Applied Intelligence (2024) 54:8982–9007
https://doi.org/10.1007/s10489-024-05645-1

Deep learning for computer vision based activity recognition and fall detection of the elderly: a systematic review

F. Xavier Gaya-Morey[1,2,3] · Cristina Manresa-Yee[1,2,3] · José M. Buades-Rubio[1,2,3]

Accepted: 25 June 2024 / Published online: 8 July 2024

Abstract

As the proportion of elderly individuals in developed countries continues to rise globally, addressing their healthcare needs, particularly in preserving their autonomy, is of paramount concern. A growing body of research focuses on Ambient Assisted Living (AAL) systems, aimed at alleviating concerns related to the independent living of the elderly. This systematic review examines the literature pertaining to fall detection and Human Activity Recognition (HAR) for the elderly, two critical tasks for ensuring their safety when living alone. Specifically, this review emphasizes the utilization of Deep Learning (DL) approaches on computer vision data, reflecting current trends in the field. A comprehensive search yielded 2,616 works from five distinct sources, spanning the years 2019 to 2023 (inclusive). From this pool, 151 relevant works were selected for detailed analysis. The review scrutinizes the employed DL models, datasets, and hardware configurations, with particular emphasis on aspects such as privacy preservation and real-world deployment. The main contribution of this study lies in the synthesis of recent advancements in DL-based fall detection and HAR for the elderly, providing insights into the state-of-the-art techniques and identifying areas for further improvement. Given the increasing importance of AAL systems in enhancing the quality of life for the elderly, this review serves as a valuable resource for researchers, practitioners, and policymakers involved in developing and implementing such technologies.

Keywords Human activity recognition · Fall detection · Ambient assisted living · Deep learning · Computer vision · Elderly

1 Introduction

The global population is experiencing rapid growth, accompanied by a significant increase in life expectancy, particularly in developed countries. Bloom and Luca [1] note that life expectancy in China and India has surged by nearly 30 years since 1950. Consequently, a substantial portion of the population in developed nations, approximately 20%, is aged 60 and above, a figure projected to surpass 30% in the next four decades.

With this demographic shift comes a growing concern for elderly[1] care, as the need for assistance and support rises proportionately. Among the myriad challenges faced

F. Xavier Gaya-Morey, Cristina Manresa-Yee and José M. Buades-Rubio contributed equally to this work

✉ F. Xavier Gaya-Morey
francesc-xavier.gaya@uib.es

Cristina Manresa-Yee
cristina.manresa@uib.es

José M. Buades-Rubio
josemaria.buades@uib.es

[1] Group of Computer Graphics, Computer Vision and AI (UGIVIA), Universitat de les Illes Balears (UIB), Carretera de Valldemossa, km 7.5, Palma 07122, Illes Balears, Spain

[2] Research Institute of Health Sciences (IUNICS), Universitat de les Illes Balears (UIB), Carretera de Valldemossa, km 7.5, Palma 07122, Illes Balears, Spain

[3] Department of Mathematics and Computer Science, Universitat de les Illes Balears (UIB), Carretera de Valldemossa, km 7.5, Palma 07122, Illes Balears, Spain

[1] In this work, the term "elderly" refers to individuals over 65 years old, as commonly used in the field of Computer Science. However, terminology may vary in other fields; for instance, the term "older adults" is often preferred in Psychology.

by the elderly, falls represent a particularly prevalent and perilous occurrence. The World Health Organization highlights alarming statistics on falls, identifying them as the second leading cause of unintentional injury deaths worldwide. Each year, an estimated 684,000 individuals succumb to fall-related injuries globally, with an additional 37.3 million falls severe enough to necessitate medical attention [2]. Apart from the physical harm incurred by the elderly, the economic ramifications are substantial, with fall-related treatment costs comprising a significant portion of healthcare expenditures in various countries such as the USA, Australia, EU15 and the United Kingdom [3].

Automated fall detection for the elderly is feasible through data collected from wearable or environmental devices, such as accelerometers, gyroscopes, and cameras. Furthermore, Human Activity Recognition (HAR) holds promise for diverse applications, ranging from automatic life-logging to identifying patterns indicative of illness [4, 5]. Vision data from cameras is increasingly utilized for fall detection and HAR tasks due to its numerous advantages over wearable devices or other sensors. These advantages include the ability to detect multiple events simultaneously, suitability for various subjects, environments, and tasks, as well as ease of installation and visual verification of data [6].

From an algorithmic standpoint, Deep Learning (DL) has revolutionized digital image processing, emerging as the state-of-the-art approach in numerous domains [7]. Over recent years, a plethora of DL architectures have been developed and evaluated for computer vision tasks, prompting the exploration of suitable models for HAR and fall detection among older adults.

In this study, we conduct a Systematic Literature Review (SLR) focusing on DL-based HAR and fall detection using vision data for elderly care. Our review strictly adheres to the guidelines outlined for conducting SLRs in Software Engineering by Kitchenham and Charters [8], providing a structured methodology and rigorous analysis. The document is structured as follows: we first delve into the background of the study, encompassing previous reviews and defining key concepts; we then enumerate the review questions; next, we elaborate on the review methods, detailing data sources, search strategy, study selection, quality assessment, and data extraction; subsequently, we analyze the resulting studies comprehensively to address the review questions; the discussion section synthesizes our findings and addresses the review questions; finally, we present the conclusions derived from the SLR.

2 Background

In accordance with the guidelines provided by Kitchenham and Charters [8], it is imperative to summarize previous reviews prior to conducting the SLR, thereby substantiating its necessity. Hence, we briefly outline related reviews and surveys from the past three years, as older reviews cannot encompass the most recent studies. The full list can be found in Table 1, providing a visual comparison of the main disparities.

Guerra et al. [9] studied the current state-of-the-art of Ambient Assisted Living (AAL) for frail individuals, including the elderly and disabled, encompassing both wearable and non-wearable solutions. They explored common steps in the Human Activity Recognition (HAR) processing chain. Similarly, Kumar et al. [10] delved into various types of data used for HAR, elucidating common datasets, approaches, and challenges, but not including the elderly population or fall detection. In [11], Tay et al. investigated abnormal behavior detection, such as fall detection, repetition of activities, and accidents. They explored multiple solutions, including visual and wearable sensors, and both conventional and Deep Learning (DL) approaches. A review by Momin et al. [12] explores activity pattern monitoring using depth sensors, considering this visual data as a privacy-preserving alternative to RGB video or images for older adults. The studies are categorized based on the computing technique utilized and the datasets used are analyzed. Olugbade et al. [13] conducted a scoping review on datasets utilized for HAR and fall detection, resulting in an extensive compilation of over 700 datasets of various modalities. Multiple taxonomies were developed to categorize the datasets by population groups, data types, and creation purposes, among others, although only four datasets include elderly subjects. Alam et al. [14] conducted a review specifically on DL-based fall detection systems, analyzing different fall types, popular datasets, evaluation metrics, and architectural variations. Rastogi et al. [15] reviewed a broader range of tasks, including falls and other relevant information extracted from video sequences, such as body shape changes, posture, and gait. Another relevant review is by Gutiérrez et al. [16], also centered on fall detection, which describes common processing steps, ML models, datasets, metrics, and tracking techniques, with most studies utilizing RGB and depth data.

While prior reviews have addressed various aspects of our research domain, notable differences underscore the necessity of our study. Table 1 sheds light upon this by displaying the pivotal aspects considered in the current review, along with whether they are addressed or not in the aforementioned reviews.

The sole review exclusively focusing on DL techniques was conducted by Alam et al. [14], which, however, omitted HAR from its scope, thus neglecting a significant portion of studies included in our analysis. In contrast, other reviews encompassed techniques employing handcrafted features or classical vision approaches, reflecting a broader scope than our exclusive focus on DL-based solutions. Furthermore,

Table 1 Comparison of previous reviews with ours

Work	Year	System	Focus			Task		Data			AAL		
			DL	Elderly	CV	FD	HAR	RGB	Depth	IR	Privacy	HW	Deployment
Ours	2024	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
[9]	2023			✓		✓	✓	✓	✓	✓	✓		
[10]	2023						✓	✓	✓				
[11]	2023			✓		✓		✓	✓	✓			
[12]	2022			✓	✓	✓	✓		✓		✓		
[13]	2022	✓				✓	✓	✓	✓	✓			
[14]	2022		✓	✓	✓	✓		✓	✓	✓	✓		
[15]	2022			✓	✓	✓	✓	✓	✓	✓			
[16]	2021			✓	✓	✓		✓	✓	✓			

By columns, important aspects taken into account in this review, and whether they are addressed or not by each review. From left to right: if the review is systematic; focuses on the exploration of DL solutions; targets elderly people as the users of the system; centers on the use of vision data; explores the Fall Detection and the Human Activity Recognition tasks; explores the use of RGB, depth or infrared data; takes privacy as a critical concern; describes the hardware used in the found studies; and if it studies the deployment of the FD and HAR systems in real environments

previous reviews often overlooked the importance of studying DL-related nuances, such as the significance of training datasets, architectural considerations, and feature extraction methods. In our review, we meticulously categorize and elucidate these nuances through a comprehensive taxonomy of identified techniques.

Another notable observation is the limited attention given to HAR in several previous reviews, with some omitting the task altogether. As a result, our review unveils a greater number of studies dedicated to fall detection and HAR in the elderly. Additionally, our analysis delves deeper into the intricacies of these tasks, providing a more comprehensive understanding.

Moreover, only a few reviews explored applications within AAL systems and the associated privacy implications. Hardware specifications, beyond the prevalent use of Kinect cameras, were rarely examined, and the effective deployment of fall detection or HAR systems was not thoroughly explored. In contrast, our review emphasizes these aspects, which are pivotal in facilitating the transference to society.

Finally, it is worth noting that, apart from [13], none of the previous reviews adhered to a systematic review process. By rigorously following the systematic review methodology outlined by Kitchenham and Charters [8], our study ensures a robust and unbiased selection and analysis of relevant studies. We conducted a comprehensive search across various databases, employing well-defined search strings aligned with our research questions. Each study underwent careful quality assessment, and strict exclusion criteria were applied to ensure the inclusion of only the most relevant and high-quality literature. This systematic approach minimizes potential biases and ensures that our review is based on a well-rounded selection of literature.

3 Review questions

As outlined in [8], specifying the research questions is a critical aspect of any systematic review, as they guide the entire methodology: from the search process identifying primary studies to address them, to the data extraction process extracting the required data items, and finally to the data analysis synthesizing the data to answer the questions. The review questions for this study are presented in Table 2.

The first research question, RQ1, aims to identify the methods used to recognize activities or detect falls among elderly individuals. The choice to specifically investigate HAR and fall detection stemmed from an exploratory initial search, where they emerged as the two most relevant recognition tasks in AAL for the elderly. Given that visual data offers numerous advantages over other sensor data types,

Table 2 Primary and secondary research questions used for this SLR

ID	Research question
RQ1	**What computer vision deep learning techniques are used for human activity recognition and fall detection on elderly people?**
RQ1.1	What is the preferred data type?
RQ1.2	What are the most extensively used architectures?
RQ1.3	What are the most extended datasets?
RQ2	**How can these tasks be deployed successfully in a real environment?**
RQ2.1	What is the most common hardware (cameras, robots, etc.)?
RQ2.2	How is privacy of the elderly preserved?

such as visual verification and simultaneous subject recognition, and DL has become the state-of-the-art approach in computer vision, conducting an in-depth analysis of the most prevalent methods with these characteristics is crucial for informing future research in this domain. Furthermore, three research subquestions are included regarding common data types (e.g., RGB, depth, thermal, etc.), DL architectures (e.g., CNN, RNN, etc.), and datasets found in the reviewed literature. These subquestions aim to delve deeper into the solution choices at different design steps, which are closely related to various requirements such as privacy preservation, result stability, and inference speed.

The second research question, RQ2, emerges as a significantly unexplored area, as highlighted in Table 1 of the Background section. Many previous reviews have focused on the recognition phase of previous studies, enumerating common methods, processing steps, and datasets. However, the effective deployment in real-world scenarios is pivotal for the transfer of such methods to society, and this aspect remains largely unexplored. Works with implementations in real environments, whether through the use of assistive robots or camera-based setups, are expected to be found among the selected studies. Therefore, it is desirable to explore their design choices, setups, and encountered challenges in greater depth. Additionally, privacy is a particularly concerning aspect to consider when dealing with users, especially when utilizing visual data from cameras, and the approaches to addressing it are of interest for future research. For these reasons, RQ2.1 and RQ2.2 delve into common hardware choices and privacy preservation strategies.

4 Review methods

In this section, we provide a detailed description of the systematic review protocol followed, based on the guidelines outlined by Kitchenham and Charters [8]. Firstly, we list and analyze the primary data sources used, providing visualization of the distribution of studies among these sources. Next, we define the search strategy, which encompasses search terms, synonyms, and time restrictions. Following this, we establish criteria for inclusion and exclusion of studies, followed by the design of a quality assessment checklist to identify and remove low-quality studies. Finally, in the data extraction and synthesis stage, we define how information from each primary study is obtained and outline the specific attributes considered of interest.

4.1 Data sources

For this systematic review, we selected five primary data sources: SCOPUS, Web of Science (WOS), IEEE Xplore Digital Library, ACM Digital Library, and PubMed.

SCOPUS and WOS were chosen as comprehensive digital libraries covering a wide range of disciplines, while IEEE Xplore focuses on engineering and technology, ACM Digital Library specializes in computer science, and PubMed is centered on biomedical studies. This selection ensures the inclusion of relevant literature from diverse domains, maximizing the breadth of content considered in our review.

The distribution of studies retrieved from each source is illustrated in Fig. 1. As depicted, the majority of studies were sourced from ACM and SCOPUS, with only a small fraction (110 out of a total of 2,616) obtained from PubMed.

4.2 Search strategy

We constructed different query strings tailored to match the syntax of each digital library while minimizing differences and employing consistent synonyms for the concepts being searched. Each query string connected the various concepts using logical AND, while synonyms for each concept were connected with logical OR. To account for inflection of certain keywords, we utilized the "*" operator after the root word to allow for any possible word endings. In the SCOPUS library, the search was restricted to titles, abstracts, or keywords due to the impracticality of retrieving results otherwise, with the majority being poorly relevant. Conversely, the entire text was searched for in the remaining databases. The primary concepts searched, along with their corresponding lists of synonyms, are as follows:

- **Task to perform (activity recognition or fall detection):** *"action recognition"* OR *"activit* recognition"* OR *"fall* detection"* OR *"behaviour recognition"* OR *"behaviour detection"* OR *"physical activity recognition"*

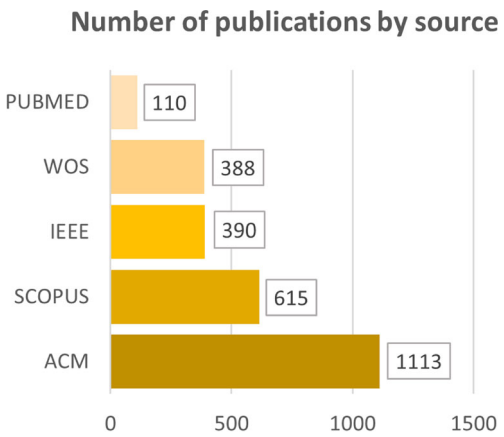


Fig. 1 Number of publications obtained from each database, before duplicate removal and study selection (2616 in total)

- **Ambient Assisted Living:** *"monitoring"* OR *"assist* living"* OR *"AAL"* OR *"smart home"* OR *"activit* of daily life"* OR *"activit* of daily living"* OR *"ADL"*
- **Target collective (elderly people):** *"elder*"* OR *"old* people"* OR *"senior"*
- **Kind of data used (Computer Vision):** *"vision"* OR *"rgb"* OR *"video"* OR *"image"* OR *"skeleton"* OR *"depth"* OR *"camera"* OR *"gesture"*

Initially, we included studies published from 2013 onwards in the search. However, upon further examination, we observed that the majority of relevant studies were published recently. Consequently, we decided to limit the review to the last five years. Figure 2 displays the accumulated relevant studies from 2013 to 2023. As depicted, only 19 relevant articles were found during the first six years, while 151 were discovered in the last five. This trend underscores the increasing significance of DL-based strategies for HAR and fall detection. By focusing on studies published in the last five years, we aim to gain a deeper analysis of recent trends.

4.3 Study selection

After collecting studies from various sources, limiting by year, and removing duplicates, exclusion criteria were applied to eliminate non-relevant studies. The exclusion criteria were as follows:

- **Deep Learning**: Studies not utilizing DL were considered irrelevant for this review. Including this criterion in the exclusion criteria rather than in the query strings enabled the inclusion of more relevant studies, since many studies did not directly reference DL but instead used the name of a specific model.
- **Language**: Studies not in English or Spanish were excluded.
- **Data Type**: Studies using data types other than RGB, depth, or IR were excluded. This includes both videos and images. Skeleton data was also included, but only if computed from the other three types of data. Studies using sensory data along with visual data were also included, allowing for multimodal approaches.
- **Accessibility**: Studies not accessible for various reasons, such as being part of paid content (e.g., book chapters), source website down, or retracted content, were excluded.
- **Redundancy**: In cases where a journal article extended a work already presented in a conference, the conference proceedings publications were omitted, as the journal article represented an extension of the same work.
- **Task**: Studies focused on tasks other than HAR or fall detection, such as velocity estimation, gait trend, level of tiredness, etc., were excluded. However, studies that did not directly perform HAR or fall detection but presented a new dataset for these tasks were included.
- **Target Collective**: Studies not centered on elderly people were excluded. Merely mentioning the elderly as one of the beneficiaries of the work was insufficient; the study had to either use data from elderly people or have them in mind when designing the experiment.
- **Works in Progress**: Conference proceedings about works in progress, containing only the initial stages of the study and lacking the experimentation phase, were excluded.
- **Quality**: Publications with very poor quality (e.g., null reproducibility, highly biased decisions, too small datasets, etc.) were excluded. More information about quality assessment can be found in Section 4.4.

The results of study collection and duplicate removal are illustrated in Fig. 3. A total of 2,616 studies were collected from the different sources using the aforementioned queries, of which 633 duplicates were detected and removed, leaving a total of 1,983 studies.

As depicted in Fig. 3, only 151 studies remained after applying the exclusion criteria, comprising 89 conference pro ceedings and 64 journal articles. The conference proceedings

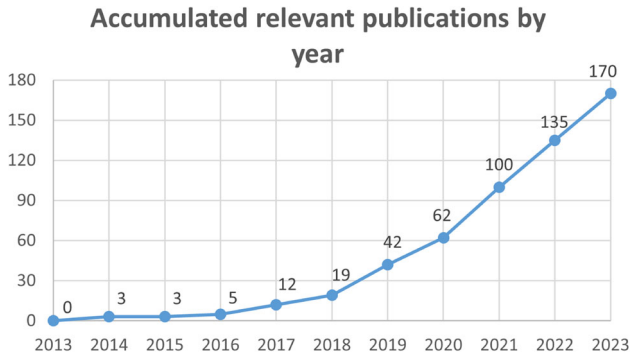


Fig. 2 Accumulated publications from 2013 to 2023 (both included) after study selection and quality assessment

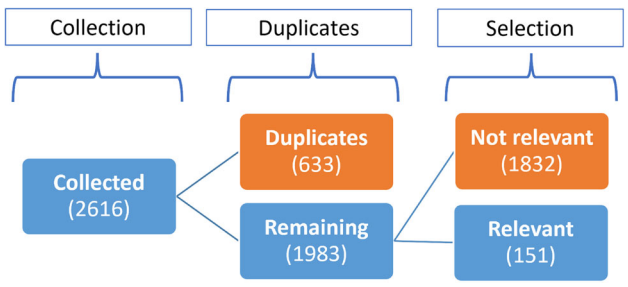


Fig. 3 Articles collected at each phase of the systematic search process, including acquisition from various sources, removal of duplicates, and study selection, which also involved quality assessment

were retained for analysis among the relevant studies, as they serve as a standard search strategy to address publication bias, which can lead to systematic bias in systematic reviews unless special efforts are made to address this issue [8].

4.4 Quality assessment

Given the absence of a universally agreed-upon definition of study "quality," the proposed guidelines in [8] were adhered to, primarily focusing on bias and validity as measures of quality. Specifically, the following aspects were taken into account:

- **Reproducibility**: Assessing whether the work can be replicated. This can be achieved by disclosing the dataset used, using external datasets, and either publishing the code used for the model or providing sufficient details to recreate the model.
- **Comparison with Other Works**: Evaluating whether the performance of the model is compared with the state-of-the-art. It's essential to ensure that comparisons are made under fair conditions, meaning that the models should be trained and tested on the same data to avoid introducing bias.
- **Use of External Datasets**: Considering whether the model is tested on external datasets to mitigate possible bias from the data and facilitate comparison with other models for the same task. Additionally, using external datasets allows other studies to utilize the results without the need to retrain the model on different data.

These aspects were included in the list of fields during the data extraction phase (last three fields), as discussed in Section 4.5. Moreover, these quality aspects were also used as exclusion criteria, as previously mentioned in Section 4.3.

In addition to these aspects, the type of study, either conference proceedings or journal articles, was also considered as a quality indicator, with journal articles typically being longer and more mature.

4.5 Data extraction and synthesis

From each study remaining after applying the exclusion criteria, various data points were extracted to summarize the content and establish taxonomies for various aspects of interest. All data were compiled into a table, with each entry containing the following fields:

- Title
- Author/s
- Type (journal article or conference proceedings)
- Publication year
- Task (HAR or fall detection)
- Data type (RGB, depth, IR or skeleton)
- Auxiliary sensor data type (accelerometers, gyroscopes, etc.)
- Camera used
- Dataset (name of external dataset/s or "custom")
- DL model/s and task (skeleton joints estimation, feature extraction, classification, etc.)
- Other ML models or computer vision techniques used
- System integration in a robot (yes/no) and which one
- System integration in a framework (yes/no)
- How is privacy preserved? (depth or IR only, low resolution, etc.)
- Reproducible (yes/no)
- Test with external datasets
- Comparison with other approaches

The complete list of relevant studies is provided in Tables 4 and 3, which display only basic information for each study. The remaining information will be synthesized in Section 5 through tables and plots, allowing for an overview of the distribution of works by used data types, DL model families, datasets, etc. Additionally, particularly relevant or interesting aspects of the works will be summarized, and important concepts will be addressed in more detail.

5 Results

This section provides an overview of the primary studies discovered through the systematic search process and presents the findings. Each study is thoroughly examined, and summaries are presented in the form of tables and graphs where applicable. Subsections are structured to address individual research questions, enhancing readability and organization.

5.1 RQ1: fall detection and human activity recognition

The review primarily focuses on two main tasks: fall detection and Human Activity Recognition (HAR). It is worth noting that fall detection can be viewed an especially important activity of HAR. As illustrated in Fig. 4, fall detection has received the most attention in the past five years, with a total of 72 studies, while HAR has been explored in 52 studies. This discrepancy highlights the significance of fall detection when concerning the elderly population. Many works emphasize the importance of accurately and swiftly identifying falls among the elderly, given the potential for injuries and health implications if prompt actions are not taken. Consequently, several studies mention integrating fall detection

Table 3 Full list of relevant studies examined in this systematic review

Ref.	Year	Task	CV Data	Ref.	Year	Task	CV Data
[17]	2023	FD	D	[18]	2022	FD	RGB
[19]	2023	FD	RGB	[20]	2022	FD	RGB/D/IR
[21]	2023	FD	RGB	[22]	2022	FD	RGB-D
[23]	2023	FD	RGB	[24]	2022	FD & HAR	RGB
[25]	2023	FD	RGB	[26]	2022	FD & HAR	RGB
[27]	2023	FD	RGB	[28]	2022	FD & HAR	RGB
[29]	2023	FD	RGB	[30]	2022	FD & HAR	RGB
[31]	2023	FD	RGB	[32]	2022	FD & HAR	RGB
[33]	2023	FD	RGB	[34]	2022	FD & HAR	RGB
[35]	2023	FD	RGB	[36]	2022	HAR	D
[37]	2023	FD	RGB	[38]	2022	HAR	D
[39]	2023	FD	RGB	[40]	2022	HAR	D
[41]	2023	FD	RGB	[42]	2022	HAR	RGB
[43]	2023	FD & HAR	D	[44]	2022	HAR	RGB
[45]	2023	FD & HAR	IR	[46]	2022	HAR	RGB
[47]	2023	FD & HAR	RGB	[48]	2022	HAR	RGB
[49]	2023	FD & HAR	RGB	[50]	2022	HAR	RGB
[51]	2023	FD & HAR	RGB	[52]	2022	HAR	RGB
[53]	2023	FD & HAR	RGB	[54]	2022	HAR	RGB
[55]	2023	FD & HAR	RGB	[56]	2022	HAR	RGB
[57]	2023	FD & HAR	RGB	[58]	2022	HAR	RGB/D
[59]	2023	FD & HAR	RGB	[60]	2022	HAR	RGB-D
[61]	2023	FD & HAR	RGB	[62]	2021	FD	RGB
[63]	2023	FD & HAR	RGB	[64]	2021	FD	RGB
[65]	2023	FD & HAR	RGB	[66]	2021	FD	RGB
[67]	2023	FD & HAR	RGB	[68]	2021	FD	RGB
[69]	2023	FD & HAR	RGB	[70]	2021	FD	RGB
[71]	2023	HAR	RGB	[72]	2021	FD	RGB
[73]	2023	HAR	RGB	[74]	2021	FD	RGB
[75]	2023	HAR	RGB	[76]	2021	FD	RGB
[77]	2023	HAR	RGB	[78]	2021	FD	RGB
[79]	2023	HAR	RGB	[80]	2021	FD	RGB
[81]	2023	HAR	RGB/D	[82]	2021	FD	RGB
[83]	2023	HAR	RGB-D	[84]	2021	FD	RGB
[85]	2023	HAR	RGB-D	[86]	2021	FD	RGB
[87]	2022	FD	D	[88]	2021	FD	RGB
[89]	2022	FD	RGB	[90]	2021	FD	RGB
[91]	2022	FD	RGB	[92]	2021	FD	RGB
[93]	2022	FD	RGB	[94]	2021	FD	RGB
[95]	2022	FD	RGB	[96]	2021	FD	RGB
[97]	2022	FD	RGB	[98]	2021	FD	RGB/D
[99]	2022	FD	RGB	[100]	2021	FD & HAR	RGB
[101]	2022	FD	RGB	[102]	2021	FD & HAR	RGB
[103]	2022	FD	RGB	[104]	2021	FD & HAR	RGB
[105]	2022	FD	RGB	[106]	2021	FD & HAR	RGB
[107]	2022	FD	RGB	[108]	2021	FD & HAR	RGB
[109]	2022	FD	RGB	[110]	2021	HAR	D
[111]	2022	FD	RGB	[112]	2021	HAR	D

Table 3 continued

Ref.	Year	Task	CV Data	Ref.	Year	Task	CV Data
[113]	2021	HAR	D	[114]	2020	HAR	RGB
[115]	2021	HAR	D	[116]	2020	HAR	RGB
[117]	2021	HAR	D	[118]	2020	HAR	RGB-D
[119]	2021	HAR	IR	[120]	2020	HAR	RGB-D
[121]	2021	HAR	RGB	[122]	2019	FD	IR
[123]	2021	HAR	RGB	[124]	2019	FD	RGB
[125]	2021	HAR	RGB	[126]	2019	FD	RGB
[127]	2021	HAR	RGB	[128]	2019	FD	RGB
[129]	2021	HAR	RGB	[130]	2019	FD	RGB
[131]	2021	HAR	RGB	[132]	2019	FD	RGB
[133]	2021	HAR	RGB/D	[134]	2019	FD	RGB
[135]	2021	HAR	RGB/D	[136]	2019	FD	RGB
[137]	2020	FD	RGB	[138]	2019	FD	RGB
[139]	2020	FD	RGB	[140]	2019	FD	RGB
[141]	2020	FD	RGB	[142]	2019	FD	RGB
[143]	2020	FD	RGB	[144]	2019	FD	RGB, IR
[145]	2020	FD	RGB	[146]	2019	FD	RGB-D
[147]	2020	FD	RGB	[148]	2019	HAR	D
[149]	2020	FD	RGB	[150]	2019	HAR	D
[151]	2020	FD	RGB	[152]	2019	HAR	D
[153]	2020	FD	RGB	[154]	2019	HAR	D
[155]	2020	FD	RGB-D	[156]	2019	HAR	RGB
[157]	2020	FD	RGB-D	[158]	2019	HAR	RGB
[159]	2020	FD & HAR	IR	[160]	2019	HAR	RGB
[161]	2020	FD & HAR	RGB	[162]	2019	HAR	RGB
[163]	2020	HAR	RGB	[164]	2019	HAR	RGB
[165]	2020	HAR	RGB	[166]	2019	HAR	RGB
[167]	2020	HAR	RGB				

The tasks of the studies may involve FD (fall detection), HAR (Human Action Recognition), or both (FD, HAR). The CV data column indicates the type of computer vision data used: RGB, D (depth), or IR (infrared). Data types are separated by commas if the study requires all of them or by slashes if only one is necessary

into systems or applications capable of alerting medical personnel [41, 49].

Only 27 out of 151 studies (approximately 18%) address both tasks simultaneously. This disparity arises from the emphasis placed on fall detection compared to other activities (such as walking or standing up), as well as the limited availability of data concerning fall scenarios, often resulting in an imbalanced problem. However, some studies manage to address both tasks. For instance, in [24, 67, 104], both tasks are computed using the UP-FALL dataset [168], which includes five types of falls and six common activities. This balanced dataset allows for the preservation of the importance of accurately detecting falls amidst other activities. A similar approach is adopted in [106], where a custom dataset with egocentric videos is utilized. Nevertheless, there are studies that treat falls as just another task to recognize [30, 47, 161].

5.2 RQ1.1: data type

Among the studies collected, three types of vision data were considered: RGB, depth, and infrared (IR). The distribution of these data types is illustrated in Fig. 5. RGB data were the most prevalent for fall detection and HAR among the elderly (132 studies), followed by depth data (30 studies), with IR data being the least utilized (6 studies). This discrepancy can primarily be attributed to the accessibility of common cameras compared to specialized ones equipped with depth or infrared sensors. Additionally, RGB cameras offer benefits such as lower costs and easier visual data inspection. Notably, infrared cameras are less frequently employed, typically positioned overhead (top-down perspective) and characterized by very low resolutions, allowing for the use of simpler CNN models [45, 122], as well as non-convolutional models like LSTM [119, 159] and Transformer [119]. Depth cam-

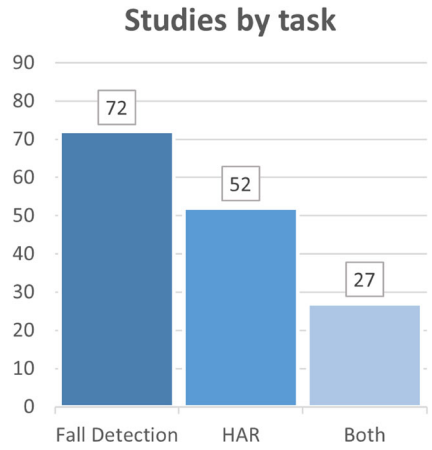


Fig. 4 Distribution of studies by target task: fall detection, HAR or both

eras are more commonly used than infrared ones, although they are often employed to extract skeleton joints rather than directly performing fall detection and HAR. Specifically, 67% of studies utilizing depth data computed skeleton joints before classification [38, 152, 152], while the remaining 33% did not [20, 148, 155].

Skeleton poses and sequences emerged as prevalent data types across the reviewed studies, with 67 studies incorporating skeleton data in some form. Given the human-centric nature of HAR and fall detection tasks, skeletal data represent logical features, offering efficient information compression while maintaining interpretability. Skeletons are typically represented as ordered sets of coordinates of body landmarks, either in 2D [24, 61, 107] or 3D [22, 38, 152] positions, depending on whether they were estimated from RGB or depth data, respectively. When skeleton estimation is performed on videos, the result is a sequence of skeleton poses with an added temporal dimension, enabling exploration of pose evolution over time intervals. 35 studies employed the evolution of one or more body landmarks for fall or HAR recognition [49, 107], while the remaining 32 studies performed recognition using static poses exclusively [42, 118].

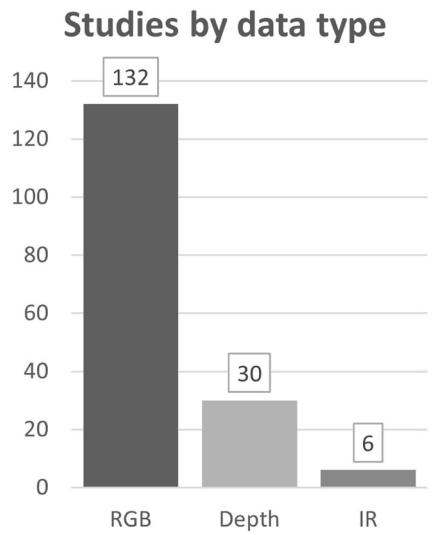


Fig. 5 Distribution of studies by data type used. Note that more than one data type was used in some studies

In addition to vision data, some studies utilized sensor data to enhance system performance, employing different models or strategies for classification and subsequently fusing the results. Fourteen studies, listed in Table 4, utilized at least one of five types of sensor data, including Inertial Measurement Unit (IMU)[2], audio, barometer, luminosity, radar, electrocardiogram (ECG), GPS, and network traffic. IMU data was the most commonly used, featured in 10 of the 14 studies, particularly for fall detection (in 6 out of 10 studies using IMUs), owing to its effectiveness in identifying abrupt movements and subsequent immobility [64, 149]. Barometer, GPS, radar, luminosity, and ECG data were consistently employed in conjunction with IMU data. Barometer and luminosity data served to acquire auxiliary or redundant information to enhance recognition consistency [74, 149]. ECG data in [47] was utilized to identify inconsistencies in recognition and trigger specific further computations. In [85], four types of data (IMU, audio, radar, and GPS), along with visual data, were used for federated learning, where independent models were trained using different data modalities.

Regarding data fusion, no instances of early fusion were found. Instead, intermediate (7 studies) and late (5 studies) fusion methods were prevalent. Late fusion involved using a model for each data modality to produce a classification result, with the final classification determined using either voting [74, 89] or weight attribution methods [44, 69, 149]. In intermediate fusion, different models extracted features from various modalities, with a final model performing classification based on concatenated feature inputs. Various final models were utilized, including CNN [64], fully connected layers [54, 72], SVM [99], and stacked classifiers [77]. In two studies, no fusion was performed, with different options provided for classification using distinct data modalities [47, 127].

5.3 RQ1.2: DL models

Table 5 provides a summary of all DL models utilized in the analyzed studies. These models are often employed for various specific tasks, including skeleton joints estimation, optical flow computation, and feature extraction. Moreover, the input data for these models encompasses not only images or videos but also features frequently computed by other DL

[2] Studies not specifying the use of an IMU but using accelerometers and gyroscopes have also been included in this group.

Table 4 Studies using multi-modal approaches and type of fusion with visual data

Ref.	Year	Task	Sensors	Fusion type
[85]	2023	HAR	IMU, Audio, Radar, GPS	Intermediate
[47]	2023	HAR, FD	IMU, ECG	None
[77]	2023	HAR	Network traffic	Intermediate
[69]	2022	HAR, FD	CSI	Late
[89]	2022	FD	IMU	Late
[44]	2022	HAR	IMU	Late
[99]	2022	FD	IMU	Intermediate
[54]	2022	HAR	IMU	Intermediate
[60]	2022	HAR	IMU, Radar	Intermediate
[127]	2021	HAR	Audio	None
[64]	2021	FD	IMU	Intermediate
[74]	2021	FD	IMU, Luminosity	Late
[149]	2020	FD	IMU, Barometer	Late
[148]	2019	HAR	Audio	Intermediate

Table 5 DL models utilized in the reviewed studies, tasks they are employed for, input data they process, and number of studies in which they are featured

Model Ref.	Name	Studies Task	Input data	N
[169]	OpenPose	2D Skeleton	RGB Image	25
[170]	AlphaPose	2D Skeleton	RGB Image	10
[171]	MediaPipe	2D Skeleton	RGB Image	4
[172]	PoseNet	2D Skeleton	RGB Image	3
[173]	MoveNet	2D Skeleton	RGB Image	2
[174]	RMPE	2D Skeleton	RGB Image	1
[175]	PoseFlow	2D Skeleton	RGB Image	1
	Baidu AI	2D Skeleton	RGB Image	1
[176]	FastPose	2D Skeleton	RGB Image	1
[177]	MobileNet	2D Skeleton, HAR, FD	RGB Image	7
[178]	DeepHAR	2D Skeleton, HAR, FD	RGB Image	1
[179]	PoseConv3D	3D Skeleton	Depth Image	1
[180]	STN	3D Skeleton	RGB-D Image	1
[181]	Autoencoder	FD	Different features	3
[182]	GAN	FD	Different features	2
[183]	Siamese CNNs	FD	RGB or Optical Flow Video	1
[97]	FallNet	FD	RGB Video	1
[184]	DeepFall	FD	RGB, Depth or IR Video	1
[185]	Sep-TCN	FD	Skeleton sequence	1
[186]	DCF-Net	Features	RGB Image	1
[187]	SqueezeNet	Features	RGB Image	1
[188]	EfficientNet	Features	RGB Image	1
[189]	C3D	Features	RGB Video	1
[190]	R-CNN	Features (OD, OS), FD	RGB Image	7
[191]	YOLO	Features (OD), HAR, FD	RGB Image	26
[67]	MSSkip	Features (OS)	RGB Image	1
[192]	PointRend	Features (OS)	RGB Image	1
[193]	LiteFlowNet	Features (Optical Flow)	RGB Video	1
[194]	Slowfast	Features, HAR	RGB Video	3

Table 5 continued

Model Ref.	Name	Studies Task	Input data	N
[195]	InceptionV3	Features, HAR	RGB or Depth Image	1
	CNN	Features, HAR, FD	Image, Video or different features	52
[196]	LSTM	Features, HAR, FD	Skel. sequence or Video features	29
[197]	VGG	Features, HAR, FD	RGB or Depth Image	13
[198]	ResNet	Features, HAR, FD	RGB or Depth Image or Skel. Pose	12
[199]	GCN	Features, HAR, FD	Skel. sequence or Video features	9
	RNN	Features, HAR, FD	Skel. sequence or Video features	7
[200]	I3D	Features, HAR, FD	RGB Video	5
[201]	GRU	Features, HAR, FD	Skel. sequence or Video features	4
[202]	Transformer	HAR	IR Image (8x8)	2
[203]	TANet	HAR	RGB Video	2
[204]	TPN	HAR	RGB Video	2
[205]	iCAN	HAR	Bounding Boxes sequence	1
[206]	Xception	HAR	RGB Image	1
[207]	TSN	HAR	RGB Video	1
[208]	VST	HAR	RGB Video	1
[209]	TimeSformer	HAR	RGB Video	1
[210]	Glimpse Clouds	HAR	Skeleton sequence	1
[211]	AIA	HAR	Video features and Bounding Boxes	1
	MLP	HAR, FD	Different features	5
[212]	AlexNet	HAR, FD	Feature Image	1
[28]	ARFD-Net	HAR, FD	Skeleton sequence	1

Abbreviations are used for Fall Detection (FD), Object Detection (OD), and Object Segmentation (OS). Models of to the same family are grouped together, including R-CNN, LSTM, YOLO, VGG, ResNet, RNN, CNN, and GCN

models, such as 2D or 3D skeleton poses and optical flow. A taxonomy of the identified DL models, based on different characteristics, is presented in Figure 6, offering a total count for each category. There is considerable diversity in the utilization of these models, regarding datasets used for evaluation, data types, and methodology. As demonstrated in the next section, in Table 6, a wide range of datasets was employed across the analyzed studies, with many utilizing custom datasets. Additionally, prominent datasets like URFD and UP-FALL offer various data types, including RGB recordings, depth, skeletons, accelerometers, etc., which may lead to data differences even when studies are evaluated on the same dataset. The methodology for training and testing DL methods also varies across studies, with some employing k-fold cross-validation, leave-one-out cross-validation, or no cross-validation at all. Consequently, due to the lack of standardized conditions for a fair comparison, quantitative metric results were not included in Table 5.

As mentioned in Section 5.2, many studies utilize skeleton joints as features for fall detection and HAR. To estimate these joints, various DL models are employed, with OpenPose [169] and AlphaPose [170] being the most prevalent (appearing in 25 and 10 studies, respectively). OpenPose utilizes a non-parametric representation (referred to as Part Affinity Fields) to detect skeleton joints from all humans in the image simultaneously, while AlphaPose performs human detection first and then predicts the skeleton joints for each individual. Subsequently, multiple models are used for fall detection and HAR with these skeleton joints:

- **Recurrent Networks:** Long-Short Term Memory (LSTM) [93, 95, 180] and Gated Recurrent Unit (GRU) [95, 114] are commonly used, with others grouped as RNN [40, 66].
- **Graph-Based Network:** The Graph-Convolutional Network (GCN) was the only one found, which treats skeletons as graphs rather than sequences [44, 107, 112]. Additionally, graph-based networks have the potential to perform collective activity recognition by leveraging interactive relations [213].
- **Convolutional Networks:** Various models like VGG architectures [127, 146, 165], MobileNet [42, 81, 141], ResNet family [81, 141], among others [118, 150, 156], are employed.

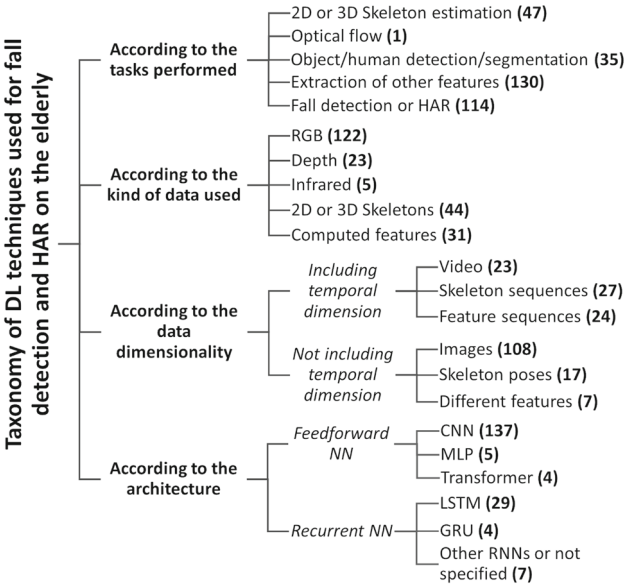


Fig. 6 Taxonomy of the DL techniques used in the found studies. The number of studies where each category was used is displayed in bold. Note that multiple models were used in many studies, and hence the same study can be counted in more than one category

Only one DL model, LiteFlowNet, was used for optical flow estimation across the studies [149]. However, 11 additional studies utilized optical flow at some stage of the recognition pipeline through non-DL-based methods [50, 55, 56, 58, 70, 83, 96, 109, 128, 146, 155].

Object detection was a prevalent task in the reviewed studies (found in 35 studies), with models from two families: R-CNN [190] and YOLO [191]. R-CNN involves a multi-step process including region proposal, feature extraction, object classification, bounding box regression, and non-maximum suppression. Conversely, YOLO focuses on real-time object detection with a single pass through the image. Both models received several ameliorations in later versions. These models were utilized for various purposes across the studies:

- Obtaining a sequence of bounding boxes from scene objects, which can serve as features in next steps [46, 52, 99].
- Triggering computation of fall detection or HAR upon detection of human presence, saving computation time [100, 105].
- Reducing data complexity by putting the focus on the target person [27, 102, 161].
- Getting features from the humans in the scene, like height-to-width ratio, used for fall detection or HAR in further steps [31, 78, 136].
- Direct detection of falls or recognition of activities [25, 76, 139].

Additionally, object segmentation plays a crucial role in several studies. The most commonly used model is Mask R-CNN [214], which extends the capabilities of the R-CNN family to object segmentation. Another notable model is PointRend [192], a neural network module that enhances the granularity of segmentation models by treating image segmentation as a rendering problem. Conversely, a novel model proposed in [67] specifically addresses object segmentation as part of the processing pipeline for fall detection and post-fall classification, named MSSkip. MSSkip builds upon common ideas from other segmentation models but incorporates multi-scale skip connections and depth-wise separable convolutions in the decoder to minimize computation. Object segmentation serves various purposes in the reviewed studies: in [103], averaged output masks are utilized as spatio-temporal features for further recognition steps; [88] performs direct classification into fall or not fall based on the segmentation of fallen individuals; segmentation masks are fed to a convolutional LSTM in [67] and to a CNN followed by an LSTM in [153] to extract spatial and temporal features for fall detection; in [108], segmentation masks are input to different machine learning models to identify falls. Conversely, in [41], segmentation is used solely to anonymize images before feeding them to an autoencoder for fall detection.

Moreover, alongside the aforementioned DL-computed features, other features are predominantly computed using convolutional models such as VGG-16, VGG-19 [27, 143, 153], ResNet [55, 145], or InceptionV3 [157]. Less frequently, non-DL-based features like Histograms of Oriented Gradients (HOG) [55, 134], Local Binary Patterns (LBP) [55, 86], and Bag of Words (BoW) [50, 138] are also utilized. Following feature extraction, multiple DL models are employed for classification. However, at this stage, it is common to use non-deep machine learning models such as Support Vector Machine [55, 58, 106, 136], Random Forest [39, 55], Decision Tree [39, 106], and KNN [106].

Furthermore, fall detection is frequently approached as a normal/abnormal classification task in the reviewed studies, with normal activities modeled and falls treated as abnormal data. This involves performing feature extraction, either using pre-trained models to extract spatio-temporal features from video/images or utilizing estimated skeleton joints, followed by training a model to identify normal activities. Various approaches are employed for this task, such as utilizing an MPED-RNN network on skeletal data [94], employing DeepFall on multiple data modalities (RGB, depth, and IR) [20], using autoencoders after obtaining spatio-temporal features from other networks [41, 111, 144], and employing Generative Adversarial Networks (GANs) by utilizing the discriminator as the normal/abnormal classifier [86, 103].

Finally, the choice of architecture in the analyzed studies often depends on the data dimensionality, with recurrent neural networks (RNNs) primarily used when considering

Table 6 Comprehensive list of publicly available datasets used in the reviewed studies, along with their basic specifications

Ref.	Dataset	Eld.	Falls	Type	Data types	Samples	Cl.	Studies
[216]	URFD	No	Yes	Video	RGB-D, Skel., IMU	140	2	40
[168]	UP-FALL	No	Yes	Video	RGB, IR, IMU	561	11	17
[217]	Le2i	No	Yes	Video	RGB	191	2	16
[218]	MultiCam	No	Yes	Video	RGB	192	2	16
[219]	NTU RGB+D	No	Yes	Video	RGB-D, Skel.	56880	60	14
[220]	FDD-Adhikari	No	No	Image	RGB-D	21499 (frames)	5	6
[221]	MSRDailyActivity3D	No	No	Video	RGB-D, Skel.	320	16	4
[222]	UTD-MHAD	No	No	Video	RGB-D, Skel., IMU	861	27	4
[223]	CAD-60	No	No	Video	RGB-D	720	12	3
[224]	ETRIActivity3D	Yes	Yes	Video	RGB-D, Skel.	112620	55	3
[225]	HMDB51	No	Yes	Video	RGB	6766	51	3
[226]	KTH	No	No	Video	RGB	2391	6	3
[118]	PRECIS HAR	No	Yes	Video	RGB-D	800	16	3
[154]	ToyotaSmartHome	Yes	No	Video	RGB-D, Skel.	16115	31	3
[227]	CAD-120	No	No	Video	RGB-D	1200	10	[36, 156]
[228]	DMLSmartActions	No	Yes	Video	RGB-D, Skel.	932	12	[47, 160]
[136]	FPDS	No	Yes	Image	RGB	2064	2	[101, 136]
[229]	HQFSD	Yes	Yes	Video	RGB	55	2	[46, 151]
[230]	NTU RGB+D 120	No	Yes	Video	RGB-D, Skel.	114480	120	[107, 135]
[231]	N-UCLA	No	No	Video	RGB-D, Skel.	100	10	[161]
[232]	UCF101	No	No	Video	RGB	13320	101	[58, 143]
[233]	UTKinect-Action3D	No	No	Video	RGB-D, Skel.	200	10	[102, 150]
[234]	UWA3DII	No	Yes	Video	RGB-D, Skel.	120	30	[107, 161]
[20]	MUVIM	Yes	Yes	Video	RGB-D, IR, IMU	244	2	[20]
[161]	ALMOND	No	Yes	Video	RGB	7565	22	[161]
[235]	BIT-interaction	No	No	Video	RGB	400	8	[162]
[236]	C-MHAD	No	Yes	Video	RGB, IMU	120	7	[44]
[97]	FallAction	No	Yes	Video	RGB	2000	20	[97]
[237]	FDD-Chen	No	Yes	Video	RGB	30	2	[126]
[238]	FDD-TST	No	Yes	Video	RGB-D, Skel., IMU	132	8	[157]
[239]	FPDS-Elderly	Yes	Yes	Image	RGB	413	2	[101]
[240]	IXMAS	No	No	Video	RGB, MHV	330	11	[161]
[241]	Kinetics 400	No	No	Video	RGB	306245	400	[129]
[242]	Kinetics 600	No	Yes	Video	RGB	495547	600	[97]
[243]	Kinetics 700-2020	No	Yes	Video	RGB	647907	700	[125]
[135]	KIST SynADL	No	Yes	Video	RGB-D, Skel.	462200	55	[135]
[244]	MMU	No	Yes	Video	RGB	51	2	[93]
[245]	NAD	No	No	Video	RGB	84	7	[135]
[246]	OOPS	No	Yes	Video	RGB	20338	2	[97]
[247]	PKU-MMD	No	Yes	Video	RGB-D, Skel.	21545	51	[121]
[248]	Stanford40	No	No	Image	RGB	9532	40	[125]
[249]	V-COCO	No	No	Image	RGB	10346	25	[52]
[101]	VWFP	No	Yes	Image	RGB	6071	2	[101]
[250]	YTBF	No	Yes	Video	RGB-D, Skel.	606	2	[151]

The columns "Eld." and "Cl." denote the presence of elderly people in the datasets and the number of classes, respectively. The "Studies" column indicates in which studies they appear, or the number of reviewed studies if it is greater than two

the temporal dimension and feedforward neural networks (FFNNs) when not. RNNs are well-suited for problems involving sequential data due to their ability to remember input data using internal memory. As such, they are often employed for fall detection and activity recognition from skeleton sequences [49, 180] and feature sequences computed frame-wise by CNNs [24, 143, 148]. While CNNs are commonly used for extracting visual features from images, transformers have also been utilized in the FFNNs category, particularly for tasks involving low-resolution images [119], 3D skeleton data [81], and video by adapting Vision Transformer (ViT) [215] to video formats [53, 79]. Additionally, multilayer perceptrons (MLPs) are consistently employed for skeleton data [34, 38, 92, 140] or visual features [108].

5.4 RQ1.3: Datasets

Table 6 provides a comprehensive list of datasets used in the reviewed studies for activity recognition and fall detection. Emphasizing the importance of reproducibility and comparability, only publicly available datasets are included, aiming to facilitate future research in the field. Each dataset is categorized based on several common characteristics:

- **Elderly:** Despite fall detection and activity recognition often targeting elderly individuals, only a small fraction of datasets (12%) include samples from this demographic. This scarcity highlights the challenge of collecting real-life data from the elderly population, especially genuine fall incidents.
- **Falls:** The majority of datasets (58%) include falls as a class, with 23% specifically focusing on binary classification between fall and not fall activities, underscoring the significance of this task in eldercare.
- **Type:** Video data is predominant (85% of datasets), aligning with the temporal nature of activities like falls, where temporal context is crucial for accurate recognition. Furthermore, video allows for the rapid acquisition of a large quantity of images in the form of frames, which can then be utilized by data-driven solutions, such as DL-based methods.
- **Data types:** While RGB data is ubiquitous, depth frames, skeleton joints, and inertial data are found in 38%, 29%, and 13% of datasets, respectively. Other data types such as infrared data and motion history volumes (MHV) are less common. The presence of RGB data in all datasets allows for the discovery of the exact conditions of the recordings (environment, perspective, users, etc.) and serves as a visual check of the data, a feature not offered by other types of data.
- **Samples:** Dataset sizes vary significantly, ranging from less than 50 samples (e.g., FDD-Chen) to over 500,000 samples (e.g., Kinetics 700-2020), reflecting the diversity in data availability.
- **Classes:** The number of classes also varies widely, from binary classification to datasets with hundreds of classes, though the latter are typically not focused on AAL.
- **Studies:** Half of the datasets are utilized in only one study, while only five are used in more than ten studies, indicating varying degrees of dataset popularity and usage.

The University of Rzeszow Fall Detection (URFD) dataset [216] stands out as the most extensively used, featuring in 40 studies [41, 89, 153]. Focused on fall detection, URFD offers 70 sequences capturing falls and activities of daily living (ADL) from two perspectives, along with various data modalities including RGB, depth, skeleton joints, and inertial data. The UP-FALL dataset [168], appearing in 17 studies [24, 39, 103], provides data from 17 subjects performing 11 activities, offering RGB video, infrared images, and inertial data for both fall detection and human activity recognition (HAR). In contrast, the Le2i dataset [217], used in 16 studies [47, 93, 137], focuses solely on fall detection, featuring 143 videos with falls and 48 with normal activities, with varying actors, scenery characteristics, and illumination conditions. Similarly, the MultiCam dataset [218], utilized in 16 studies [27, 30, 72], provides RGB video from 24 sequences captured from eight perspectives, facilitating the study of falls and confounding events. The NTU RGB+D dataset [219], used in 14 studies [112, 118, 131], offers a vast collection of samples from 40 subjects performing 60 activities, recorded using Kinect cameras, thus providing RGB video, depth images, and skeleton joints. An extended version of this dataset also exists: the NTU RGB+D 120 dataset [230], which expands upon it by adding 60 additional classes. However, it is only utilized in two of the reviewed studies [107, 135]. The remaining datasets were utilized fewer than 10 times, with approximately half of them being employed in only one study.

While most datasets are collected from real environments, two exceptions are noted: [101] and [135], offering synthetic images and videos, respectively. Despite the advantages of synthetic data, such as ease of acquisition and controlled conditions, models trained solely on synthetic data may lack adaptability to real-world scenarios.

Notably, some studies opted for custom datasets instead of utilizing existing ones. Figure 7 illustrates the proportion of studies using custom, external, or both types of datasets. Only 19 studies provided evaluations on both custom and external datasets, with a greater frequency of evaluations conducted solely on external datasets (86 studies) compared to those exclusively using custom datasets (46 studies).

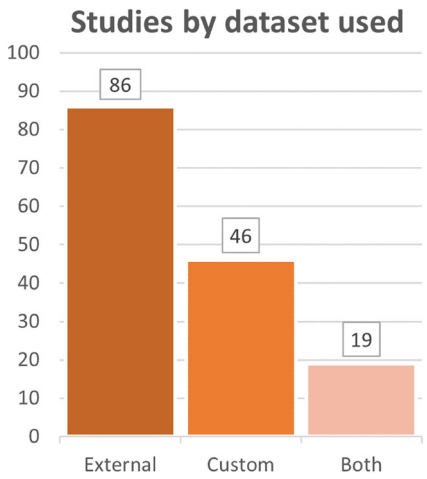


Fig. 7 Distribution of studies by dataset used

5.5 RQ2: Framework integration

In 18 of the reviewed articles, frameworks were proposed to integrate the tasks of HAR or fall detection into real environments, addressing various aspects such as security, utilization of cloud services, client-server configuration, network communications, IoT devices, etc. Below, we provide brief descriptions of the proposed frameworks.

In [42], a custom robot is suggested to integrate the HAR task into the environment, alongside other functionalities like language processing to enable chatbot interactions. In [161], a camera system is employed to capture visual data, which is then sent to a central server for computation. Subsequently, notifications, reports, and alerts are dispatched to a designated "guardian".

In [74], a Docker-based system is proposed to manage the flow between various programs involved in fall detection, distributing resources, and regulating communications. Docker is also utilized in [78], where the NAO robot is suggested for data acquisition and user interaction to prevent falls. In [30, 32], an intermediary step between recording and DL computation is introduced to preprocess video data and reduce bandwidth consumption.

In [18, 33, 49, 52, 58, 93, 105], the proposed frameworks integrate the collection of visual data through camera monitoring systems, centralized server-based recognition of fall detection or various activities, and trigger various responses based on the severity of the situation, such as contacting health services. For instance, [33] utilizes the third-party service 'Twilio' to send phone messages in case of a fall, while in [105], the system transfers recordings to a computer for human inspection upon fall detection.

In [123, 127], activity recognition results, along with recorded video data, are transmitted to a mobile application used for monitoring system users. Similar capabilities are offered in [63], with the addition of face blurring anonymization. [77] conducts all experiments in a connected environment, exploring the use of network traffic from multiple smart appliances combined with visual data to recognize various activities. Additionally, to assess the transferability of their approach across environments, they experimented with a smart residential apartment.

In [85], federated learning is employed to ensure privacy preservation of users. The system incorporates three sensor modalities (depth, mmWave radar, and audio) and was tested in the homes of 16 elderly subjects.

5.6 RQ2.1: hardware

A list of the hardware used in the reviewed studies (when mentioned) is presented in Table 7. Specialized cameras such as thermal, depth, and wearable cameras, as well as social assistive robots, were included. Information regarding datasets not created in the reviewed studies was excluded. Hardware related to computation or common RGB cameras was omitted due to the wide range of possibilities available in these areas.

For depth video retrieval, the most commonly used camera is the Microsoft Kinect (7 studies), followed by the Orbbec Astra Pro (3 studies), and Intel RealSense (1 study). These cameras share similar specifications, offering RGB-D recording using an IR camera for the depth channel, which provides accurate depth estimation at short distances. Additionally, they enable reliable 3D skeleton joint estimation.

There is less consensus in the use of thermal cameras, with multiple camera models employed. Consequently, there is considerable variation in the retrieved data, including differences in resolution, sensitivity to temperature, maximum and minimum effective distances, etc.

Only five studies deployed HAR or fall detection in an AAL system using a social assistive robot. Among these, two studies utilized the Pepper robot, one employed the NAO robot, and the remaining studies used custom-made robots.

5.7 RQ2.2: privacy protection

Figure 8 illustrates the various privacy protection methods identified in the reviewed studies. Among the 151 studies reviewed, 75 did not address privacy concerns, opting for the use of unmodified RGB video or images of elderly users. Among the remaining studies, the majority employed skeleton data computed from RGB images, while four offered specific methods to anonymize RGB data, and others chose to utilize thermal or depth data instead.

The most effective privacy-preserving methods avoid the deployment of RGB cameras in AAL settings. This is typically achieved through the use of visual data types that do not allow for subject identification, such as thermal and

Table 7 Special cameras and social robots found in the reviewed studies

Type	Model	Brand	Studies
RGB-D Camera	Astra Pro	Orbbec	[20, 117, 118]
	Kinect v1	Microsoft	[98, 148, 152, 154]
	Kinect v2	Microsoft	[17, 38, 43]
	RealSense D415	Intel	[83]
Thermal Camera	AMG8831	Panasonic	[119]
	FLIR ONE	iTherml	[20]
	HTPA32x32d	Heimann Sensor	[122]
	MLX90640	Melexis	[45]
	MLX90641	Melexis	[159]
	PI450	Optris	[144]
Wearable RGB Camera	OnReal G1	Fondi	[55]
LiDAR	Horizon LiDAR	Livox	[113]
Social Assistive Robot	Dori	Custom made	[42]
	LOLA	Custom made	[136]
	NAO	Aldebaran URG	[78]
	Pepper	Aldebaran URG	[116, 158]

depth imaging. Among the collected studies, five exclusively employed thermal data [20, 45, 119, 122, 159]. In all cases, DL-based methods utilized CNNs to extract visual features and perform classification. Additionally, 21 studies utilized solely depth data, with 17 of them using it to estimate 3D skeleton poses, as demonstrated in [38, 43, 81, 152]. Notably, Microsoft Kinect was utilized in all 17 studies to estimate skeletons from depth maps through randomized decision forests [251], leaving RGB data unused for this estimation. Four studies exploited depth data without skeleton estimation, instead relying on the extraction of human silhouettes [148] and visual features using CNNs [20, 113, 117].

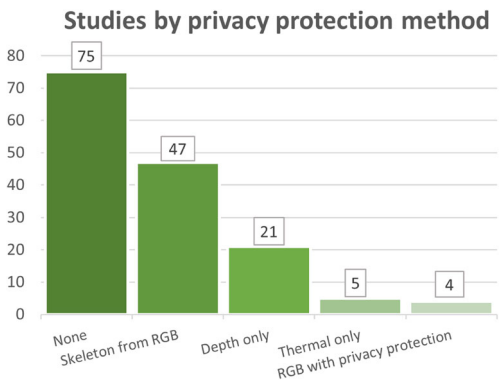


Fig. 8 Distribution of studies by method used to preserve privacy. The total does not add up to 151 studies because in some studies different options were given

A total of 51 studies utilized RGB data at some stage, applying anonymization techniques. In contrast to the aforementioned studies, the input data used by these studies can be used to identify subjects, as conventional video recording is involved at the beginning of the processing pipeline. Among these, 47 studies relied on 2D skeleton estimation methods like OpenPose [169] and AlphaPose [170] to protect privacy, removing visual data that can be used to identify users, as illustrated in [24, 44, 104, 107, 137]. There were four studies in which privacy was protected through other methods. In [144], an IR camera is used to detect the face region of frames and remove it from the RGB frames. In [86], the RGB frames are modified in such a way that individuals cannot be identified, while fall detection can still be applied effectively. In [55], a wearable camera providing a first-person perspective is used to avoid recording the user of the system. Human silhouettes are computed in [41] and used for future recognition steps.

6 Discussion

This section utilizes the discovered results and the responses provided to the review questions to underscore common strengths and weaknesses of the reviewed studies. It also compiles a comprehensive list of recommendations for future reference based on the findings of this systematic literature review. In Figure 9, the search process and key findings from the reviewed studies are summarized.

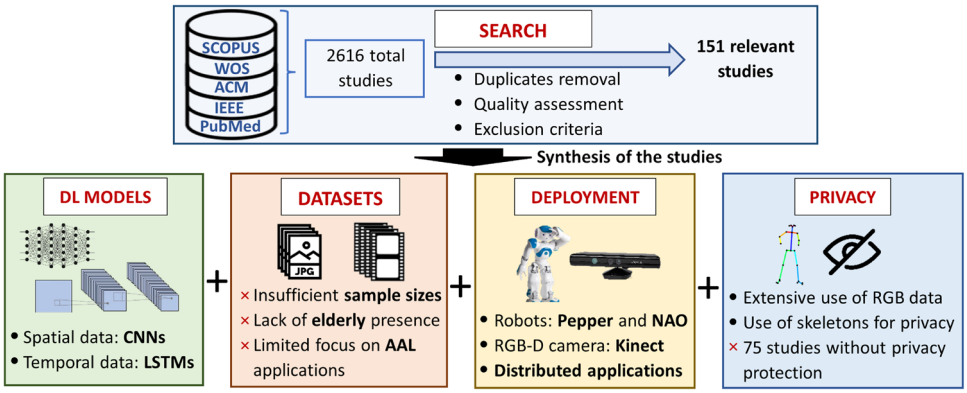


Fig. 9 Summary of the search process and found results

6.1 Strengths and weaknesses of the reviewed studies

Upon reviewing the 151 relevant studies and addressing the research questions, the main strengths and weaknesses observed are discussed in this subsection, which we believe can provide valuable insights for future studies in the field.

A notable benefit of utilizing skeleton joints is their ability to significantly reduce data size compared to raw image or video data, while also offering user anonymization, maintaining data interpretability, and achieving satisfactory results in fall detection and HAR. Furthermore, there is a growing number of methods to derive human skeletons from RGB or depth data, with 13 different skeleton estimation DL models identified in the reviewed studies (as shown in Table 5).

The primary strength of studies employing only depth or infrared data lies in the privacy protection they afford, as RGB footage is not recorded at any point in the system pipeline. However, these studies also face two major weaknesses: a reduced amount of data for detection or recognition tasks, particularly pronounced in the case of IR recordings where resolution tends to be much lower, and less interpretable data, which may pose challenges when manual intervention is required to address errors.

Among the reviewed studies, 27 perform both fall detection and HAR tasks (refer to Fig. 4). This integration is particularly significant, as it is often desirable to detect accidental falls while conducting HAR on elderly individuals. It is important to note that while fall detection can be integrated as another class during HAR, it should be computed separately due to its critical nature. Therefore, most studies including fall detection implement it differently than the recognition of other classes.

Numerous studies have overlooked the temporal dimension when conducting HAR, thereby constraining the task significantly. This omission poses a significant weakness, particularly when incorporating activities that are challenging to distinguish without temporal data or are more effectively recognized with it, such as sitting/getting up or putting on/off clothes. Nonetheless, confining the analysis to spatial data typically offers the advantage of being faster and more straightforward.

Regarding the choice of model architecture, convolutional models were found to be predominant. Their primary strengths lie in their effectiveness in processing spatial data and their extensive history, which has led to numerous improvements and architectural refinements across various fields and tasks. Given their suitability for image-based tasks, convolutional models are widely preferred and even have 3D versions tailored for video processing. In contrast, recurrent models excel in handling sequential data, thus complementing the capabilities of CNNs by facilitating the tracking of computed features across different frames. Multi-layer perceptrons, however, do not yield favorable results with spatial or sequential data; they are typically employed for classification based on computed features, akin to fully connected layers in a convolutional neural network. Transformer-based architectures, being relatively newer, are not as ubiquitous. Despite their promise in handling sequential and vision data, their large parameter count presents challenges in training and deploying them on low-specification systems. Nonetheless, they have showcased significant potential across various domains.

Given that fall detection and HAR for the elderly aim to assist this population in AAL settings, studies offering frameworks for deploying systems in real environments are of particular interest. Eighteen studies, described in Section 5.5, fall into this category.

Utilizing only an external dataset may impact the applicability of the technique to specific situations or environments, but it allows for the comparison of different methods on the same data. Conversely, relying solely on a custom dataset yields the opposite effects. The primary drawback associated with using only custom-made datasets is the external validity of the findings, as it becomes challenging to compare results with other studies, especially if the custom data is not disclosed. Including an evaluation on external datasets not only distinguishes studies from previous ones but also enables

future studies to build on the obtained performance. While the majority of the reviewed studies evaluate on existing datasets for fall detection and HAR, 46 exclusively perform evaluations on new custom datasets (as depicted in Fig. 7), limiting the reliability of the results without comparisons with existing techniques or models. Conversely, 19 studies utilize both custom and external datasets, leveraging the strengths of each approach: specialization on custom data and comparison with other methodologies.

From a data perspective, three common weaknesses are evident in the datasets utilized: the absence of elderly individuals, a limited number of samples, and the inclusion of numerous classes unrelated to activities of daily living (ADL), which may render them less suitable for fall detection and HAR among elderly populations. Primarily, the majority of datasets (88%) lack elderly participants, presenting challenges during deployment as they represent the target users of the system but are not represented in the training data. In this regard, datasets such as ETRIActivity3D, ToyotaSmartHome, MUVIM, and FPDS-Elderly would be more suitable. Additionally, a limited number of samples may prove insufficient for DL models to generalize effectively. Three of the four most extensive datasets contain fewer than 200 samples, while the remaining dataset contains fewer than 600, with approximately half of the utilized datasets containing less than 1,000 samples. Instead, datasets such as ETRIActivity3D, NTU RGB+D (or NTU RGB+D 120), or ToyotaSmartHome, all offering more than 10,000 samples, would yield better generalization results. Lastly, datasets should be tailored to focus on ADL rather than general HAR to avoid unnecessary classes for monitoring elderly individuals in their daily lives. For instance, Kinetics (400, 600, or 700) or UCF101 would not be suitable for the considered tasks as they comprise videos collected from the internet, potentially containing irrelevant activities and cuts.

6.2 Recommendations for future works

Based on the results of this SLR, a series of particularly important considerations, in our understanding, should be taken into account when conducting new studies on the topic.

First and foremost, it is crucial to assess user privacy. As observed, the approach to privacy protection will likely influence the type of data used, ranging from conventional RGB data to modified RGB, depth, IR, or skeleton data, which prevent user identification in the footage. Therefore, we recommend considering privacy protection as a fundamental aspect from the outset of the study.

Selecting an appropriate DL model for fall detection and HAR requires consideration of the deployment conditions. For embedded systems or edge-deployments, such as in social robots or mobile applications, compact models are preferred, such as MobileNet or EfficientNet—well-known CNNs specifically tailored for such devices. These models can be augmented with recurrent models like LSTM to accommodate temporal data. Conversely, if model size is not a constraint, 3D CNNs like I3D, TPN, TANet, SlowFast, and C3D are suitable for video data, while GCN can be applied to skeleton data. Alternatively, Transformer-based architectures like TimeSformer or VST are also an option for processing video input data.

For model evaluation, utilizing a publicly available dataset is essential to enable comparison with existing models or techniques. Prominent datasets for fall detection include URFD, UP-FALL, MultiCam, and Le2i, while for HAR, UP-FALL and NTU RGB+D are commonly used. However, we encourage the adoption of ETRIActivity3D or ToyotaSmartHome, which offer a more extensive collection of video samples and include elderly participants. Both datasets support HAR, with ETRIActivity3D additionally containing falls and providing multiple perspectives from elderly users, diverse classes (at least 30), and various data modalities, including RGB, depth, and skeleton joints.

In cases where a custom dataset is provided, authors are encouraged to make it publicly available. This facilitates its use in future studies, either directly or by merging it with other datasets to form a larger dataset, enhances the reproducibility of experiments, and enables comparison with newer models or techniques. RGB-D cameras, such as Microsoft Kinect, Orbbec Astra Pro, and Intel RealSense, are recommended for collecting custom datasets as they facilitate experimentation with various types of data, with depth data offering privacy preservation capabilities.

When deploying the system in a real environment, the most common approach, as indicated by the reviewed studies, involves establishing a camera setup within the environment. This setup records data and transmits it to a central server for processing. It is also the most cost-effective option, depending on factors such as camera type, resolution, and processing requirements. Alternatively, for those preferring to use an assistive robot, both NAO and Pepper robots are viable solutions. These commercial robots come equipped with cameras, speakers, microphones, and other necessary components, offering customizable options to adapt to different projects and environments.

7 Conclusions

In this systematic literature review, we have investigated fall detection and human activity recognition for the elderly, with a particular focus on deep learning techniques applied to computer vision data. Our study aimed to address two primary research questions related to the implementation of DL methods for these tasks and their deployment in real-world environments, considering hardware and privacy concerns.

Throughout the review process, we analyzed 151 relevant studies, providing a structured overview of the main findings to facilitate accessibility for practitioners and researchers. The findings offer valuable insights into the effective implementation of DL techniques for fall detection and HAR in elderly care, which are becoming increasingly important in the context of Ambient Assisted Living (AAL) systems.

Privacy emerged as a common concern, with 50% of the reviewed studies lacking any measures to address it. The most prevalent privacy protection method identified was the use of skeleton joints estimation, employed in 45% of the studies.

Convolutional DL models were found to be predominant, owing to their effectiveness in processing spatial data and extensive history of refinement. However, we observed a lack of consideration for the temporal dimension in many studies, which limits the recognition of some activities.

Regarding datasets, we identified three common weaknesses: the absence of elderly individuals, a limited number of samples, and the inclusion of numerous irrelevant classes for the AAL systems. We recommend datasets such as ETRI-Activity3D and ToyotaSmartHome, which offer extensive samples and include elderly participants.

Moving forward, we emphasize the importance of privacy assessment from the outset of studies and recommend selecting appropriate DL models based on deployment conditions. Utilizing publicly available datasets for model evaluation is crucial, and authors are encouraged to make custom datasets publicly available to enhance reproducibility and facilitate future research.

In terms of deployment, camera setups within the environment were the most common approach identified, offering cost-effectiveness and flexibility. Alternatively, assistive robots like NAO and Pepper provide customizable options for deployment in various projects and environments.

Overall, this SLR provides a comprehensive overview of recent advancements in DL-based fall detection and HAR for the elderly, offering valuable insights for researchers, practitioners, and policymakers involved in developing and implementing AAL technologies.

Author Contributions F.Xavier Gaya-Morey: Conceptualization, Methodology, Validation, Investigation, Writing - Original Draft, Writing - Review & Editing Preparation, Visualization.
Cristina Manresa-Yee: Conceptualization, Methodology, Writing - Review & Editing Preparation, Supervision, Project administration, Funding acquisition.
Jose M. Buades-Rubio: Conceptualization, Methodology, Writing - Review & Editing Preparation, Supervision, Project administration, Funding acquisition.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. Grant PID2019-104829RA-I00 funded by MCIN/AEI/10.13-039/ 501100011033, project EXPLainable Artificial INtelligence systems for health and well-beING (EXPLAINING). Grant PID2022-136779OB-C32 funded by MCIN/AEI/ 10.13039/501100011033 and by ERDF A way of making Europe, project Playful Experiences with Interactive Social Agents and Robots (PLEISAR): Social Learning and Intergenerational Communication. F. X. Gaya-Morey was supported by an FPU scholarship from the Ministry of European Funds, University and Culture of the Government of the Balearic Islands.

Data Availability All relevant data for the study can be found, structured and organized in the form of tables, throughout this document. No additional data were generated.

Declarations

Competing of interest The authors have no competing interests to declare that are relevant to the content of this article.

Ethical and informed consent for data used This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Bloom DE, Luca DL (2016) Chapter 1 - the global demography of aging: Facts, explanations, future. vol 1, pp 3–56. North-Holland
2. WHO (2021) World Health Organization fact sheets: Falls. https://www.who.int/en/news-room/fact-sheets/detail/falls
3. Heinrich S, Rapp K, Rissmann U, Becker C, König H-H (2010) Cost of falls in old age: a systematic review. Osteoporos Int 21:891–902
4. Climent-Pérez P, Spinsante S, Mihailidis A, Florez-Revuelta F (2020) A review on video-based active and assisted living technologies for automated lifelogging. Expert Systems with Applications. 139
5. Khodabandehloo E, Riboni D, Alimohammadi A (2021) Healthxai: Collaborative and explainable ai for supporting early diagnosis of cognitive decline. Futur Gener Comput Syst 116:168–189
6. Nizam Y, Jamil MMA (2020) Classification of daily life activities for human fall detection: A systematic review of the techniques and approaches. Stud Syst Decis Control 273:137–179
7. Walsh J, O' Mahony N, Campbell S, Carvalho A, Krpalkova L, Velasco-Hernandez G, Harapanahalli S, Riordan D (2019) Deep learning vs. traditional computer vision
8. Kitchenham B, Charters, S (2007) Guidelines for performing systematic literature reviews in software engineering. 2
9. Guerra BMV, Torti E, Marenzi E, Schmid M, Ramat S, Leporati F, Danese G (2023) Ambient assisted living for frail people through human activity recognition: state-of-the-art, challenges and future

directions. Frontiers in Neuroscience. 17. https://doi.org/10.3389/fnins.2023.1256682
10. Kumar R, Kumar S (2023) A survey on intelligent human action recognition techniques. Multimed Tools Appl. https://doi.org/10.1007/s11042-023-17529-6
11. Tay NC, Connie T, Ong TS, Teoh ABJ, Teh PS (2023) A review of abnormal behavior detection in activities of daily living. IEEE Access. 11:5069–5088. https://doi.org/10.1109/ACCESS.2023.3234974
12. Momin MS, Sufian A, Barman D, Dutta P, Dong M, Leo M (2022) In-home older adults' activity pattern monitoring using depth sensors: A review. Sensors. 22
13. Olugbade T, Bieńkiewicz M, Barbareschi G, D'amato V, Oneto L, Camurri A, Holloway C, Björkman M, Keller P, Clayton M, Williams ACDC, Gold N, Becchio C, Bardy B, Bianchi-Berthouze N (2022) Human movement datasets: An interdisciplinary scoping review. ACM Comput, Surv, p 55
14. Alam E, Sufian A, Dutta P, Leo M (2022) Vision-based human fall detection systems using deep learning: A review. Comput Biol Med 146:105626
15. Rastogi S, Singh J (2022) Human fall detection and activity monitoring: a comparative analysis of vision-based methods for classification and detection techniques. Soft Comput 26:3679–3701
16. Gutiérrez J, Rodríguez V, Martin S (2021) Comprehensive review of vision-based fall detection systems. Sensors. 21(3)
17. Sudasinghe SATN, Sooriyabandara IKS, Banadara AHMDPM, Rajendran H, Jayasekara AGBP (2023) Vision attentive robot for elderly room, pp 19–24. https://doi.org/10.1109/MERCon60487.2023.10355403
18. Zhang Y, Zheng X, Liang W, Zhang S, Yuan X (2022) Visual surveillance for human fall detection in healthcare iot. IEEE Multimed 29:36–46
19. Wang X, Zheng X, Liu J, Yuan B, Zhao L, Sun J (2023) Abnormal behavior detection for patients in nursing rehabilitation center. J Appl Sci Eng (Taiwan). 26:925–933. https://doi.org/10.6180/jase.202307_26(7).0003
20. Denkovski S, Khan SS, Malamis B, Moon SY, Ye B, Mihailidis A (2022) Multi visual modality fall detection dataset. IEEE Access 10:106422–106435
21. Marshal S, Raj SA, Jebaseeli TJ, Niranjan S (2023) An image-based fall detection system for the elderly using yolov5, pp 493–498.https://doi.org/10.1109/ICACRS58579.2023.10404248
22. Li S, Man C, Shen A, Guan Z, Mao W, Luo S, Zhang R, Yu H (2022) A fall detection network by 2d/3d spatio-temporal joint models with tensor compression on edge. ACM Trans. Embed. Comput, Syst, p 21
23. Eltahir MM, Yousif A, Alrowais F, Nour MK, Marzouk R, Dafaalla H, Elnour AAH, Aziz ASA Hamza MA (2023) Deep transfer learning-enabled activity identification and fall detection for disabled people. Computers, Materials and Continua 75:3239–3255. https://doi.org/10.32604/cmc.2023.034037
24. Inturi AR, Manikandan VM, Garrapally V (2022) A novel vision-based fall detection scheme using keypoints of human skeleton with long short-term memory network. Arabian Journal for Science and Engineering
25. Ke Y, Yao Y, Xie Z, Xie H, Lin H, Dong C (2023) Empowering intelligent home safety: Indoor family fall detection with yolov5, pp 942–949. https://doi.org/10.1109/DASC/PiCom/CBDCom/Cy59711.2023.10361490
26. Suarez JJP, Orillaza N, Naval P (2022) Afar: A real-time vision-based activity monitoring and fall detection framework using 1d convolutional neural networks. Association for Computing Machinery, New York, NY, USA, pp 555–559
27. Agrawal M, Agrawal S (2023) Enhanced deep learning for detecting suspicious fall event in video data. Intell Autom Soft Comput 36:2653–2667. https://doi.org/10.32604/iasc.2023.033493
28. Yadav SK, Luthra A, Tiwari K, Pandey HM, Akbar SA (2022) Arfdnet: An efficient activity recognition & fall detection system using latent feature pooling. Knowledge-Based Systems. 239
29. Dakare AA, Wu Y, Hashimoto N, Kumagai T, Miura T (2023) Fall detection inside an autonomous driving bus: - examination of image processing algorithms-, pp 1–4. https://doi.org/10.1109/ICCE56470.2023.10043518
30. Rajavel R, Ravichandran SK, Harimoorthy K, Nagappan P, Gobichettipalayam KR (2022) Iot-based smart healthcare video surveillance system using edge computing. J Ambient Intell Humanized Comput 13:3195–3207
31. M, PV, Shekar M, B SLP, Ngadiran R, Ravindran S (2023) Fall detection system for monitoring elderly people using yolov7-pose detection model, pp 1–6. https://doi.org/10.1109/IC2E357697.2023.10262506
32. Wang B, Wu X, Gong M, Zhao J, Sun Y (2022) Lightweight network based real-time anomaly detection method for caregiving at home. Institute of Electrical and Electronics Engineers Inc., Hangzhou, China, pp 1323–1328
33. Paul SK, Zisa AA, Walid MAA, Zeem Y, Paul RR, Haque MM, Hamid ME (2023) Human fall detection system using long-term recurrent convolutional networks for next-generation healthcare: A study of human motion recognition, pp 1–7. https://doi.org/10.1109/ICCCNT56998.2023.10308247
34. Xie L, Sun Y, Chambers JA, Naqvi SM (2022) Privacy preserving multi-class fall classification based on cascaded learning and noisy labels handling. Institute of Electrical and Electronics Engineers Inc., Linköping, Sweden, pp 1–6
35. Fan S, Li M, Han C (2023) Intelligent video monitoring for real-time detection and recognition of elderly falls on the embedded platform, pp 630–635. https://doi.org/10.1109/ICIPCA59209.2023.10257766
36. Patsch C, Zakour M, Chaudhari R (2022) Automatic recognition of human activities combining model-based ai and machine learning. SCITEPRESS, Setubal, Portugal, pp 15–22
37. Cheng B, Su Y, Cai Y (2023) Research on real-time human fall detection method based on yolov5-lite, pp 218–221. https://doi.org/10.1109/ICEICT57916.2023.10245195
38. Guerra BMV, Schmid M, Beltrami G, Ramat S (2022) Neural networks for automatic posture recognition in ambient-assisted living. Sensors 22
39. Inturi AR, Manikandan VM, Kumar MN, Wang S, Zhang Y (2023) Synergistic integration of skeletal kinematic features for vision-based fall detection. Sensors 23. https://doi.org/10.3390/s23146283
40. Sun H, Chen Y (2022)Real-time elderly monitoring for senior safety by lightweight human action recognition, vol 2022-May, pp 1–6. IEEE Computer Society, Lincoln, NE, USA
41. Wahla SQ, Ghani MU (2023) Visual fall detection from activities of daily living for assistive living. IEEE Access 11:108876–108890. https://doi.org/10.1109/ACCESS.2023.3321192
42. Kim J-W, Choi Y-L, Jeong S-H, Han J (2022) A care robot with ethical sensing system for older adults at home. Sensors (Basel) 22:7515
43. Jain, A., Akerkar, R., Srivastava, A.: Privacy-preserving human activity recognition system for assisted living environments. IEEE Transactions on Artificial Intelligence, pp 1–15 (2023) https://doi.org/10.1109/TAI.2023.3323272
44. Lin F, Wang Z, Zhao H, Qiu S, Shi X, Wu L, Gravina R, Fortino G (2022) Adaptive multi-modal fusion framework for activity monitoring of people with mobility disability. IEEE J Biomed Health Inf 26:4314–4324

45. Rezaei A, Stevens MC, Argha A, Mascheroni A, Puiatti A, Lovell NH (2023) An unobtrusive human activity recognition system using low resolution thermal sensors, machine and deep learning. IEEE Trans Biomed Eng 70:115–124. https://doi.org/10.1109/TBME.2022.3186313
46. He J, Xiang M, Zhao X (2022) An elderly indoor behavior recognition method based on improved slowfast network, vol 2216. IOP Publishing Ltd, San Francisco, CA, USA
47. Yazici A, Zhumabekova D, Nurakhmetova A, Yergaliyev Z, Yatbaz HY, Makisheva Z, Lewis M, Ever E (2023) A smart e-health framework for monitoring the health of the elderly and disabled. INTERNET OF THINGS. 24 https://doi.org/10.1016/j.iot.2023.100971
48. Zhang C, Yang X (2022) Bed-leaving action recognition based on yolov3 and alphapose. Association for Computing Machinery, New York, NY, USA, pp 117–123
49. Zhang Y, Liang W, Yuan X, Zhang S, Yang G, Zeng Z (2023) Deep learning based abnormal behavior detection for elderly healthcare using consumer network cameras. IEEE Transactions on Consumer Electronics 1. https://doi.org/10.1109/TCE.2023.3309852
50. Prasad SK, Ko Y-B (2022) Deep learning based human activity recognition with improved accuracy, vol 2022-October, pp 1492–1495. IEEE Computer Society, Jeju Island, Republic of Korea
51. Gao P (2023) Development of yolo-based model for fall detection in iot smart home applications. Int J AdvComput Sci Appl 14:1118–1125.https://doi.org/10.14569/IJACSA.2023.01410117
52. Achirei S-D, Heghea M-C, Lupu R-G, Manta V-I (2022) Human activity recognition for assisted living based on scene understanding. Applied Sciences (Switzerland) 12
53. Gaya-Morey FX, Manresa-Yee C, Buades-Rubio JM (2023). Explainable activity recognition for the elderly. https://doi.org/10.1145/3612783.3612790
54. Islam MM, Nooruddin S, Karray F (2022) Multimodal human activity recognition for smart healthcare applications, vol 2022-October, pp 196–203. Institute of Electrical and Electronics Engineers Inc., Prague, Czech Republic
55. Wang X, Talavera E, Karastoyanova D, Azzopardi G (2023) Fall detection with a nonintrusive and first-person vision approach. IEEE Sensors J 23:28304–28317. https://doi.org/10.1109/JSEN.2023.3314828
56. Isoi H, Takefusa A, Nakada H, Oguchi M (2022) Performance of domain adaptation schemes in video action recognition using synthetic data. Association for Computing Machinery, New York, NY, USA, pp 70–79
57. Rashid H-A, Mohsenin T (2023) Hac-pocd: Hardware-aware compressed activity monitoring and fall detector edge poc devices, pp 1–5. https://doi.org/10.1109/BioCAS58349.2023.10389023
58. Ji Q (2022) The design of the lightweight smart home system and interaction experience of products for middle-aged and elderly users in smart cities. Comput Intell Neurosci 2022:1279351
59. Luo, B.: Human fall detection for smart home caring using yolo networks. Int J Adv Comput Sci Appl14:59–68 (2023) https://doi.org/10.14569/IJACSA.2023.0140409
60. Ouyang X, Shuai X, Zhou J, Shi IW, Xie Z, Xing G, Huang J (2022) Cosmo: Contrastive fusion learning with small data for multimodal human activity recognition. Association for Computing Machinery, New York, NY, USA, pp 324–337
61. Ravankar A, Rawankar A, Ravankar AA (2023) Real-time monitoring of elderly people through computer vision. Artif Life Robot 28:496–501. https://doi.org/10.1007/s10015-023-00882-y
62. Galvao YM, Portela L, Ferreira J, Barros P, Fagundes OADA, Fernandes BJT (2021) A framework for anomaly identification applied on fall detection. IEEE ACCESS 9:77264–77274
63. Singh IS, Kaza P, Hosler Iv, PG, Chin ZY, Ang KK (2023) Real-time privacy preserving human activity recognition on mobile using 1dcnn-bilstm deep learning. In: Proceedings of the 2023 5th International Conference on Image, Video and Signal Processing. IVSP '23, pp 18–26. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3591156.3591159
64. Galvo YM, Ferreira J, Albuquerque VA, Barros P, Fernandes BJT (2021) A multimodal approach using deep learning for fall detection. Expert Systems with Applications 168
65. Sukreep S, Dajpratham P, Nukoolkit C, Yamsaengsung S, Khajontantichaikun T, Mongkolnam P, Jaiyen S, Chongsuphajaisiddhi V (2023) Recognizing fall risk factors with convolutional neural network, pp 391–396. https://doi.org/10.1109/JCSSE58229.2023.10202147
66. Kang Y-K, Kang H-Y, Kim J-B (2021) A study of fall detection system using context cognition method. Institute of Electrical and Electronics Engineers Inc., Ho Chi Minh City, Vietnam, pp 79–83
67. Mobsite S, Alaoui N, Boulmalf M, Ghogho M (2023) Semantic segmentation-based system for fall detection and post-fall posture classification. Engineering Applications of Artificial Intelligence 117 https://doi.org/10.1016/j.engappai.2022.105616
68. Raj A, Singh D, Prakash C (2021) Active human pose estimation for assisted living. Association for Computing Machinery, New York, NY, USA, pp 110–116
69. Zhou F, Zhu G, Li X, Li H, Shi Q (2023) Towards pervasive sensing: A multimodal approach via csi and rgb image modalities fusion. In: Proceedings of the 3rd ACM MobiCom workshop on integrated sensing and communications systems. ISACom '23, pp 25–30. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3615984.3616505
70. Yang Y, Ren H, Li C, Ding C, Yu H (2021) An edge-device based fast fall detection using spatio-temporal optical flow model. Institute of Electrical and Electronics Engineers Inc., New York, NY, USA, pp 5067–5071
71. Zaghdoud A, Jemai O (2023) A metaplastic neural network technique for human activity recognition for alzheimer's patients, pp 1–6. https://doi.org/10.1109/INISTA59065.2023.10310437
72. Sultana A, Deb K, Dhar PK, Koshiba T (2021) Classification of indoor human fall events using deep learning. Entropy (Basel) 23
73. Siow CZ, Chin WH, Zhang Y, Naoyuki K (2023) A one-shot learning method for human activity recognition using extracted essential poses from push-fixed adaptive resonance theory, pp 569–575. https://doi.org/10.1109/ICMLC58545.2023.10327934
74. Divya V, Sri RL (2021) Docker-based intelligent fall detection using edge-fog cloud infrastructure. IEEE Internet Things J 8:8133–8144
75. Shejy G, Chattani B, Batheja V, Patwa M, Deshmukh S (2023) Activity monitoring and unusual activity detection for elderly homes, pp 1–5. https://doi.org/10.1109/ICACTA58201.2023.10393535
76. Chen Y, Du R, Luo K, Xiao Y (2021) Fall detection system based on real-time pose estimation and svm. Institute of Electrical and Electronics Engineers Inc., Nanchang, China, pp 990–993
77. Liu S, Mangla T, Shaowang T, Zhao J, Paparrizos J, Krishnan S, Feamster N (2023) Amir: Active multimodal interaction recognition from video and network traffic in connected environments. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 7. https://doi.org/10.1145/3580818
78. Killian L, Julien M, Kevin B, Maxime L, Carolina B, Mélanie C, Nathalie B, Sylvain G, Sebastien G (2021) Fall prevention and detection in smart homes using monocular cameras and an interactive social robot. Association for Computing Machinery, New York, NY, USA, pp 7–12
79. Ammar SB, Ho T-C, Karray F, Gueaieb W (2023) Hacer: An integrated remote monitoring platform for the elderly, pp 1–6. https://doi.org/10.1109/iMETA59369.2023.10294645

80. Ge W, Luo X, Tao R, Shi Y (2021) Human fall detection algorithm based on mixed attention mechanism. Association for Computing Machinery, New York, NY, USA, pp 32–37
81. Snoun A, Bouchrika T, Jemai O (2023) Deep-learning-based human activity recognition for alzheimer's patients' daily life activities assistance. Neural Comput Appl 35:1777–1802. https://doi.org/10.1007/s00521-022-07883-1
82. Pita MSU, Alon AS, Melo PMB, Hernandez RM, Magboo AI (2021) Indoor human fall detection using data augmentation-assisted transfer learning in an aging population for smart homecare: A deep convolutional neural network approach. Institute of Electrical and Electronics Engineers Inc., New York, NY, USA, pp 64–69
83. Raza MA, Chen L, Li N, Fisher RB (2023) Eatsense: Human centric, action recognition and localization dataset for understanding eating behaviors and quality of motion assessment. Image and Vision Computing 137. https://doi.org/10.1016/j.imavis.2023.104762
84. Vaiyapuri T, Lydia EL, Sikkandar MY, Diaz VG, Pustokhina IV, Pustokhin DA (2021) Internet of things and deep learning enabled elderly fall detection model for smart homecare. IEEE Access 9:113879–113888
85. Ouyang X, Xie Z, Fu H, Cheng S, Pan L, Ling N, Xing G, Zhou J, Huang J (2023) Harmony: Heterogeneous multi-modal federated learning through disentangled model training. In: Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services. MobiSys '23, pp 530–543. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3581791.3596844
86. Liu J, Tan R, Han G, Sun N, Kwong S (2021) Privacy-preserving in-home fall detection using visual shielding sensing and private information-embedding. IEEE Trans Multimed 23:3684–3699
87. Fayad M, Hachani M-Y, Mostefaoui A, Chouali S, Yahiaoui R (2022) Elderly fall detection: A lightweight kinect based deep learning approach. Association for Computing Machinery, New York, NY, USA, pp 89–95
88. Zherdev D, Zherdeva L, Agapov S, Sapozhnikov A, Nikonorov A, Chaplygin S (2021)Producing synthetic dataset for human fall detection in ar/vr environments. Applied Sciences (Switzerland) 11
89. Meraikhi SA, Al-Rajab M (2022) A multimodal approach of machine and deep learnings to enhance the fall of elderly people. J Inf Technol Manag 14:168–184
90. Feng X, Jiang W (2021) Research on human fall detection based on tiny-yolov3 algorithm. Association for Computing Machinery, New York, NY, USA, pp 1326–1330
91. Chen P-C, Chang C-H, Chan Y-W, Tasi Y-T, Chu WC (2022) An approach to real-time fall detection based on openpose and lstm, pp 1573–1578
92. Xie L, Yang Y, Zeyu F, Naqvi SM (2021) Skeleton-based fall events classification with data fusion. Institute of Electrical and Electronics Engineers Inc., Karlsruhe, Germany
93. Anwary AR, Rahman MA, Muzahid AJM, Ashraf AWU, Patwary M, Hussain A (2022) Deep learning enabled fall detection exploiting gait analysis. Annu Int Conf IEEE Eng Med Biol Soc 2022:4683–4686
94. Fatima M, Yousaf, MH, Yasin A, Velastin SA (2021) Unsupervised fall detection approach using human skeletons, pp 1–6
95. Lau XL, Connie T, Goh MKO, Lau SH (2022) Fall detection and motion analysis using visual approaches. Int J Technol 13:1173–1182
96. Berlin SJ, John M (2021) Vision based human fall detection with siamese convolutional neural networks. Journal of Ambient Intelligence and Humanized Computing
97. Nigam N, Dutta T, Verma D (2022) Fall-perceived action recognition of persons with neurological disorders using semantic supervision. IEEE Transactions on Cognitive and Developmental Systems 1
98. Li X, Chen W (2021) Fall recognition algorithm for the elderly based on home service robot. Institute of Electrical and Electronics Engineers Inc., Zhengzhou, China, pp 329–335
99. Aarthi MS, Juliet S (2022) Intelligent fall detection system based on sensor and image data for elderly monitoring, pp 1259–1265
100. Fernando YPN, Gunasekara KDB, Sirikumara KP, Galappaththi UE, Thilakarathna T, Kasthurirathna D (2021) Computer vision based privacy protected fall detection and behavior monitoring system for the care of the elderly, vol. 2021-September, pp 1–7. Institute of Electrical and Electronics Engineers Inc., New York, NY, USA
101. Carrara F, Pasco L, Gennaro C, Falchi F (2022) Learning to detect fallen people in virtual worlds. Association for Computing Machinery, New York, NY, USA, pp 126–130
102. Tseng H-T, Hsieh C-C, Hsu T-Y (2021) Elder action recognition based on convolutional neural network and long short-term memory, pp 1–2
103. Galvão YM, Portela L, Barros P, Araújo Fagundes RA, Fernandes BJT (2022) Onefall-gan: A one-class gan framework applied to fall detection. Engineering Science and Technology, an International Journal
104. Ramirez H, Velastin SA, Meza I, Fabregas E, Makris D, Farias G (2021) Fall detection and activity recognition using human skeleton features. IEEE Access 9:33532–33542
105. Zheng H, Liu Y, Wu X, Zhang Y (2022) Realization of elderly fall integration monitoring system based on alphapose and yolov4. Institute of Electrical and Electronics Engineers Inc., Hangzhou, China, pp 604–620
106. Wang X, Talavera E, Karastoyanova D, Azzopardi G (2021) Fall detection and recognition from egocentric visual data: A case study. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 12661 LNCS, pp 431–443
107. Zahan S, Hassan GM, Mian A (2022) Sdfa: Structure aware discriminative feature aggregation for efficient human fall detection in video. IEEE Transactions on Industrial Informatics, pp 1–9
108. Hasib R, Khan KN, Yu M, Khan MS (2021) Vision-based human posture classification and fall detection using convolutional neural network, pp 74–79
109. Rajalaxmi RR, Gothai E, Suganth V, Vignesh S, Varun T (2022) Vision based fall detection using optimized convolutional neural network. Institute of Electrical and Electronics Engineers Inc., Coimbatore, India, pp 1–6
110. Wang J, Yang W (2021) Action recognition based on cross spatial temporal graph convolution. Association for Computing Machinery, New York, NY, USA
111. Anitha G, Priya SB (2022) Vision based real time monitoring system for elderly fall event detection using deep learning. Comput Syst Sci Eng 42:87–103
112. Budisteanu E-A, Mocanu IG (2021) Combining supervised and unsupervised learning algorithms for human activity recognition. SENSORS 21
113. Tu L, Ouyang X, Zhou J, He Y, Xing G (2021) Feddl: Federated learning via dynamic layer sharing for human activity recognition. Association for Computing Machinery, New York, NY, USA, pp 15–28
114. Jaouedi N, Perales FJ, Buades JM, Boujnah N, Bouhlel MS (2020) Prediction of human activities based on a new structure of skeleton features and deep learning model. Sensors (Switzerland). 20:1–15
115. Tianming Z, Pengbiao Z, Peng X, Bintao W (2021) Multi-stream cnn-lstm network with partition strategy for human action recognition. Association for Computing Machinery, New York, NY, USA, pp 431–435

116. Lang X, Feng Z, Yang X (2020) Research on human-robot natural interaction algorithm based on body potential perception. Association for Computing Machinery, New York, NY, USA, pp 260–264
117. Lumetzberger J, Raoofpour A, Kampel M (2021) Privacy preserving getup detection. Association for Computing Machinery, New York, NY, USA, pp 234–243
118. Popescu A-C, Mocanu I, Cramariuc B (2020) Fusion mechanisms for human activity recognition using automated machine learning. IEEE Access 8:143996–144014
119. Badarch L, Gochoo M, Batnasan G, Alnajjar F, Tan T-H (2021) Ultra-low resolution infrared sensor-based wireless sensor network for privacy-preserved recognition of daily activities of living. Institute of Electrical and Electronics Engineers Inc., Boston, MA, USA
120. Mathe E, Tranou A, Spyrou E, Perantonis S (2020) Human action recognition with deep learning techniques. Association for Computing Machinery, New York, NY, USA
121. Giannakos I, Mathe E, Spyrou E, Mylonas P (2021) A study on the effect of occlusion in human activity recognition. Association for Computing Machinery, New York, NY, USA, pp 473–482
122. Rafferty J, Medina-Quero J, Quinn S, Saunders C, Ekerete I, Nugent C, Synnott J, Garcia-Constantino M (2019) Thermal vision based fall detection via logical and data driven processes. Institute of Electrical and Electronics Engineers Inc., Honolulu, HI, USA, pp 35–40
123. Awal MI, Iksan LH, Fhamy RZ, Basuki DK, Sukaridhoto S, Wada K (2021) Action recognition with spatiotemporal analysis and support vector machine for elderly monitoring system. Institute of Electrical and Electronics Engineers Inc., Surabaya, Indonesia, pp 470–475
124. Wang H, Gao Z, Lin W (2019) A fall detection system based on convolutional neural networks. Association for Computing Machinery, New York, NY, USA, pp 242–246
125. Sivakumar M, Iswarya E, Malusha K, Priyadharshini TY (2021) Computer vision based wellness analysis of geriatrics. Institute of Electrical and Electronics Engineers Inc., Coimbatore, India, pp 1762–1765
126. Wang F, Liu J, Hu GD (2019) A novel indoor human fall detection method based on an end-to-end neural network and bagged tree classifier. Association for Computing Machinery, New York, NY, USA, pp 384–389
127. Iksan LH, Awal MI, Fhamy RZ, Pratama AA, Basuki DK, Sukaridhoto S (2021) Implementation of cloud based action recognition backend platform, pp 1–6
128. Brieva J, Ponce H, Moya-Albor E, Martinez-Villasenor L (2019) An intelligent human fall detection system using a vision-based strategy. Institute of Electrical and Electronics Engineers Inc., Los Alamitos, California, USA, pp 1–5
129. Nambissan GS, Mahajan P, Sharma S, Gupta N (2021) The variegated applications of deep learning techniques in human activity recognition. Association for Computing Machinery, New York, NY, USA, pp 223–233
130. Hassan MFA, Hussain A, Saad MHM, Yusof Y (2019) Convolution neural network-based action recognition for fall event detection. International Journal of Advanced Trends in Computer Science and Engineering 8
131. Tan T-H, Hus J-H, Liu S-H, Huang Y-F, Gochoo M (2021) Using direct acyclic graphs to enhance skeleton-based action recognition with a linear-map convolution neural network. Sensors 21
132. Mohamed NA, Zulkifley MA, Kamari NAM (2019) Convolutional neural networks tracker with deterministic sampling for sudden fall detection. Institute of Electrical and Electronics Engineers Inc., New York, NY, USA, pp 1–5
133. Byeon Y-H, Kim D, Lee J, Kwak K-C (2021) Body and hand-object roi-based behavior recognition using deep learning. Sensors 21:1–23
134. Kumar D, Ravikumar AK, Dharmalingam V, Kafle VP (2019) Elderly health monitoring system with fall detection using multi-feature based person tracking. Institute of Electrical and Electronics Engineers Inc., Atlanta, GA, USA, pp 1–9
135. Hwang H, Jang C, Park G, Cho J, Kim I-J (2021) Eldersim: A synthetic data generation platform for human action recognition in eldercare applications. IEEE Access 1
136. Maldonado-Bascón S, Iglesias-Iglesias C, Martín-Martín P, Lafuente-Arroyo S (2019) Fallen people detection capabilities using assistive robot. Electronics (Switzerland) 8
137. Han K, Yang Q, Huang Z (2020) A two-stage fall recognition algorithm based on human posture features. Sensors (Switzerland) 20:1–21
138. Ferooz F, Ashraf MA, Hussain W, Butt AH, Khan YD (2019) Person fall recognition by using deep learning: Convolutional neural networks and image category classification using bag of feature, pp 1–6
139. Chiang JWH, Zhang L (2020) Deep learning-based fall detection, vol 12, pp 891–898. WORLD SCIENTIFIC PUBL CO PTE LTD. Singapore
140. Safarzadeh M, Alborzi Y, Ardekany AN (2019) Real -time fall detection and alert system using pose estimation, pp 508–511
141. Serpa YR, Nogueira MB, Neto PPM, Rodrigues MAF (2020) Evaluating pose estimation as a solution to the fall detection problem. Institute of Electrical and Electronics Engineers Inc., New York, NY, USA, pp 1–7
142. Huang Z, Liu Y, Fang Y, Horn BKP (2019) Video-based fall detection for seniors with human pose estimation. Institute of Electrical and Electronics Engineers Inc., Boston, MA, USA
143. Berardini D, Moccia S, Migliorelli L, Pacifici I, Massimo PD, Paolanti M, Frontoni E(2020) Fall detection for elderly-people monitoring using learned features and recurrent neural networks. Experimental Results 1
144. Ma C, Shimada A, Uchiyama H, Nagahara H, Taniguchi R-i (2019) Fall detection using optical level anonymous image sensing system. Opt Laser Technol 110:44–61
145. Romaissa BD, Mourad O, Brahim N, Yazid B (2020) Fall detection using body geometry in video sequences, pp 1–5
146. Cameiro SA, Silva GPD, Leite GV, Moreno R, Guimaraes SJF, Pedrini H (2019) Multi-stream deep convolutional network using high-level features applied to fall detection in video sequences, vol 2019-June, pp 293–298. IEEE Computer Society, New York, NY, USA
147. Wang X, Jia K (2020) Human fall detection algorithm based on yolov3. Institute of Electrical and Electronics Engineers Inc., Beijing, China, pp 50–54
148. Siriwardhana C, Madhuranga D, Madushan R, Gunasekera K (2019) Classification of activities of daily living based on depth sequences and audio. Institute of Electrical and Electronics Engineers Inc., Kandy, Sri Lanka, pp 278–283
149. Lv X, Gao Z, Yuan C, Li M, Chen C (2020) Hybrid real-time fall detection system based on deep learning and multi-sensor fusion. Institute of Electrical and Electronics Engineers Inc., Shenzhen, China, pp 386–391
150. Phyo CN, Zin TT, Tin P (2019) Deep learning for recognizing human activities using motions of skeletal joints. IEEE Trans Consum Electron 65:243–252
151. Li J, Xia S-T, Ding Q (2020) Multi-level recognition on falls from activities of daily living. Association for Computing Machinery, New York, NY, USA, pp 464–471
152. Saini R, Kumar P, Kaur B, Roy PP, Dogra DP, Santosh KC (2019) Kinect sensor-based interaction monitoring system using

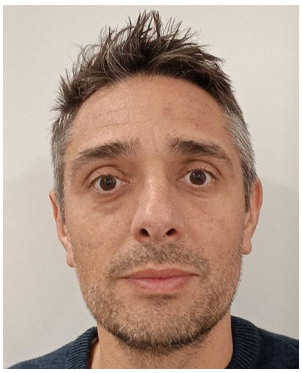
the blstm neural network in healthcare. Int J MachLearn Cybern 10:2529–2540
153. Chen Y, Li W, Wang L, Hu J, Ye M (2020) Vision-based fall event detection in complex background using attention guided bi-directional lstm. IEEE Access 8:161337–161348
154. Das S, Dai R, Koperski M, Minciullo L, Garattoni L, Bremond F, Francesca G (2019) Toyota smarthome: Real-world activities of daily living, vol. 2019-October, pp 833–842. Institute of Electrical and Electronics Engineers Inc., Seoul, Korea (South)
155. Khraief C, Benzarti F, Amiri H (2020) Elderly fall detection based on multi-stream deep convolutional networks. Multimed Tools Appl 79:19537–19560
156. Phyo CN, Zin TT, Tin P (2019) Complex human-object interactions analyzer using a dcnn and svm hybrid approach. Applied Sciences (Switzerland) 9
157. Kharazian Z, Rahat M, Fatemizadeh E, Nasrabadi AM (2020) Increasing safety at smart elderly homes by human fall detection from video using transfer learning approaches. Research Publishing Services, Venice, Italy, pp 2774–2780
158. Nan M, Ghiță AS, Gavril A-F, Trascau M, Sorici A, Cramariuc B, Florea AM (2019) Human action recognition for social robots, pp 675–681
159. Tateno S, Meng F, Qian R, Li T (2020) Human motion detection based on low resolution infrared array sensor, pp 1016–1021
160. Mehr HD, Polat H (2019) Human activity recognition in smart home with deep learning approach. IEEE, New York, NY, USA, pp 149–153
161. Buzzelli M, Albé A, Ciocca G (2020) A vision-based system for monitoring elderly people at home. Applied Sciences (Switzerland) 10
162. Jalal A, Mahmood M, Hasan AS (2019) Multi-features descriptors for human activity tracking and recognition in indoor-outdoor environments, pp 371–376
163. Tan T-H, Gochoo M, Chen H-S, Liu S-H, Huang Y-F (2020) Activity recognition based on dcnn and kinect rgb images, pp 1–4
164. Ding Q, Yang F, Li J, Wu S, Zhao B, Wang Z, Xia S-T (2019) Rtadi: Fast real-time video representation for multi-view human fall detection. Institute of Electrical and Electronics Engineers Inc., New York, NY, USA, pp 13–18
165. Atikuzzaman M, Rahman TR, Wazed E, Hossain MP, Islam MZ (2020) Human activity recognition system from different poses with cnn. Institute of Electrical and Electronics Engineers Inc., New York, NY, USA, pp 1–5
166. Priya GGL, Jain M, Santosh KC, Mouli PVSSRC (2019) Temporal super-pixel based convolutional neural network (ts-cnn) for human activity recognition in unconstrained videos. Commun Comput Inf Sci 1035:255–264
167. Gul MA, Yousaf MH, Nawaz S, Rehman ZU, Kim H (2020) Patient monitoring by abnormal human activity recognition based on cnn architecture. Electronics (Switzerland) 9:1–14
168. Martínez-Villaseñor L, Ponce H, Brieva J, Moya-Albor E, Núñez-Martínez J, Peñafort-Asturiano C (2019) Up-fall detection dataset: A multimodal approach. Sensors 19
169. Cao Z, Hidalgo Martinez G, Simon T, Wei S, Sheikh YA (2019) Openpose: Realtime multi-person 2d pose estimation using part affinity fields. IEEE Transactions on Pattern Analysis and Machine Intelligence
170. Fang H-S, Li J, Tang H, Xu C, Zhu H, Xiu Y, Li Y-L, Lu C (2022) Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. IEEE Transactions on Pattern Analysis and Machine Intelligence
171. Lugaresi C, Tang J, Nash H, McClanahan C, Uboweja E, Hays M, Zhang F, Chang C-L, Yong MG, Lee J, Chang W-T, Hua W, Georg M, Grundmann M (2019) MediaPipe: A Framework for Building Perception Pipelines. arXiv
172. Papandreou G, Zhu T, Chen L-C, Gidaris S, Tompson J, Murphy K (2018) Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model. In: Computer Vision – ECCV 2018, pp 282–299. Springer, Cham
173. TensorFlow: MoveNet. https://github.com/tensorflow/tfjs-models/tree/master/pose-detection/src/movenet
174. Fang H-S, Xie S, Tai Y-W, Lu C (2017) Rmpe: Regional multi-person pose estimation. In: 2017 IEEE international conference on computer vision (ICCV), pp 2353–2362
175. Xiu Y, Li J, Wang H, Fang Y, Lu C (2018) Pose flow: Efficient online pose tracking. In: BMVC
176. Zhang J, Zhu Z, Zou W, Li P, Li Y, Su H, Huang G (2019) FastPose: Towards Real-time Pose Estimation and Tracking via Scale-normalized Multi-task Networks
177. Howard A, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications
178. Luvizon DC, Picard D, Tabia H (2018) 2D/3D Pose Estimation and Action Recognition Using Multitask Deep Learning. https://doi.org/10.1109/CVPR.2018.00539
179. Duan H, Zhao Y, Chen K, Lin D, Dai B (2022) Revisiting skeleton-based action recognition. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 2959–2968. https://doi.org/10.1109/CVPR52688.2022.00298
180. Li S, Man C, Shen A, Guan Z, Mao W, Luo S, Zhang R, Yu H (2022) A fall detection network by 2d/3d spatio-temporal joint models with tensor compression on edge. ACM Trans Embed Comput Syst 21
181. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Sci 313(5786):504–507
182. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville, A., Bengio, Y (2014) Generative Adversarial Nets. Curran Associates, Inc
183. Koch GR (2015) Siamese neural networks for one-shot image recognition
184. Nogas J, Khan S, Mihailidis A (2020) Deepfall: Non-invasive fall detection with deep spatio-temporal convolutional autoencoders. Journal of Healthcare Informatics Research. 4
185. Hampiholi B, Jarvers C, Mader W, Neumann H (2020) Depthwise separable temporal convolutional network for action segmentation. In: 2020 International conference on 3D vision (3DV), pp 633–641
186. Wang Q, Gao J, Xing J, Zhang M, Hu W (2017) DCFNet: Discriminant Correlation Filters Network for Visual Tracking. arXiv
187. Iandola F, Han S, Moskewicz M, Ashraf K, Dally W, Keutzer K (2016) Squeezenet: Alexnet-level accuracy with 50x fewer parameters and ¡0.5mb model size
188. Tan M, Le QV (2020) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks
189. Tran D, Bourdev L, Fergus R, Torresani L, Paluri M (2015) Learning spatiotemporal features with 3d convolutional networks, pp 4489–4497
190. Girshick R, Donahue J, Darrell T, Malik J (2013) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition
191. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: Unified, real-time object detection, pp 779–788
192. Kirillov A, Wu Y, He K, Girshick R (2020) Pointrend: Image segmentation as rendering. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 9796–9805. https://doi.org/10.1109/CVPR42600.2020.00982
193. Hui T-W, Tang X, Loy CC (2018) LiteFlowNet: A lightweight convolutional neural network for optical flow estimation

194. Feichtenhofer C, Fan H, Malik J, He K (2019) SlowFast Networks for Video Recognition. https://doi.org/10.1109/ICCV.2019.00630
195. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2818–2826
196. Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory. Neural Comput 9(8):1735–1780
197. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
198. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778
199. Kipf TN, Welling M (2017) Semi-supervised classification with graph convolutional networks
200. Carreira J, Zisserman A (2017) Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. https://doi.org/10.1109/CVPR.2017.502
201. Cho K, Merrienboer B, Bahdanau D, Bengio Y (2014) On the properties of neural machine translation: Encoder-decoder approaches
202. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, Kaiser L, Polosukhin I (2017) Attention is all you need
203. Liu Z, Wang L, Wu W, Qian C, Lu T (2021) Tam: Temporal adaptive module for video recognition. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp 13688–13698. https://doi.org/10.1109/ICCV48922.2021.01345
204. Yang C, Xu Y, Shi J, Dai B, Zhou B (2020) Temporal pyramid network for action recognition. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 588–597. https://doi.org/10.1109/CVPR42600.2020.00067
205. Gao C, Zou Y, Huang J (2018) iCAN: instance-centric attention network for human-object interaction detection. BMVA Press
206. Chollet F (2017) Xception: Deep learning with depthwise separable convolutions. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), pp 1800–1807. https://doi.org/10.1109/CVPR.2017.195
207. Wang L, Xiong Y, Wang Z, Qiao Y, Lin D, Tang X, Van Gool L (2016) Temporal segment networks: Towards good practices for deep action recognition. In: Leibe B, Matas J, Sebe N, Welling M (eds) Computer Vision - ECCV 2016, pp 20–36. Springer, Cham
208. Liu Z, Ning J, Cao Y, Wei Y, Zhang Z, Lin S, Hu H (2022) Video swin transformer. In: 2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 3192–3201. https://doi.org/10.1109/CVPR52688.2022.00320
209. Bertasius G, Wang H, Torresani L (2021) Is space-time attention all you need for video understanding? In: ICML, vol 2, pp 4
210. Baradel F, Wolf C, Mille J, Taylor G (2018) Glimpse clouds: Human activity recognition from unstructured feature points, pp 469–478
211. Tang J, Xia J, Mu X, Pang B, Lu C (2020) Asynchronous interaction aggregation for action detection. In: Computer Vision – ECCV 2020, pp 71–87. Springer, Cham
212. Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. Commun ACM 60(6):84–90. https://doi.org/10.1145/3065386
213. Lu L, Lu Y, Yu R, Di H, Zhang L, Wang S (2020) Gaim: Graph attention interaction model for collective activity recognition. IEEE Trans Multimed 22(2):524–539. https://doi.org/10.1109/TMM.2019.2930344
214. He K, Gkioxari G, Dollár P, Girshick R (2020) Mask r-cnn. IEEE Trans Pattern Anal Mach Intell 42(2):386–397. https://doi.org/10.1109/TPAMI.2018.2844175
215. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N (2021) An image is worth 16x16 words: Transformers for image recognition at scale. In: International conference on learning representations
216. Kwolek B, Kepski M (2014) Human fall detection on embedded platform using depth maps and wireless accelerometer. Comput Methods Programs Biomed 117:489–501
217. Charfi I, Miteran J, Dubois J, Atri M, Tourki R (2013) Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and adaboost-based classification. J Electron Imaging 22:41106
218. Auvinet E, Rougier C, Meunier J, St-Arnaud A, Rousseau J (2011) Multiple cameras fall data set
219. Shahroudy A, Liu J, Ng T, Wang G (2016) Ntu rgb+d: A large scale dataset for 3d human activity analysis
220. Adhikari K, Bouchachia H, Nait-Charif H (2017) Activity recognition for indoor fall detection using convolutional neural network. Institute of Electrical and Electronics Engineers Inc., New York, NY, USA, pp 81–84
221. Wang J, Liu Z, Wu Y, Yuan J (2012) Mining actionlet ensemble for action recognition with depth cameras, pp 1290–1297
222. Chen C, Jafari R, Kehtarnavaz N (2015) Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor, pp 168–172
223. Sung J, Ponce C, Selman B, Saxena A (2012) Unstructured human activity detection from rgbd images, pp 842–849
224. Jang J, Kim D, Park C, Jang M, Lee J, Kim D (2020) Etri-activity3d: A large-scale rgb-d dataset for robots to recognize daily activities of the elderly, pp 10990–10997
225. Kuehne H, Jhuang H, Garrote E, Poggio T, Serre T (2011) Hmdb: A large video database for human motion recognition, pp 2556–2563
226. Schuldt C, Laptev I (2004) Caputo B. Recognizing human actions: a local svm approach vol 3, pp 32–363
227. Koppula H, Gupta R, Saxena A (2012) Learning human activities and object affordances from rgb-d videos. The International Journal of Robotics Research 32
228. Amiri SM, Pourazad MT, Nasiopoulos P, Leung VCM (2013) Non-intrusive human activity monitoring in a smart home environment, pp 606–610
229. Vanrumste B, Debard G, Croonenborghs T, Mertes G, Baldewijns G (2016) Bridging the gap between real-life data and simulated data by providing a highly realistic fall dataset for evaluating camera-based fall detection algorithms. Healthcare Technology Letters 3
230. Liu J, Shahroudy A, Perez M, Wang G, Duan L-Y, Kot A (2019) Ntu rgb+d 120: A large-scale benchmark for 3d human activity understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp 1–18
231. wang J, Nie X, Xia Y, Wu Y, Zhu S (2014) Cross-view action modeling, learning, and recognition. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition
232. Soomro K, Zamir A, Shah M (2012) Ucf101: A dataset of 101 human actions classes from videos in the wild. CoRR
233. Xia L, Chen C-C, Aggarwal JK (2012) View invariant human action recognition using histograms of 3d joints, pp 20–27
234. Rahmani H, Mahmood A, Huynh D, Mian A (2016) Histogram of oriented principal components for cross-view action recognition. IEEE Trans Pattern Anal Mach Intell 38:2430–2443
235. Kong Y, Jia Y, Fu Y (2012) Learning human interaction by interactive phrases, pp 300–313
236. Wei H, Chopada P, Kehtarnavaz N (2020) C-mhad: Continuous multimodal human action dataset of simultaneous video and inertial sensing. Sensors 20
237. Chen Z (2019) Fall detection dataset

238. Gasparrini S, Cippitelli E, Gambi E, Spinsante S, Wåhslén J, Orhan I, Lindh T (2016) Proposal and experimental evaluation of fall detection solution based on wearable and depth data fusion, vol 399, pp 99–108
239. Maldonado-Bascón S, Iglesias-Iglesias C, Martín-Martín P, Lafuente-Arroyo S (2021) Elderly Dataset
240. Weinland D, Ronfard R, Boyer E (2006) Free viewpoint action recognition using motion history volumes. Comp Vision Image Underst 104:249–257
241. Kay W, Carreira J, Simonyan K, Zhang B, Hillier C, Vijayanarasimhan S, Viola F, Green T, Back T, Natsev P, Suleyman M, Zisserman A (2017) The kinetics human action video dataset. arXiv
242. Carreira J, Noland E, Banki-Horvath A, Hillier C, Zisserman A (2018) A Short Note about Kinetics-600. arXiv
243. Smaira L, Carreira J, Noland E, Clancy E, Wu A, Zisserman A (2020) A Short Note on the Kinetics-700-2020 Human Action Dataset. arXiv
244. Chua J-L, Chang Y, Lim W (2013) A simple vision-based fall detection technique for indoor video surveillance. Signal, Image and Video Processing, p 9
245. Chen Y, Yu L, Ota K, Dong M (2018) Robust activity recognition for aging society. IEEE J Biomed Health Inform 22:1754–1764
246. Epstein D, Chen B, Vondrick C (2020) Oops! predicting unintentional action in video. https://doi.org/10.1109/CVPR42600.2020.00100
247. Liu C, Hu Y, Li Y, Song S, Liu J (2017) PKU-MMD: A Large Scale Benchmark for Continuous Multi-Modal Human Action Understanding. arXiv
248. Yao B, Jiang X, Khosla A, Lin A, Guibas L, Li F-F (2011) Human action recognition by learning bases of action attributes and parts, pp 1331–1338
249. Gupta S, Malik J (2015) Visual Semantic Role Labeling. arXiv
250. Fan Y, Levine M, Gongjian W, Qiu S (2017) A deep neural network for real-time detection offalling humans in naturally occurring scenes. Neurocomputing. 260
251. Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, Moore R, Kipman A, Blake A (2011) Real-time human pose recognition in parts from single depth images. CVPR 2011:1297–1304. https://doi.org/10.1109/CVPR.2011.5995316

F. Xavier Gaya-Morey is a Ph. D. candidate and professor at the University of the Balearic Islands. He holds a bachelor's degree in computer engineering and a master's degree in data science and computer vision. His research is centered on the areas of explainable artificial intelligence and computer vision, with special focus on their applications to improve the life quality of the older adults.

Cristina Manresa-Yee received her degree in Computer Science and her Ph. D. in Computer Science from the University of Balearic Islands. She is currently an Associate Professor at the University of the Balearic Islands. Her research interests include human-computer interaction, computer vision and explainable AI.



José M. Buades-Rubio received his degree in Computer Science and his Ph. D. in Computer Science from the University of Balearic Islands. He is currently an Associate Professor at the University of the Balearic Islands. His research interests include computer graphics, computer vision and artificial intelligence.