vol. 12, no. 10, p. 4988, 2022.

[46] Z. Zhang, W. Wang, A. An, Y. Qin, and F. Yang, "A human activity recognition method using wearable sensors based on convtransformer model," *Evolving Systems*, vol. 14, no. 6, pp. 939–955, 2023.

[47] J. Wu and Q. Liu, "A novel spatio-temporal network of multi-channel cnn and gcn for human activity recognition based on ban," *Neural Processing Letters*, vol. 55, no. 8, pp. 11 489–11 507, 2023.

[48] L. Bai, L. Yao, X. Wang, S. S. Kanhere, B. Guo, and Z. Yu, "Adversarial multi-view networks for activity recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 2, pp. 1–22, 2020.

[49] J. Ye, H. Hu, G.-J. Qi, and K. A. Hua, "A temporal order modeling approach to human action recognition from multimodal sensor data," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 13, no. 2, pp. 1–22, 2017.

[50] A. Sarkar, S. S. Hossain, and R. Sarkar, "Correction to: Human activity recognition from sensor data using spatial attention-aided cnn with genetic algorithm," *Neural Computing and Applications*, vol. 35, no. 16, pp. 12 239–12 239, 2023.

[51] S. Reshmi and E. Ramanujam, "An ensemble maximal feature subset selection for smartphone based human activity recognition," *Journal of Network and Computer Applications*, vol. 226, p. 103875, 2024.

[52] Y.-H. Tseng and C.-Y. Wen, "Hybrid learning models for imu-based har with feature analysis and data correction," *Sensors*, vol. 23, no. 18, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/18/7802