



Artificial Intelligence for skeleton-based physical rehabilitation action evaluation: A systematic review

Sara Sardari^{a,b,*}, Sara Sharifzadeh^c, Alireza Daneshkhah^a, Bahareh Nakisa^b, Seng W. Loke^b, Vasile Palade^a, Michael J. Duncan^d

^a Centre for Computational Science & Mathematical Modelling, Coventry University, Coventry, UK

^b School of Information Technology, Faculty of Science Engineering and Built Environment, Deakin University, Geelong, Vic, Australia

^c Department of Computer Science, Swansea University, Swansea, UK

^d Centre for Sport, Exercise and Life Sciences, Coventry University, Coventry, UK

ARTICLE INFO

Keywords:

Activity evaluation
Activity recognition
Computer vision
Deep learning
Physical rehabilitation
Skeleton data

ABSTRACT

Performing prescribed physical exercises during home-based rehabilitation programs plays an important role in regaining muscle strength and improving balance for people with different physical disabilities. However, patients attending these programs are not able to assess their action performance in the absence of a medical expert. Recently, vision-based sensors have been deployed in the activity monitoring domain. They are capable of capturing accurate skeleton data. Furthermore, there have been significant advancements in Computer Vision (CV) and Deep Learning (DL) methodologies. These factors have promoted the solutions for designing automatic patient's activity monitoring models. Then, improving such systems' performance to assist patients and physiotherapists has attracted wide interest of the research community. This paper provides a comprehensive and up-to-date literature review on different stages of skeleton data acquisition processes for the aim of physio exercise monitoring. Then, the previously reported Artificial Intelligence (AI) - based methodologies for skeleton data analysis will be reviewed. In particular, feature learning from skeleton data, evaluation, and feedback generation for the purpose of rehabilitation monitoring will be studied. Furthermore, the associated challenges to these processes will be reviewed. Finally, the paper puts forward several suggestions for future research directions in this area.

1. Introduction

With the recent advances in medical science and related technologies, the elderly population in developed countries, such as the UK and Australia, is growing. According to the Australian Bureau of Statistics [1], the proportion of older adults in Australia (aged 65 and over) is predicted to grow from 15% (3.8 million) of the whole population in 2017 to 22% (8.8 million) in 2025. Moreover, according to the Office for National Statistics [2], due to the recent advances in healthcare, the population of people over 60 in the UK is increasing from 14.9 million in 2014 to 18.5 million in 2025. According to Cameron and Kurle [3], the possibility of being physically disabled due to different medical conditions, such as stroke or a hip fracture, is higher for older adults. Therefore, a new challenge is emerging for the healthcare systems in developed countries, since the increase in the aging rate is associated with the decline in the physical ability of the aging population. Therefore, they are going to make up one of the largest groups of people participating in physical rehabilitation programs.

Patients with physical disabilities are usually prescribed (by physiotherapists or occupational therapists) to attend different rehabilitation programs either in medical agencies (hospitals) or at home. Each of these methods of rehabilitation at the home or a medical center has its advantages and limitations. During the inpatient rehabilitation period, the performances of the patients are being monitored by the experts, and they are provided with prompt feedback. However, depending on their performance, the patients might need to attend the program for several sessions. Attending these programs with the supervision of an expert is expensive, time-consuming, and tedious, since it involves transportation and inpatient medical services. Therefore, there is a preference for a majority of patients for home-based rehabilitation [3]. Moreover, the inpatient programs usually include long waiting lists due to factors such as shortage of staff and long waiting time for treatment leading to poor health improvement in patients [4]. In addition, the trend of home-based rehabilitation has increased after the prolonged COVID-19 pandemic in 2020 and the closure of rehabilitation centers

* Corresponding author at: Centre for Computational Science & Mathematical Modelling, Coventry University, Coventry, UK.

E-mail address: sardaris@coventry.ac.uk (S. Sardari).

or their limited programs [5]. According to Frigerio et al. [6], the implementation of tele-rehabilitation during the COVID-19 lockdown has gained excellent satisfaction from the patients' side and is a promising tool to be used after the pandemic.

In addition to all of the aforementioned positive impacts of home-based rehabilitation, it is worth mentioning that the quality of the rehabilitation program plays an important role in the extent of the neuroplasticity achieved by the patients [7]. However, the lack of feedback and tedious home settings can have a demotivating effect on the patients in home-based rehabilitation programs and may affect the final outcome [8]. According to Gelaw et al. [9], to yield positive results from home-based physical rehabilitation programs, patient enthusiasm and continuous follow-up from the side of an expert are essential. To follow up on the progress of the patients, the therapists use web-based tele-rehabilitation programs in which, they monitor the actions and provide online feedback. However, due to several challenges such as shortage of staff and lack of time, there are challenges in ensuring consistency of online monitoring [10]. Therefore, the need for computer-aided home-based rehabilitation programs for inexpensive and private training sessions with online feedback is increasing. These computer-aided therapy sessions will utilize the sensors and human activity analysis algorithms to guide the patients in performing the actions properly, and assist the healthcare providers in monitoring patients' recovery status. The specific aim of this paper is to investigate different automatic human activity analysis techniques in the literature for monitoring home-based rehabilitation exercises, and to explore the challenges and limitations of such techniques for further research recommendations.

In general, human activity analysis is one of the most important and challenging areas in AI. It involves analyzing human body movements based on the motions of different body joints, skeleton, and muscles [11]. Based on the complexity of the action, these movements can be interpreted as different gestures, human-human interactions, group actions, and behaviors [12]. Analyzing these activities can provide useful information about the personality of individuals, their physiological and psychological states, and possibly their targets and intentions. Recently, there is a growing interest in developing and using automatic human activity analysis systems, which can assist experts in different tasks for health-care [13,14], surveillance in public places [15,16], and developing driver-less systems [17,18]. However, developing an effective system for human activity analysis highly depends on how accurate are the motion tracking, data pre-processing, representation learning, and evaluation techniques [11]. Therefore, while looking for a problem definition for activity analysis, four main questions arise: "Q1: What is the task that we are targeting in human activity analysis?", "Q2: What are the actions and input modalities for an automated system?", "Q3: What are the automatic learning strategies that we can consider for the problem?", and "Q4: What are the evaluation techniques for the performance of an automated system?". An overview of these questions as a pipeline of this study follows:

Q1: Human activity analysis encompasses several general tasks, such as Human Activity Recognition (HAR), Human Activity Detection (HAD), Human Activity Prediction (HAP), and finally Human Activity Evaluation (or assessment) (HAE). One of the popular fields of study is the traditional HAR problem, which involves action classification, based on a developed system that can assign class labels to different action categories, based on the different input modalities. HAR has been widely explored by researchers in various domains of healthcare [19,20], driverless cars [21], surveillance systems for public areas/home/organizations [22], smart home/city [23–25], etc. HAD aims to assign starting and ending points (labels) to the performed actions. Assigning these points to an untrimmed video has attracted much attention due to its real-world applications in detecting and managing abnormal dangerous situations in traffic or public [26,27]. HAP refers to developing a model which can predict future actions (states) of a series of actions based on the previous incomplete observations.

Predicting the next states of many real-world actions and behaviors can prevent hazardous situations, such as careless driving, terrorist attacks, or even fall prediction in the daily living of elderly people [11,28,29]. Finally, in contrast to all of the tasks mentioned above, HAE aims to assess the performed actions by individuals based on some reference correct actions and provide some feedback (such as scores) to improve the quality of the actions. According to Lei et al. [11], this field of study has begun to attract many researchers in the community because of its important real-world applications, such as skill training for different expertise learners [30,31], sports activity assessment [32,33], and physical activity rehabilitation [34,35]. In order to create an automatic physical activity monitoring system for a rehabilitation period, the HAE is the most important one to be considered. In other words, an ideal automatic monitoring device must be able to evaluate the action properly and then provide feedback on how the action can be more accurately performed.

Q2: As mentioned above, human motion analysis has multiple applications which include different real-world situations. Depending on which real-world problem we aim to solve by motion tracking and analysis, the individuals perform different activities, ranging from simple daily actions to complex and specific actions (such as sports activity and rehab prescriptions). According to Yadav et al. [36], depending on the complexity of the action and application we are aiming, the automatic human action analysis systems usually need large datasets containing different useful modalities, which aim to represent the performed actions in the best way. According to Sun et al. [37], human actions can be represented using several modalities such as vision-based (RGB videos/images, depth videos/images, skeleton/joint data sequences, InfraRed (IR) sequences) [28,38], wearable-based [39], radar-based [40,41], audio-based [42], and Wifi-based [43]. With the wide variety and accessibility of the sensors for capturing these modalities in the past decade, the investigations on designing automatic HAR/ HAE systems based on these data are growing [37]. However, all of these modalities capture various information about an action. Therefore, they have different levels of strengths and limitations, which are illustrated in Fig. 1. The most important factors which should be considered while selecting a modality for capturing actions are the sensor's cost, appropriate resolution depending on the application and the target activity, privacy-preserving, visual interpretability, and robustness towards any changes in the data collection conditions.

Comparing different techniques of data collection in Fig. 1 illustrates the fact that the skeleton/joint modality, which includes a sequence of the coordinates of human body joints, might be the best option considering all of the factors. Skeletal data has drawn much attention for the task of human activity recognition by many researchers [44–46], because of certain advantages that it has compared to other methods. According to Shi et al. [44], skeleton-based activity recognition stands out from the rest of the vision-based recognition methods because it shows robustness towards changes in body scales, speed of the performed activity, camera viewpoints, and the interference of backgrounds. This modality preserves visible information somehow, however, it is an affordable privacy-preserving technique to capture important structural body motion information. Considering the aforementioned advantages, this study specifically aims to dive deep into different capturing techniques for skeletal data and previous studies utilizing this modality.

Q3: One of the most important challenges in the human motion analysis pipeline is how to develop a system that has a robust representation/feature learning framework. The performance of any automatic recognition and evaluation model highly depends on the quality of features extracted that represent the data [47]. There are two general approaches for data representation and feature learning, i.e. hand-crafted features learning, and automatic feature learning using DL techniques [48]. Some of the popular methods for classical hand-crafted feature extraction are based on different modalities of depth, RGB, and skeleton data. For example, Depth Motion Map (DMM), Histogram of



Fig. 1. Strength and limitations of different vision-based and non-vision-based modalities in human action analysis.

Gradients (HoG), and local binary features [49] can be extracted from depth data. In other studies conducted by Xia et al. [50], 3D Histogram of Oriented Displacement (HOD) and Accumulation of Motion Energy (AME) are extracted as action features from the skeleton data. Using the handcrafted features techniques for data representation might need expert's knowledge or specific algorithms for each problem, which might lead us to less generalization in various problems. This means that by using a specific algorithm for extracting features in a specific modality for a specific problem, you may not be able to use the same pipeline for another problem. Motivated by the successful performance results of the use of DL techniques in various studies [51,52], this paper will mostly investigate different studies which used this method for skeleton data in the past ten years.

Q4: Finally, the evaluation techniques and criteria in the two tasks of HAR and HAE differ from each other, since they address two different problems of classification and regression, respectively. To the best of our knowledge, previous studies in the vision-based rehabilitation field rarely explored different evaluation methods for HAE problems. Therefore, throughout this study, we will explore different evaluation techniques, and examine their challenges, advantages, and limitations.

Exploring these four fundamental questions for the problem of automatic physical rehabilitation monitoring creates a workflow which is illustrated in Fig. 2. Examining each part of this pipeline helps us to find

existing challenges and limitations in the literature, and then find possible solutions for addressing them. In the data acquisition stage, sensor selection, ethics considerations, activity selection, and experiment design are the most important tasks to perform. After data collection, modality capturing and action labeling are the challenging stages. Next, the researchers should design the proper classification/regression model based on the HAE/HAR problem. Finally, the evaluation metrics and methods for the designed system should be considered to produce accurate results. This paper is organized based on exploring different solutions to these four questions and the way different related papers address this workflow.

The remainder of this paper is organized as follows: Section 2 clarifies the methodology for this literature review and its contributions compared to other related survey papers. In Section 3, we discuss the types of impairments and the target body parts for different rehabilitation exercises prescribed by the medical experts. Section 4 investigates the related challenges of data collection. Section 5 explores the methods for capturing skeleton data through sensing hardware. Section 6 provides a comparative analysis of different public datasets and discusses their limitations and strengths. Section 7 provides information of the AI-based methods for representation learning on the skeleton data and how these models can be evaluated. Section 8 provides information on how other studies evaluated the activities and annotated them.

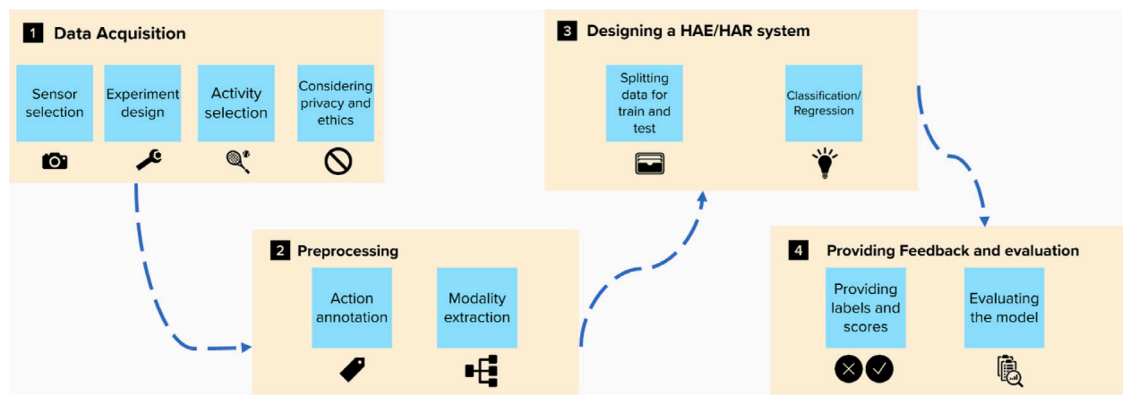


Fig. 2. Designing a human activity analysis system for automatically monitoring physical rehabilitation follows this framework from left to right.

Section 9 provides a brief discussion of the current challenges detected in the literature. Finally, Section 10 concludes the paper.

2. Methodology of the review and its contributions

In the last decade, several surveys and literature reviews have been published, aiming to review generally the vision-based studies for automatic physical activity recognition and evaluation for the rehabilitation period. However, each of them summarizes different scopes of studies and their limitations. In 2004, Zhou et al. [53] discussed different visual or non-visual human motion tracking sensors for rehabilitation exercises and compared these technologies. However, this study fails to discuss any AI-based algorithms for developing automatic recognition and assessment methods. In 2014, Webster et al. [54] investigated the applications of Microsoft Kinect sensors in elderly care, stroke rehabilitation, fall detection, and Kinect-based gaming. In the study conducted by Da Gama et al. [55], the authors mostly focused on providing a formulation of monitoring the progress in the rehabilitation using various techniques such as angle flexion, euclidean distance, etc. However, according to Debnath et al. [56], both of the previous studies which either study the recognition (prediction) or evaluation techniques for rehabilitation, have a clinical perspective for evaluation. To solve this problem, Sathyanarayana et al. [57] discussed the vision-based algorithms from the computer vision perspective for evaluation. In a recent study, Ahad et al. [58] provided a short review of vision-based action understanding for applications in assistive healthcare. They investigated general vision-based sensors (such as Vicon optical tracking system and depth sensors) and environmental scenarios (such as lighting conditions and background settings) for data collection, the challenges ahead of these data collection scenarios, and some benchmark datasets. However, this work lacks information about further technical methods for representation learning (such as different DL techniques, which can be utilized for this purpose) and evaluation methods for scoring the activities. The latest literature review for computer vision-based algorithms for rehabilitation and assessment is conducted in [56]. This paper discusses a wide range of general vision-based techniques for either recognizing or assessing rehabilitation exercises. However, due to the generality of this study, this paper reviews the previous studies in the field without covering the important materials such as the significance and limitations of different sensors, techniques, and scenarios for data collection. The possible physical rehabilitation exercises were not discussed also as a guideline for future work. In addition, there is very limited discussion on how different AI-based methods and evaluation techniques can improve the performance and feedback responses in the system. To solve all of these issues this paper covers a wide range of studies specific to skeleton-based activity assessment for rehabilitation problem. This study contributes the following:

- This paper comprehensively reviews the skeleton-based data collection procedures in relation to the sensor and physical activity selection. Different challenges for proper data collection are identified and the limitations of the previous related public datasets are discussed.
- This study specifically aims to provide an up-to-date and holistic literature review on the AI-based skeleton data analysis methods for the physical rehabilitation problem. To the best of our knowledge, this is the first time a study has been conducted on the strength and the gaps of HAE methods provided for this specific problem which paves the path for further studies.
- The evaluation techniques are comprehensively explored for (1) general automatic scoring systems and (2) part-based assessment for each activity. Furthermore, the gaps and limitations of those methods are discussed. This adds to the novelty of this paper compared to the previously conducted literature review papers.

This study includes a systematic literature review and it encompasses the most recent studies (between 2011 and 2022), related to developing AI-based technologies for automatic physical activity evaluation on rehabilitation. The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) [59] checklist is utilized to conduct the step-by-step research methodology. In the identification stage an appropriate search for articles has been performed through Google Scholar, Scopus, and Science Direct, PubMed databases as illustrated in Fig. 3. Based on the research question of building skeleton-based automatic human activity evaluation systems for rehabilitation problems, we utilized Boolean search strings such as “Activity Recognition”, “Skeleton-based activity assessment”, “Human activity evaluation”, “Kinect sensors”, and “Rehabilitation” in different combinations. For example, when using the Science Direct database for a combination of “Skeleton-based activity assessment for rehabilitation”, the number of retrieved articles was 1439. The number of articles retrieved from Google Scholar, Science Direct, Scopus, and PubMed are 6012, 2312, 1273, and 1310, respectively. To avoid duplication among all of the retrieved 10907 articles, Mendeley software was utilized and this resulted in the final 6091 articles. In the next stage, the articles were screened to omit the unrelated studies to the content of the research question based on several conditions. For this step, titles, abstracts, language (only English is included), number of citations, and journal (or conference) quality were considered which led to only 942 papers. To check the eligibility for inclusion, full texts of records were screened and several records were excluded due to their irrelevance to the aim and scope of this review paper, duplicated information from another reviewed literature, and lack of detailed discussion and evaluation. To double-check the precision of this search, a software called Publish-or-Perish (PoP) was used. It provides a specific search on specified keywords which results in limited, but more related search results. Finally, the 32 most related papers to the scope of this study including studies related

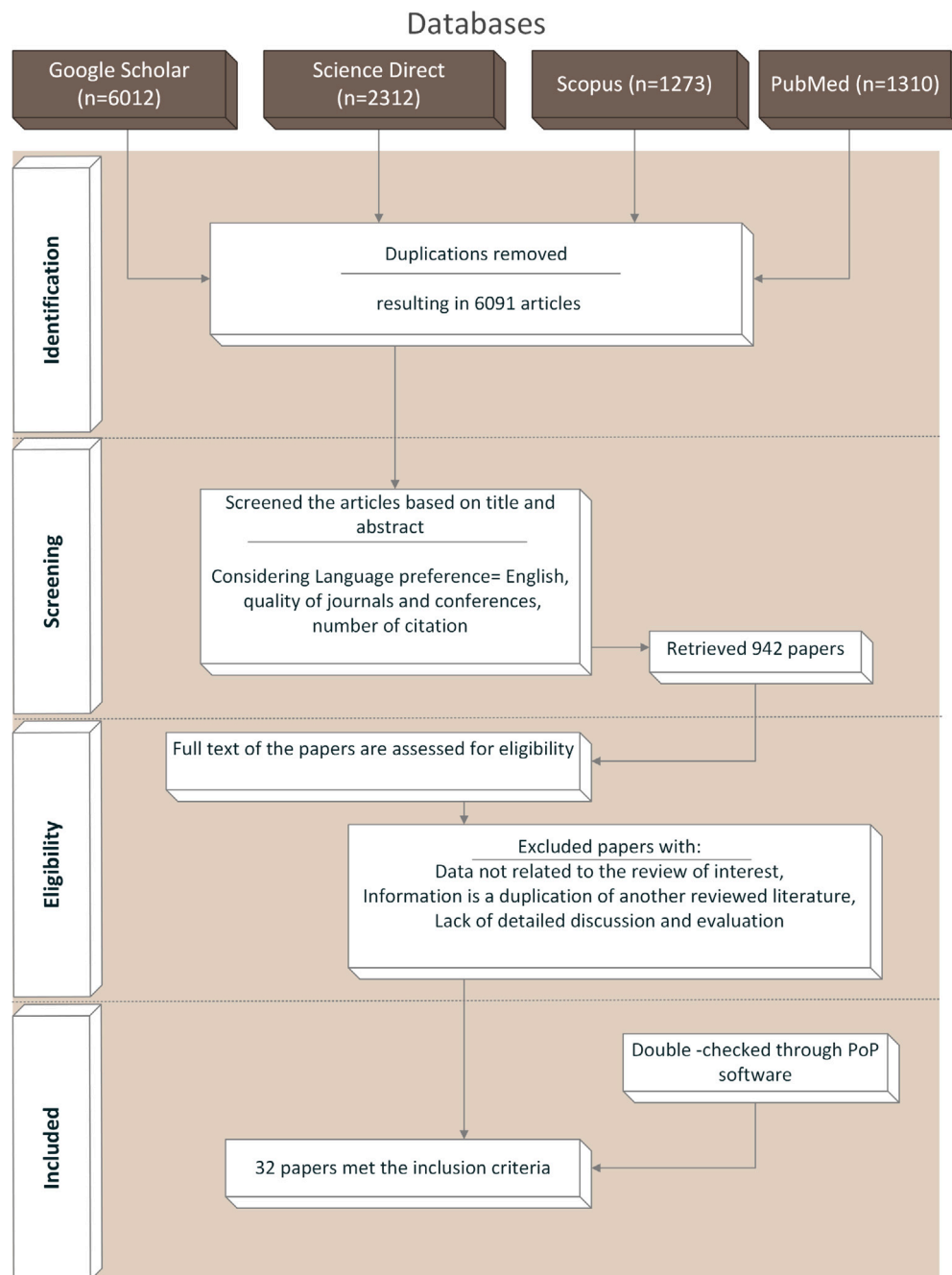


Fig. 3. The retrieval methodology used for finding, evaluating, and including the related articles in this review.

to previous literature reviews [53–58], datasets (Table 3), and AI-based methodologies (Table 4), were retrieved and included in the review. It is worth mentioning that one review study from 2008 [53] is added to this paper because of the importance of the paper to the coherence of this systematic review.

3. Rehabilitation and physical exercises

Physical disability and impairment are defined as limitations in the individual's physical functionality, mobility, or stamina which can be temporary or permanent for the long term and hinder them from daily normal activities [60]. In general, physical disabilities include activity, mobility, visual, or auditory impairments, or chronic pains causing difficulty in functioning. They may occur (especially in older

individuals) due to different neurological conditions such as stroke and Parkinson's disease or different injuries such as spinal cord injuries, brain injuries, and hip fractures. To be specific, the physical impairments can be categorized into two general groups: musculoskeletal and neuromusculoskeletal [61]; musculoskeletal disabilities affect the joint, skeleton, and muscle movements directly due to different reasons such as back and neck pain, osteoarthritis, and bone fractures and injuries. The neuromusculoskeletal group includes impairments caused by neurological conditions such as stroke, cerebral palsy, poliomyelitis, spinal cord/brain injuries, and Parkinson's disease. These types of disorders affect the nervous system which controls the muscles and bones and their interaction with the brain [62]. Fig. 4 illustrates that these disabilities might occur in both upper and lower limbs which are divided by the hip joint [63].

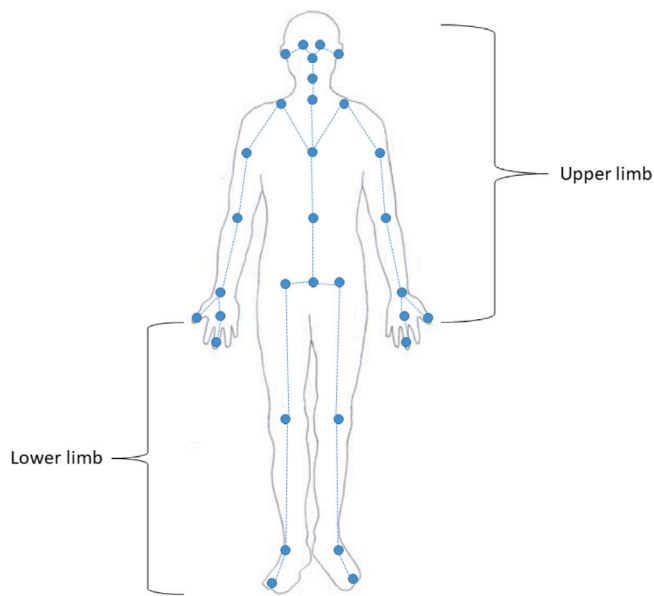


Fig. 4. The 32 body joints and skeletons from both upper limb and lower limb captured by Azure Kinect.

To overcome the challenges of different impairments in daily life, physical rehabilitation programs are provided by the healthcare systems in most developed countries. The role of rehabilitation programs is to improve the physical functionality of temporary cases of disabilities and to define a need and care routine for permanent types of disabilities. Along with pharmacological treatments most of the physical rehabilitation programs encompass different physical exercise therapies which aim to prepare patients with disabilities for normal daily activities. These exercises may be prescribed by the expert tool/weight-free or with the use of therabands or weights, based on the need and facilities of the patients and the type of impairments. The healthcare provider team usually monitors the physical activities based on several scoring and evaluation questionnaires and methods [64–66].

There is a vast set of exercises for the purpose of physical rehabilitation. However, some of these exercises are more common and suitable for the purpose of data collection for developing a vision-based HAR/HAE system. These actions can be performed without using any tools and weights and they are visually understandable for further recognition and evaluation. Table 1 illustrates some of the physical activities, target disability, and target body parts. The first 6 exercises in the table are targeting upper limb and the next 4 of them are for lower limb impairments. It is worth mentioning that some of these exercises are targeting general impairments, which means that they are most common in any rehabilitation program, regardless of the impairment. This means that creating datasets using these exercises is more helpful for creating a generalized automatic HAR/HAE system. In addition, the datasets which have been created targeting both upper and lower limbs, are considering more different skeletons and muscles which leads to a generalized dataset.

The figures of the actions illustrated in Table 1 are perfectly interpretable for gesture description. Several exercises are for the rehabilitation of the impaired upper limb. As an example, elbow flexion and extension consist of moving the elbow joint, starting from a straight elbow to a bent one. Shoulder flexion is moving the shoulder while keeping the arm straight in front of the body. Shoulder abduction consists of the movement of the arm raised away from the body's side while keeping it straight. To perform shoulder forward elevation the participant needs to clap the hands together and lift the arms up above the head while keeping arms and elbows straight. Shoulder extension is

another exercise starting the arm beside the body and finishing behind the body while keeping the posture straight.

Some exercises are prescribed for the improvement of mobility in the lower limb. For example, the side tap is a way to improve the balance in the body by training the patient to move one leg to the other side of the body, while maintaining the balance. The description and guidance of the physical exercises and their targeted type of impairments have been explored and mentioned in detail in a website [67], which is developed and gathered by a large team of physiotherapists in Sydney, Australia [68].

Rehabilitation period exercises can be conducted either in a medical clinic or hospital with the direct supervision of a healthcare provider, or in a home-based situation, where the patients perform the prescribed actions in the home. There are several factors contributing to the failure of clinic-based programs in providing full or partial recovery for patients. The expensive treatments, lack of young workforce assisting in these programs, transportation problems, the comfortable situation of home-based rehabilitation for some older adults, and occurring pandemics such as COVID-19 hinder many disabled individuals from continuing to attend these programs. In the case of home-based rehabilitation exercises, most of the patients are noncompliant with the prescribed activities due to the lack of activity monitoring and feedback [58]. With the advent of computer vision systems and AI techniques, which leads to automatic rehabilitation period monitoring, the challenges of traditional clinic-based and home-based rehabilitation programs can be overcome. Telerehabilitation with a good strategy for choosing data modality, vision-based sensor, and AI-based techniques can assist the medical sector in monitoring the rehabilitation and progress of patients. In the next section, we discuss the challenges that skeleton data can overcome compared to other vision-based data collection methods, and why this modality is preferable compared to other vision-based modalities.

4. Challenges of data collection











While the interest of researchers in creating vision-based public datasets for patient action recognition and evaluation has skyrocketed in previous years, there are several technical and ethical issues that need to be considered before creating a scenario for the calibrations. These issues might hinder the dataset from being accurate and generalized for further technical research and evaluation on them. In this section, we are going to briefly discuss these challenges and difficulties of vision-based datasets, and how skeleton modality can be a good substitute for all other vision-based modalities by ameliorating some of the limitations. Some of the challenges are general for any data collection related to this domain, some of them are specific to vision-based methods, and some of them are being solved by using the skeleton data as the modality.

Privacy preserving: Even though the vision-based modalities are favorable for their highly informative features and being captured in a non-intrusive manner, they can create issues regarding privacy preservation. Specifically, the RGB and depth images/videos contain confidential face information, which creates reluctance for the individuals to participate in the data collection. This information should be confidentialised using some face-blurring algorithms to avoid the risk of identifying, which adds another step to the preprocessing phase. This issue creates a challenge for dataset availability for data formats such as RGB data. However, using modalities such as skeleton data is highly privacy-preserving because it just contains information like body joint positions which cannot be used for identifying the participants.

Ethical integrity and intellectual property rights: Another important issue that should be considered in any type of data collection is preserving the ethical integrity of the procedure. According to Facca et al. [75], using digital sensing technologies for collecting data on health-related subjects is challenging. These additional challenges compared to other HAR/HAE data collections originated from the fact that the procedure

Table 1

List of exercises prescribed by experts in rehabilitation programs. The figures are extracted from several resources [67,69,70].

Gesture visual description	Gesture Name	Target impairment
	Elbow flexion and extension [71]	General/ Spinal cord injury
	Shoulder Flexion [71]	General/ Spinal cord injury
	Shoulder Abduction [69,71]	General
	Shoulder Forward Elevation [71]	General
	Shoulder Extension [69]	General
	Shoulder internal rotator with theraband [69]	Spinal cord injury
	Standing up and sitting down [72]	General/ Impaired balance for elderly
	Walking on staircase [73]	General/ Impaired balance for elderly
	Deep squat [69,74]	General
	Stepping to targets(side tap) [71]	Impaired balance for elderly/ lower limb incomplete Tetraplegia

includes real-life patients and disabled people, which is a sensitive group. In order to utilize patients in the process, the data collection procedure needs more ethical screening from the different organizations including the hospital. This can also raise problems related to intellectual property rights for different organizations and hospitals. This problem is usually solved by engaging healthy participants and asking them to perform the correct activity and then mimicking the patients performing the same action for the data collection [58]. Although

the collected data is not as realistic as the previous methodology, it is enough for developing different AI methods and evaluating their performance.

Dataset Diversity: This issue should be considered before collecting the data, in order to create generalized data containing participants with different genders, ages, clothes, physical stamina, and ability [58]. The data collection should be performed in multiple episodes or repetitions, for multiple actions, on different days and situations, containing

subjects from various groups which is a challenging task to be performed by a single team of researchers and needs to be performed in several hospitals and institutions in parallel. This can lead to the ethical issues mentioned above, because of the sensitivity of the data. This challenge might be one of the major reasons for the lack of public diverse datasets for physical activity recognition in rehabilitation. Most of the public datasets for this field mentioned in Section 2.4 contain a limited number of repetitions, subjects, and general diversity of the data. However, NTU-RGBD Dataset [46] is one of the most diverse and popular general action recognition datasets. This dataset contains 60 classes of single-person actions (such as drinking water, falling down) and two-person actions (such as hugging, walking towards, or high-five), captured from 40 participants. A large number of participants and actions helps this dataset to contain diverse samples with high numbers. In another study conducted by researchers in Osaka University [76] for gait recognition, gait videos of 10,307 subjects (a balanced number of males and females with various ages, ranging from 2 to 87 years) were collected [77].

Ambiance calibration: The variations in the ambiance of the environment selected for performing action hugely affect the quality and diversity of the data. The actions can be performed in different indoor/outdoor, lighting, and temperature conditions. Most of the sensors are sensitive to these conditions and might perform poorly in some of these situations. According to Shahrourdy et al. [46] a large number of variations can be created by capturing the data in different backgrounds, in order to create ambiance inconsistency and provide a robust system. In the case of some sensors such as Kinect, they are limited to indoor scenes, because of the operational limitations for lighting in this sensor [55] and that should be considered While creating datasets using this sensor.

Dataset variation: According to Miron et al. [71], another essential issue to be considered while collecting data is intra-class and inter-class variations. Each physical activity prescribed for the rehabilitation period can be performed with different variations in speed and participants, which defines intra-class variations. There are also variations between different actions which makes it harder for any HAR system to differentiate the actions. Skeleton data is somehow robust towards any differences in speed of the actions and participant's body scale because frames captured from sensors like Kinect are first converted to a series of feature vectors regardless of orientation, position, and the speed of action [78]. This makes the skeleton data modality favorable for data collection.

Data imbalance: In some data collection scenarios with binary action classification with discrete labels as "correct" or "incorrect", there is a chance that the final real-life dataset is highly imbalanced (means that the distribution of samples from both classes is not equal) [71]. This happens because some patient participants are not able to perform some gestures because of their medical condition or they are unable to perform an action with several repetitions. To solve this problem during the data collection both the correct and incorrect actions (imitating the patient's movements) can be performed by healthy participants. In the case of tackling the imbalanced dataset, methods such as undersampling and oversampling can be used after the data collection, which vary in different ranges [47].

5. Skeleton data acquisition

As mentioned before there are several vision-based data acquisition methods which include RGB, depth map, IR sequences, and skeleton data. Based on many advantages that skeleton data have compared to other vision-based modalities (like robustness for noisy background, privacy preservation, computationally efficient compared to RGB data to be processed, etc.), they are preferred by many of the studies in the scope of physical rehabilitation assessment. With the advent of pose estimation algorithms and accurate and accessible sensors, collecting skeleton data is much easier and more popular nowadays. Generally,

there are two methods for skeleton data acquisition, which include direct use of any sensing hardware, and indirect methods which include pose estimation algorithms for capturing skeleton information from RGB data [79]. However, since our aim is to use accurate sensors for capturing 3-dimensional skeleton data from people, we will discuss the prior method in this subsection.

To capture the skeleton joint data many direct approaches (using a sensor directly to capture the skeleton data) have been used by several researchers. Optical Motion Capture System (OptiTrack MoCap) sensors, such as Vicon, which are marker-based methods have been used by several studies in the scope of rehabilitation [35,69,80]. In this approach, some reflective markers can be attached to several body joints and the patient's movements are tracked by some trackers (cameras). Then, with some processing of the data in the computer, the 3D joint positions are captured [81]. The OptiTrack method for capturing 3D skeleton data is known for its accuracy in capturing the exact position and better processing capability [81]. However, due to the higher cost of acquiring the sensors for capturing data, many researchers use pose estimation algorithms or other cheap skeleton data-capturing sensors.

With the advent of the Microsoft Kinect XBOX 360, the technology of 3D sensing was transformed to a huge extent. This sensing technology was introduced for the purpose of the gaming industry originally. However, it caught the attention of the research community very quickly and it was used in various research fields like gesture recognition [82], pose detection [83], object detection citemanap2015object, sign language recognition [84], virtual reality applications [85], and rehabilitation [56]. The first version of the Microsoft Kinect Sensors (Kinect V1) was meant to be used in the gaming industry and then used by researchers in various research fields as a sensing method. The second version of Kinect (Kinect V2), which was for Windows, had better resolution than the previous one and also has been used for scientific research. Both of these devices have one depth and one RGB camera. In 2019 Azure Kinect sensors were introduced by Microsoft for scientific purposes, mainly for computer vision and speech analysis applications [86]. Among the three sensing technologies for 3D imaging of Time-of-Flight (TOF), Stereo vision, and structured light, the MS Kinect V1 uses structured light technology, in which the device projects some known signal to the object and inspects pattern distortion on the signal received back. This method is suitable for indoor activity monitoring because the pattern distortion is highly sensitive to environmental interference [87]. MS Kinect V2 and Azure Kinect utilize the TOF method of sensing in which the camera sends out IR lights, and records the time or distance it takes for the IR light sent out to return back. The dataset collected by this sensor is limited to indoor scenes, because of the operational limitations of the IR sensor in the sense of light [55]. This method compared to the structured light is more robust to changing lighting conditions [88]. Compared to the first two versions of Kinect, Azure Kinect has several advantages, including better depth resolution, a lighter device (Azure Kinect is a lot smaller than the first two versions), and is more accurate in positioning the skeleton data [86]. The Skeleton joint positions captured by the SDK designed for each of the devices of Azure Kinect, Kinect V2, and Kinect V1, are 32, 25, and 20 joints, respectively. To the best of our knowledge no other previous studies have used the Azure Kinect sensor for capturing physical rehabilitation exercises and due to the reasonable prices of this sensor compared to other accurate methods such as OptiTracks, this sensor should be explored in future studies. Table 2 illustrates several other depth sensors which can capture depth 3D sensors with their related features. Some of the sensors such as Intel RealSense D455 use stereoscopic technology for capturing the depth data by using several cameras distant from each other like the human eyes. These methods need separate SDKs for capturing the skeleton data from the depth data.

Table 2

Several depth sensors and their features such as depth image resolution and frame rate of capturing image.

Depth Sensors	Technology	Depth FOV	frame rate (FPS)	Depth image resolution	Compatibility	Price
Kinect V1	Structured light	H:57°,V:43°	30	320 × 240 px	Windows 7,8 (USB 2.0)	≈ \$150
Kinect V2	TOF	H:70°, V:60°	30	512 × 424px	Windows 8 and higher (USB 3.0)	≈ \$160
Azure Kinect	TOF	NFOV: 65° WFOV: 120°	30	NFOV: 640 × 576 WFOV: 512 × 512	Windows 10 64-bit (USB 3.0)	≈ \$400
Intel RealSense L515	TOF	H:70°, V:55°	30	1024 × 768 px	USB 3.0	≈ \$350
Intel RealSense D455	Stereoscopic	H:87°, V:58°	Up to 90	1280 × 720 px	USB 3.0	≈ \$240
Intel RealSense D435	Stereoscopic	H:87°, V:58°	Up to 90	1280 × 720 px	USB 3.0	≈ \$180
Orbbec Astra	Structured light	H:57°, V:45°	30	640 × 480 px	Windows 7 and higher (USB 2.0)	≈ \$160
Asus Xtion Pro	Structured light	H:58°, V:45°	30	640 × 480px	USB 2.0	≈ \$150

6. Comparative analysis of available datasets

Observing the previous studies on collecting data for rehabilitation exercises, confirms the fact that there are a few publicly available skeleton-based datasets collected for targeting different impairments. For example, in a study conducted by Ar and Akgul [89], the authors used a Microsoft Kinect sensor to capture the RGB and depth videos of the participants performing several rehabilitation exercises for knee and shoulder rehabilitation. However, this study's major limitation is that it does not include joint and skeleton information, which can be useful profoundly in HAR/HAE tasks. The reason behind the lack of publicly available datasets for rehabilitation exercises is the privacy issue for patients and the property rights of organizations [44]. Even the previously described rich datasets like NTU RGB-D [46] are for daily activities and they are not including rehabilitation exercises which are complex activities. Table 3 illustrates some of the characteristics of the datasets which include skeleton data as one of the vision-based modalities collected. These datasets encompass different vision-based modalities to provide sufficient information for evaluating different automatic systems trained on them. However, there are several limitations for the existing datasets (such as using old low-resolution sensors, capturing data in single-view, and targeting specific populations or body limbs) which need to be addressed.

Most of the datasets collected previously are targeting some specific impairments and their therapy activities. One of the famous datasets created and published in 2018 is UI-PRMD (University of Idaho-Physical Rehabilitation Movement Data) [69] which is captured to address the lack of publicly available datasets for therapy movements. One of the strengths of this dataset is that it includes 10 general rehabilitation exercises and is not targeting any specific impairment group. They asked 10 healthy individuals to perform both correct actions and incorrect actions (simulating the patients) for 10 repetitions. This dataset includes the positions and angles of body joints as skeleton data. Although the present paper is exploring skeleton data as a sufficient modality for recognition and evaluation, using multi-modality techniques can improve the performance of any HAR/HAE system. However, the UI-PRMD is an example of studies not providing any further vision-based modalities as a data format.

Another recent dataset collected and published by Miron et al. [71] is utilizing one Kinect V1 sensor to record skeleton data from 29 subjects (15 patients and 14 healthy people) performing 9 general rehabilitation exercises. This dataset provides skeleton data and the depth images captured and not the RGB streams. Other than having limited modalities, this dataset is suitable for HAR tasks since it only provides labels for "correct" and "incorrect" gestures.

The University of Bristol's (Sensor Platform for HEalthcare in Residential Environment) SPHERE-Staircase2014 [73], SPHERE-Walking 2015 [72], SPHERE-SitStand2015 [72] are a series of datasets including the normal and impaired version of each of the walking, walking on the staircase and sitting and standing movements. The actions in these series of datasets have been performed in both normal and abnormal gait (simulating the patients with stroke and Parkinson's disease with the supervision of a physiotherapist) in front of either Kinect or ASUS Xmotion RGB-D camera. Although these datasets are a great source for motion quality evaluation, they are specific to certain targeted actions and the datasets are not generalized.

The KIMORE dataset is another recent dataset, published in 2019 [90], addressing the limited participants problem. In this study, 44 healthy and 34 unhealthy subjects perform 5 repetitions of 5 physical activities for back pain rehabilitation. Kinect V2 was used for action recording and the depth streams and joint positions and joint orientations were extracted using the sensor. The RGB images are also captured, however, they are not publicly available. This dataset could solve some problems related to a limited number of participants and capture different modalities; however, this study includes only a limited series of actions related to a specific target impairment (back pain).

AHA-3D [91] is a dataset captured in 2018 for assessing the lower body fitness in seniors while performing exercises of chair-stand, feet up and go, step test, and unipedal stance. A Kinect V2 and an RGB camera are used to capture the information related to these actions from 11 young and 10 elderly people. Although this dataset has several vision-based modalities which are useful for creating a powerful multi-modal HAR/HAE system, this dataset lacks in the number of action classes and the number of subjects. The TRSP [92] dataset was created to address the lack of an appropriate dataset for detecting compensatory motions during the rehabilitation period of stroke patients. Such a data set is useful in developing an automatic system for coaching stroke survivors in proper positioning. A Kinect V2 was used to capture the skeleton data from four compensatory movements performed by 19 participants. This dataset was also created for a specific purpose and includes limited actions, participants, and modality.

Although some of the limitations of these studies are mentioned specifically for each of them, there are some important general limitations that have not been considered by any of these datasets. For example, these datasets could be captured using sensors with higher accuracy compared to sensors like Kinect V2, such as MS Azure Kinect. Another problem they are facing is that they have used single sensors, without changing their point of view, and position of them, which can highly impact the RGB and depth data collected. The environmental calibrations of the lab used for data capturing have not been considered and the data was collected in constant environmental situations

Table 3
Public datasets for physical rehabilitation exercises.

Dataset (year)	Target group	Participants	Sensors	Physical activity	Objective	Collected modalities	Limitations
IRDS (2021) [71]	General	29 subjects (15 patients, 14 healthy people)	Kinect V1	several repetitions of 9 general rehabilitation exercises (both upper and lower limb)	Classification of correct/incorrect actions regression (predicting a score for actions)	Skeleton data, depth images	Limited number of modality, limited number of subjects, discrete labels only suitable for HAR research, imbalance data
KIMORE (2019) [90]	Back pain, Stroke, Parkinson's	78 subjects (34 patients, 44 healthy)	Kinect V2	5 repetitions of 5 exercises for back pain (both upper and lower limb)	regression (predicting a score for actions)	Skeleton data, depth images, RGB(not public)	Specific target population, specific physical activities, limited number of actions
UI-PRMD (2018) [69]	General	10 healthy subjects (performing both correct and incorrect actions)	Kinect and VICON	10 repetition of 10 general rehabilitation exercises (both upper and lower limb)	regression (predicting a score for actions)	Skeleton data	Limited number of modality, limited number of subjects
SPHERE-Staircase (2014) [73]	Walking-up stairs gait	12 participants performing normal and abnormal gait	ASUS Xmotion RGB-D camera	48 sequences of 1 action including walking up the stairs (lower limb)	Classifica-tion/regression	Depth streams and Skeleton data	Limited to one action, limited number of modalities, limited number of subjects, specific physical activities, limited to certain limb rehab
SPHERE-walking (2015) [72]	Walking gait, simulating stroke and Parkinson's patients	10 participants performing normal and abnormal gait	ASUS Xmotion RGB-D camera	40 sequences of 1 action including walking (lower limb)	Classifica-tion/regression	Depth streams and Skeleton data	Limited to one action, limited number of modalities, limited number of subjects, specific physical activities (limited number of actions), limited to certain limb rehab
SPHERE-SitStand (2015) [72]	Sitting and standing gait	10 participants performing normal and abnormal gait	Kinect V2	109 sequences of 1 action including walking (lower limb)	Classifica-tion/regression	Depth streams and Skeleton data	Limited to one action, Limited number of modality, limited number of subjects, specific physical activities, limited to certain limb rehab
TRSP (2017) [92]	compensatory motion detection in stroke patients	19 subjects (10 healthy , 9 stroke patients)	Kinect V2	4 compensatory movements (upper limb)	Classification	Skeleton data	Limited to one action, Limited number of modality, limited number of subjects, limited to certain limb rehab
AHA-3D (2018) [91]	Assessing senior lower body fitness levels	21 subjects (11 young, 10 elderly individuals)	Kinect V2/ RGB camera	79 sequences of 4 actions (lower limb)	Classifica-tion/regression	Skeleton data, depth, RGB images	limited number of subjects, specific physical activities, limited number of actions, limited to certain limb rehab

(without changing the background or the lighting or the temperature). In addition, some of the actions such as rehabilitation exercises related to neck joint recovery are not considered in the exercise setting. All of these raise the need to create a general dataset that can solve some of the problems mentioned above in the future. The more variations in subjects and camera views and backgrounds, the more accurate will be the evaluation of different techniques developed on the same dataset. The introduction of a new activity recognition dataset for the purpose of monitoring actions during the rehabilitation period will enable the research community to apply different new AI techniques and explore their potential and performance.

7. AI methods for representation (feature) learning and evaluation

After generating a proper dataset, including a balanced number of samples for both correct and incorrect activities, the next important step is the design of the analysis pipeline. The objective of each study plays an important role in designing the pipeline for proposing a methodology. Reviewing the literature shows that different research projects were performed for developing automatic rehabilitation systems and each pursued specific objectives. This diversity in their objective makes the comparison difficult. In addition, the studies in this field

Table 4
Skeleton-based methodologies proposed for automatic physical rehabilitation monitoring task.

Study (year)	Dataset	Targeting population	Sensors	Feature	Modeling strategy	Objective	Contributions	Limitations
Chang et al. (2011) [93]	Non-public (2 young patients performing rehab exercises for 34 days)	Upper limb motor impairment	Kinect	Posture based on the Joint angles	Kinerehab	Using Kinect posture recognition for counting the correct exercises	Confirms that the Kinect-based interventions enhance patients' motivation for rehab, and improve their performance over time	Non-public dataset, Preliminary research, Not an AI-based HAE technique, Not providing continuous score for the patients to monitor their subtle improvements
Chang et al. (2013) [94]	Non-public (2 young patients performing rehab exercises for 34 day)	Severe cerebral palsy with upper limb motor impairment	Kinect	Posture based on the Joint angles	Kinerehab	Using Kinect posture recognition for counting the correct exercises	Confirms that the Kinect-based interventions enhance patients' motivation for rehab, and improve their performance over time	Non-public dataset, Preliminary research, Not an AI-based HAE technique, Not providing continuous score for the patients to monitor their subtle improvements
Lin et al. (2013) [95]	Non-public (2 patients with upper limb disability performing Tai-Chi rehab exercises)	Upper limb impairment	Kinect	Normalized coordinates of joints	Kinerehab	Using skeleton data provided by Kinect and perform statistical analysis on them	Grading the actions using ME of skeleton data and prove the contribution of rehabilitation exercises in recovery	Non-public dataset, Preliminary research, Not an AI-based HAE technique, very limited understanding of the improvement in actions
Exell et al. (2013) [96]	Non-public (3 patients performing in 18 interventions)	Upper limb stroke rehabilitation	Kinect and stimulation glove	Joint angle trajectory	FES method for rehab and screening using the data collected from sensors	comparison of patients performance before and after FES with the reference actions using the plots for the joint angle trajectory changing in time	Illustrated the success of the proposed system for improving the patients' movement and during reach and grasp activities	Non-public dataset, Not an AI-based HAE technique
Su et al. (2014) [97]	Non-public (3 shoulder rehab exercises performed 6 subjects)	Shoulder rehab exercises	Kinect	DTW vector captured from skeleton data	Dynamic Time Warping (DTW) algorithm and fuzzy logic	Using DTW and Fuzzy Neural system provided a performance evaluation technique	Provided a good action scoring technique aligning 80.01% of the time with the experts' scores	Non-public dataset, Not an AI-based HAE technique, Requires domain knowledge to design fuzzy rules for new exercises.
Benettazzo et al. (2015) [98]	Non-public (2 shoulder rehab exercises performed by 10 participants)	Shoulder rehab exercises	Kinect	Joint position distance from the reference action	ANN posture recognition	Providing audio feedback for the actions using AI	Provided good detection of the exercises and evaluate reliably their correctness	Non-public dataset, very basic AI-based HAE/HAR technique
Antunes et al. (2016) [99]	Non-public (Modify action, Weight and balance) and public SPHERE-Walking [72]	Stroke, General rehab	Kinect, Asus Xtion PRO	Normalized and temporarily aligned skeleton data using DTW	Using mathematical approaches like Euclidean distance	Providing visually human interpretable feedback for patients	Provided an interpretable physical action assessment method	Not an AI-based HAE technique

(continued on next page)

are rather new and the field has not been adequately explored, which also highlights the potential for further research in this field.

For example, in one study, Chang et al. [93] have used Kinect sensors to leverage the human pose estimation capabilities of the Kinect

SDK for counting the correct exercises performed by the participants in physical rehabilitation (They call this system Kinerehab). This method was only proposed to do a performance evaluation of two young adults with upper limb impairments performing the exercises without

Table 4 (continued).

Study (year)	Dataset	Targeting population	Sensors	Feature	Modeling strategy	Objective	Contributions	Limitations
Eichler et al. (2018) [100]	Non-public (12 patients with stroke, 10 healthy subjects performing FMA movements)	Stroke, upper limb rehab	2 Kinect V1	FMA related features	Action scoring using SVM and RF	Classifying actions into correct/incorrect based on the FMA feature set	Provided a good precision (100% accuracy for two actions) based on just handcrafted features	Non-public dataset, very basic AI-based HAE technique, utilizing handcrafted features, Utilizing discrete values for action scoring and classifying the samples into correct/incorrect
Li et al. (2018) [101]	UI-PRMD	General rehab	Kinect, VICON	Scaled and mean shifted joint angle trajectory, RMS based soft-labels	Different GAN structures	Modeling and evaluation of actions using GAN	Public dataset, First attempt for modeling and evaluation of actions through GAN	Not providing continuous score for the patients to monitor their subtle improvements
Williams et al. (2019) [102]	UI-PRMD	General rehab	Kinect, VICON	Dimensionality reduced skeleton data with AE	GMM model for scoring the actions	Automatic assessment of physical activities	Public dataset, AI method for scoring the action, proved that dimension reduction methods such as AE perform better compared to other methods (PCA)	Using this technique might miss the information about the correlation of joints, no specific info about the contribution of joints in the action scoring is provided
Liao et al. (2020) [66]	UI-PRMD, KIMORE	General rehab, Back pain, Stroke, Parkinson's	Kinect, VICON	Dimensionality reduced skeleton data with AE	GMM Log-likelihood method for scoring the actions, spatio-temporal method for training an automatic scoring model	Automatic assessment of physical activities using spatio-temporal technique	Public dataset, AI method for scoring the action, good performance 0.02527 MAD on UI-PRMD, 0.03786 MAD on KIMORE	Using spatial-temporal technique might miss the info about the correlation of joints, no specific info about the contribution of joints in the action scoring is provided
Kim et al. (2021) [103]	IRDS	General rehab	Kinect	Heatmaps of the skeleton joints	Pre-trained ResNet	Patient identification through physical activity	Public dataset Good performance for classification (around 98% for a specific gesture)	Lack of HAE technique, for scoring the actions Using heatmap technique might miss the info about the correlation of joints
Chowdhury et al. (2021) [104]	KIMORE	Back pain, Stroke, Parkinson's	Kinect V2	Handcrafted features (angles and distance between joints), raw skeleton data	LSTM model with hand crafted features (LSTM-HF) and LSTM-GCN	Automatic assessment of physical activities using two different techniques	Public dataset, comparing the average in every fold of RMSE in all actions by LSTM-GCN (0.191) and LSTM-HF (0.290) shows LSTM-GCN is better	No specific information about the contribution of joints in the action scoring is provided
Albert et al. (2021) [105]	KIMORE	Back pain, Stroke, Parkinson's	Kinect V2	Normalized joint positions	Utilizing GAN for data augmentation	Classifying the actions into healthy/patient using a classifier on original data and augmented data	Public dataset, illustrated that the model trained on the augmented data have better F-measure over all of the classes compared to original data	Not providing continuous score for the patients to monitor their subtle improvements, no specific information about the contribution of joints in the action scoring is provided

(continued on next page)

Table 4 (continued).

Study (year)	Dataset	Targeting population	Sensors	Feature	Modeling strategy	Objective	Contributions	Limitations
Raihan et al. (2021) [106]	KIMORE	Back pain, Stroke, Parkinson's	Kinect V2	Extracted features using 1D LBP from skeleton data	Genetic algorithm-optimized CNN	Automatic assessment of physical actions	Public dataset, AI method for scoring the action, Good scoring performance (MAD 0.01337 on the validation set for KIMORE)	Using this feature extraction technique might miss the info about the correlation of joints, no specific info about the contribution of joints in the action scoring
Du et al. (2021) [80]	Non-public (2 patients with upper limb disability performing Tai-Chi rehab exercises)	Upper limb impairment	Kinect	Raw skeleton data represented as graph	GCN with self-supervised regularization	Automatic assessment of physical activities	Public dataset, proposed method (with an average of MAE for all exercises = 0.021) better than other previous methods	No specific info about the contribution of joints in the action scoring
Deb et al. (2022) [35]	UI-PRMD, KIMORE	General rehab, Back pain, Stroke, Parkinson's	Kinect, VICON	Raw skeleton data represented as graph	STGCN with self attention model	Automatic assessment of physical activities with variable length, providing explainable feedback	Public dataset, Defining explainability of the model for better feedback using attention model, considering the variable length of the movements, better performance compared to previous methods	Not providing individual scores for each of the joints
Mottaghi et al. (2022) [107]	KIMORE	Back pain, Stroke, Parkinson's	Kinect V2	Features provided by the dataset	Deep Mixture Density NN	Automatic assessment of physical actions	Public dataset, AI method for scoring the action, Good scoring performance	Using CNN-LSTM might miss the info about the correlation of joints, no specific info about the contribution of joints in the action scoring

proposing any AI-based HAE technique. In another research, Chang et al. [94] used the same methodology to evaluate the performance of the two young patients and provided them with feedback. This proposed system provided 3 Degrees of Freedom (DoF) for performing physical rehabilitation exercises in the upper limb, which included 1 DoF for elbows and 2 DoF for shoulders. That is an upgraded version of the previous research with 1 DoF. These studies can be considered preliminary research on the use of Kinect sensors for rehabilitation purposes and do not include major AI-based proposed methodology. In addition, counting the correct exercises utilized in these papers does not provide any continuous score for the patients to know how close they are to getting the action correct. However, one of the major results of these studies is to confirm that Kinect-based interventions enhance patients' motivation for rehabilitation, and improve their performance over time. According to Debnath et al. [56], to have a better scoring function for physical activities, Exell et al. [96] compared the joint angle trajectories. In this study, the authors used Functional Electrical Stimulation (FES) which is used in stroke rehabilitation as a way of assisting patients to improve their body movements. The Kinect sensor and a stimulation glove are utilized for data collection. The comparison of patients' performance before and after FES with the reference actions using the plots for the joint angle trajectory changing in time illustrated the success of the proposed system for improving the patients' movement during reach and grasp activities.

Mean joint angle error can also be used as a way to grade an action, which has been used in another study related to the former

ones mentioned above as conducted by Lin et al. [95]. In this study, the authors asked 2 patients with upper bone impairment to perform a Tai-Chi regimen for upper limb rehabilitation which includes 10 standing and 18 sitting actions illustrated in Fig. 5. This paper includes comprehensive information about the skeleton data normalization and performing an action scoring technique that they have utilized. The actions were graded through a strategy and feedback was provided for the participant, to suggest a repetition on performing the action or not.

In the study conducted by Su et al. [97], the authors utilized the DTW and a fuzzy neural system to perform better scoring of actions and provide interpretable feedback based on the speed and the DTW distance of the actions performed by the participants from standard action. Benettazzo et al. [98] utilized joint position Euclidean distance from the reference action as a feature set for providing audio feedback for performance evaluation. All of these methods were proposed in a way that they mostly aimed to produce feedback based on the skeleton data extracted from the Kinect sensors and their differences from the reference actions. However, one of the most important actions that can be performed is to use AI-based techniques (instead of mathematical difference techniques) to automatically score the actions, which makes the progress of decision-making faster using their pattern recognition ability. To solve this issue many studies changed their perspective to build an AI-base automatic scoring system.

In general, one of the most important phases in building any automatic recognition/evaluation system is to find the best representation of the data, which mainly includes finding and extracting the most

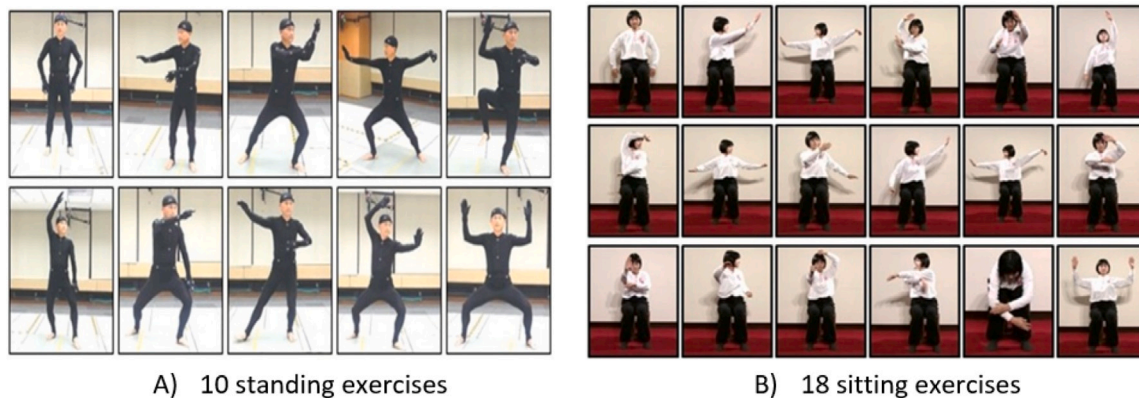


Fig. 5. Tai-Chi rehabilitation exercise regimen including (A) 10 standing exercises and (B) 18 sitting exercises [95]. For more details, please refer to the cited paper.

relevant features. There are two general approaches to performing feature extraction, which is hand-crafted feature extraction, or automatic feature extraction mainly using DL strategies.

The conventional approaches for skeleton-based activity recognition are mostly based on extracting hand-crafted features and then applying some ML methods to them [108,109]. In the area of activity assessment for physical rehabilitation, one of the hand-crafted feature-based methods is introduced by Eichler et al. [100], in which the patients and healthy participants perform the Fugl-Meyer Assessment (FMA) physical activities as a clinically approved intervention for people surviving from a stroke. Two Kinect sensors were utilized for action recording and one medical professional provided FMA scores for the actions. Some features relating to the speed of actions and statistical values (such as mean, max, and variance) of different measurements of angle and distance of the skeleton data were used as the feature set for representing the data. Then, C4.5 (as a decision tree method), Support Vector Machine (SVM), and a Random Forest (RF) classifiers were used to classify samples into patient and healthy (based on the FMA score, where 0–1 is the score for the patient and 2–3 is the score for the healthy participant). In another attempt, Antunes et al. [99] provided a visually human-interpretable feedback system that uses three different datasets to capture the skeleton data, then performs some pre-processing on the data to align them temporarily and spatially using Dynamic Time Warping (DTW), and finally provide feedback based on the Euclidean distance of the joints to the reference action to provide a score.

However, recently there are some proposed methodologies using a Deep Learning approach on the raw collected data for the same purpose. These methods mostly include three main Neural Network architectures, i.e. Recurrent Neural Networks (RNNs) [45,46,110], Convolutional Neural Networks (CNNs) [111–113], and Graph Neural Networks (GNNs) [44,114–116]. For each of these methods, the coordinates of the joints should be represented differently, such as vector sequences, pseudo-images, and graphs, respectively. According to Shi et al. [44], in the field of HAR, sequence-based techniques utilize RNN-based architectures and feed the skeleton data as a sequence of joints (time-series sequences), to capture the temporal features of the data. CNN-based frameworks can capture the spatial features of the skeleton pseudo-image representation of the skeleton data and perform an image classification task. In some studies, instead of representing skeleton data as sequences or pseudo-images, authors used graph-based models in which the skeleton data is represented as a graph. In the graph representation, joints are vertices and bones are edges. According to Shi et al. [44], the reason for the popularity of graph-based techniques for modeling skeleton data is that compared to the sequence-based methods and image-based representation, the graph-based methods are more reasonable since the skeleton in the human body is naturally organized as a graph. There are some kinematic dependencies between skeleton bones and joints, and GNN models by applying special convolutions

on over graph edges corresponding to the joints can capture these dependencies [117,118].

Deep Learning techniques for physical rehabilitation exercise evaluation have been explored recently in a small number of papers. In the study conducted by Williams et al. [102], the authors utilized an autoencoder (AE) for dimensionality reduction and a Gaussian Mixture Model (GMM) to derive a parametric probabilistic movement model of the density of the movements to evaluate the human movements in physical rehabilitation exercises. MSE, MAE, and MPE for two exercises of deep squat and standing shoulder abduction with four approaches of scoring (GMM, DTW, Mahalanobis distance, and Euclidean distance) were presented in this paper. This paper showed that the AE model produces better results compared to other dimensionality reduction methods, such as Principal Component Analysis (PCA).

Also, Liao et al. [66] proposed a pipeline with three important components of dimensionality reduction for skeleton data, the scoring method for the actions, and the spatio-temporal-based methodology for scoring the actions. This paper investigated dimensionality reduction for skeleton data using AEs (including 3D data of 15 to 40 skeleton joints regarding the sensor type) which is rarely investigated by other studies. The authors proposed a Gaussian Mixture Model (GMM) based model for scoring the actions. Finally, a spatio-temporal architecture, including 1D CNNs and Long Short-Term Memory (LSTM) layers, was used to perform the regression. Kim et al. [103] performed a patient identification using a pre-trained ResNet architecture on the heatmaps extracted from the skeleton data of the healthy and patient people in the public IRDS dataset. This method illustrated good performance in classifying the patients. However, it lacks scoring of the actions, which can help the patients to understand to what extent they are performing the actions well. One of the latest research conducted by Mottaghi et al. [107], proposed a pipeline called Deep Mixture Density Network (DMDN) including CNN, and LSTM layers for capturing spatio-temporal features of the motion by adding mixture density layers to predict the scores for the skeleton data in the KIMORE dataset. The metrics of Root Mean Square Error (RMSE) and Spearman correlation coefficient of the validation dataset for each action were provided by the authors and according to the results, the DMDN provides good performance compared to Liao et al. [66] in some of the exercises.

In 2021 Raihan et al. [106] utilized a mixture of both hand-crafted features and Deep Learning methodologies to propose a genetic algorithm-optimized CNN model trained on the 1D LBP (Local Binary Pattern) feature sets extracted from the skeleton data from KIMORE dataset. The resulting Mean Absolute Deviation (MAD) for the testing set in the KIMORE dataset illustrates that the method has a better regression performance compared to the method proposed by Liao et al. [66]. Chowdhury et al. [104] conducted research on comparing the performance of two pipelines including feeding the handcrafted features provided by the KIMORE dataset to an LSTM neural network

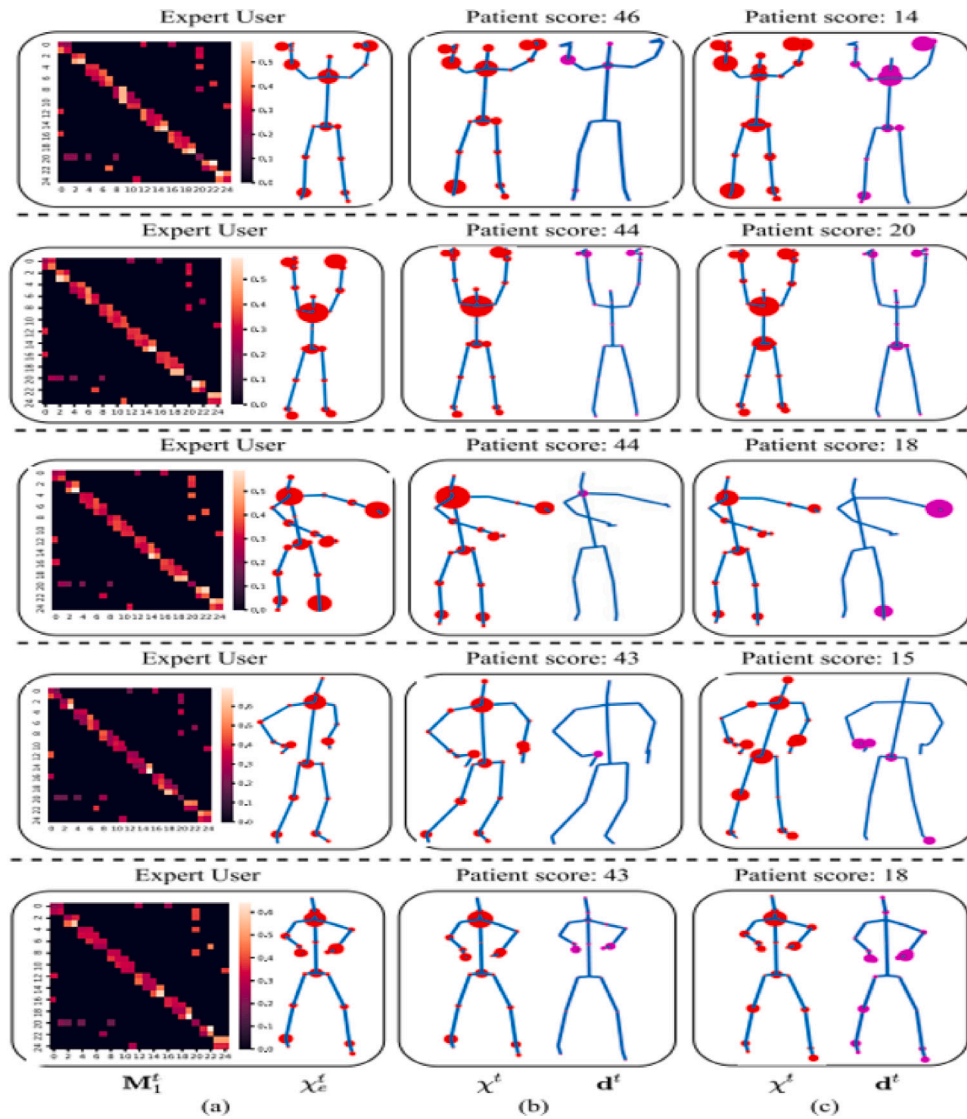


Fig. 6. Attention maps provided in the paper [35] for five exercises. To represent the importance of each joint, the circles are shown bigger when they have higher importance. The figure represents (a) the Average attention map (left) and joint role or importance (right) of expert users. In columns (b) and (c), the left figures illustrate the role (or importance) of different joints in scoring, when the score gets high or low respectively, and the right figures show the difference in the role of joints from the reference movement (where the violet circles are bigger, the patients needs to pay more attention to perform better action).

(LSTM-HF) and feeding the raw skeleton data as with a graph representation to a Graph Convolutional Network (GCN)-LSTM architecture. The RMSE reported as an average of cross-validation of every fold for each exercise illustrates that LSTM-GCN (average RMSE=0.191) performs better compared to LSTM-HF (average RMSE=0.290). The results in this paper prove the fact that the GCN technique can capture better spatio-temporal features of the human body compared to the handcrafted features provided by the experts. In similar research, Du et al. [80] utilized a GCN with a self-supervised regularization on the UI-PRMD dataset to show that the GCN can capture spatial information of the human body. The mean absolute error (MAE) between the predicted score values and the ground truth performance scores on the validation set for the 10 exercises shows that the proposed method (with an average of MAE for all exercises of 0.021) performs better than other methods such as Liao et al. [66] (with an average of MAE for all exercises=0.025).

One of the major limitations of the previous studies is that the HAE systems are not able to provide interpretable and explainable feedback for the patients to know which joints are the most contributing (salient) ones in the decision-making progress of the system. Providing an explainable methodology can help the patients to improve their actions by

paying more attention to the special joint movements resulting in low scores and assist the patients in monitoring their actions and trusting a transparent model instead of a black box. Another important limitation of the previous works is that in order to feed the action performed by a participant to a CNN or LSTM model, they had to convert the captured videos to fixed-length ones which contradict real-world situations since the actions can be performed with different speeds and repetitions. To address both of these problems and create a model with better performance, Deb et al. [35] proposed a Spatio-Temporal GCN (STGCN) with a self-attention layer. This paper provides a comparison of different methods such as [66,116,119–121] with the evaluation criteria such as MAD, Mean Absolute Percentage Error (MAPE) and RMSE scores. Comparing these criteria for all of the 10 exercises in UI-PRMD and five exercises in KIMORE illustrates that the proposed method performs better in scoring for most of the exercises. The attention map illustrating the importance of the joints in scoring each action is given in Fig. 6. To the best of our knowledge, this is the first attempt in providing explainable scores for actions in physical rehabilitation assessment and this direction needs to be explored further.

Due to the fact that it is challenging to create a large dataset in the medical domain (including the scope of this paper), which is essential

for deep learning models to learn the pattern in data, some studies tried to solve this situation with data augmentation methodologies. Albert et al. [105] proposed a Generative Adversarial Network (GAN) with CNN and LSTM layers for producing sufficient synthetically augmented data. They illustrated that a fully convolutional network classifier trained on the augmented data can classify the samples into patient and healthy better than the original data. Li et al. [101] investigated different types of GAN models, such as Deep Convolutional GANs (DCGAN) [122], Wasserstein GAN [123], and Recurrent GAN [109] for both data augmentation and performance evaluation. However, the classification accuracy of the GANs is assessed based on a series of introduced soft labels for the action sequences.

8. Evaluation methods

In this subsection two levels of evaluation criteria selection for the skeleton data analysis is discussed. In the first level, evaluation methods for the human subjects' activities are discussed which plays an important role in the final HAE system performance. The second level encompasses the evaluation techniques proposed by different studies to assess the performance of ML/DL-based HAE systems compared to other pipelines.

8.1. Evaluation methods for human subjects' actions

In this subsection, we discuss performance evaluation approaches for the human subjects' actions utilized in different previous studies. In the level of the participants' action evaluation, which includes the "degree of correctness" of the physical activities performed by the subjects, the actions can be annotated with discrete and continuous scores. In other words, the approach of scoring the actions may frame the problem as either classification or regression [56]. The action evaluation methodology plays an important role in the validation and interpretability of the whole HAE system. Table 5 includes some of the most common methods for scoring the actions which we will discuss in the following.

According to Mangal et al. [64], generally human motion scoring can be explored in two main categories, (1) rule-based and (2) template-based approaches. Rule-based approaches (or clinical scoring) are providing scores for the actions based on a set of rules provided by the clinicians who assess the movement with tools and questionnaires. In other words, some of the previous studies preferred to use the knowledge and experience of the physiotherapists in scoring during the data collection stage. Some of the very basic related methodologies such as counting the correct exercises [93,94] have been proposed previously to evaluate patients' improvement performance by comparing the number of correct exercises before and after performing some physical activities to the correct actions performed by the experts. This method lacks a very important characteristic of an automatic assessment model, which is the interpretability of the scoring methodology. The HAE systems designed based on this scoring method are unable to assist the expert in monitoring the subtle improvements in the performance. FMA [124] and Unified PD Rating Scale (UPDRS) [125] are some of the clinical scoring methodologies utilized by different authors for action assessment [100,126]. As another example, clinicians monitored the actions performed by the healthy and patient participants in the KIMORE dataset [90] through a questionnaire called the Exercise Accuracy Assessment Questionnaire (EAAQ) [127], which is illustrated in Fig. 7. According to this assessment system finally, each action is quantified through three scores of the clinical Total Score (TS) as the sum of all of the ten identified scores; the clinical Primary Outcome (PO) score as the sum of the scores of the first three questions; and the clinical Control Factors (CF) as the sum of the last seven questions. In the data collection related to the IRDS dataset [71], the authors utilized the expert knowledge to provide scores as labels of correct/incorrect to the participants performing the actions.

These methodologies are able to provide powerful and real-world scores because of using the experts' knowledge. However, there are several limitations to this data annotation method. First, in most of the data collection procedures access to different experts from different disciplines (such as both computer science and medical science) is limited. In addition, in some cases, the scoring methodologies that the medical experts use might vary based on different tools and questionnaires. This makes the data more specific to a certain tool and questionnaire results and hinders the researchers from finding a more generalized HAE pipeline for action assessment. It fails in generalizing the model for new physical activities not clinically scored and not introduced to the model before. Moreover, the reliability of the scores provided is highly dependent on the experience, knowledge, and possible bias of the expert scoring the actions. Therefore, we recommend that future researchers in the related area provide an automatic procedure to create generalized annotations. To reach this goal, it is preferable to use a template-based scoring approach in which actions are being assessed compared to a reference perfect action.

The template-based scoring approach can be classified into two groups of model-free (direct matching) and model-based group of metrics [66]. The model-free approach includes applying a distance function between the sequences of actions performed by the participant and the reference action. Utilizing distance functions as scoring criteria assists us in providing a generalized qualification method, which can be used for new types of physical activities. For example, to provide a more generalized and interpretable score for assessing the actions performed by the patients, some studies proposed grading the actions through Mean Absolute Error (distance) (MAE) or MAD [95]. For example, Lin et al. [95] used joint position (after scaling them) and angle mean error as a measuring method for monitoring the progress of patients. They used the distance/error (denoted by d) function illustrated in Eq. (1) to find the distance of 3D joint positions of the reference (R_i) and patients (P_i) movements considering that the Kinect sensor can capture n joints:

$$d = \frac{1}{n} \sum_{i=1}^n |R_i - P_i| \quad (1)$$

Then, they provided a set of discrete scores ranging from 0 to 2, in which 0 means the ME for both of the joint positions and angles was not higher than a threshold, 1 means the ME for either joint position or angle was higher than a threshold, and 2 means that the ME for both of the joint positions and angles was higher than a threshold. Although this methodology improved the understanding of the performance of the patients slightly, since it provides a discrete score, changes in the improvement of the actions are not noticeable. In addition, these scores are not taking into account the whole temporal sequence of the action being performed from the starting point to the end of the action. In general, methods like MAE and Euclidean [95,96,98,128] distance for comparing the two time series are not suitable because they are not considering the variations in the length of the time series vector (length of recording). For this reason, methods like DTW are being used as a distance metric for time series recordings with different lengths [97]. In general, DTW is a method for recovering the optimal temporal alignment of two sequences of time series with different and variable lengths [129]. This method and other versions of it have been used in several papers as a pre-processing phase to align two human actions with different lengths [99]. In specific, according to Zhou and De la Torre [130] given two time series of $X = [x_1, \dots, x_n]$ and $Y = [y_1, \dots, y_m]$, DTW is a technique to align X and Y with different lengths of n and m such that, the following sum of square cost error is minimized. This method can also be used for scoring the actions.

Compared to model-less approaches, the model-based metrics use probabilistic methods for modeling the skeleton motion data and employ the log-likelihood for performance evaluation [66]. According to Mangal et al. [64] this approach is advantageous since it generates a generalized score for any type of action with good accuracy. Hidden

EXERCISE ACCURACY ASSESSMENT QUESTIONNAIRE

Taking into account the description and aim of the exercise and observing the whole exercise (all repetitions), please answer the questions choosing one of the following options:

☐ 1=Never
 ☐ 2=Rarely
 ☐ 3=Sometimes
 ☐ 4=Often
 ☐ 5=Always

- 1) Is the primary objective of the exercise achieved (i.e., extension of the upper limbs, trunk rotation with upper limbs elevated to 90°, squatting, etc.) ?
- 2) Is the exercise repeated in a constant manner?
- 3) Is the amplitude of the movement complete?
- 4) Is the posture of the head correct?
- 5) Is the posture of the right arm correct?
- 6) Is the posture of the left arm correct?
- 7) Is the posture of the trunk correct?
- 8) Is the posture of the pelvis correct?
- 9) Is the posture of the right leg correct?
- 10) Is the posture of the left leg correct?

Fig. 7. The questionnaire used in the KIMORE dataset for scoring participant's action performance [90].

Table 5

List of common action evaluation methods and their limitations and strengths.

Subject performance evaluation method	Some of the papers using the criteria	Pros	Cons
Counting the correct exercises (discrete scoring)	[93,94]	Preliminary methodology	Not monitoring the subtle improvements in the actions
MAE of joint positions	[95]	Providing a better understanding of the action compared to the discrete scores	Not taking the whole performance of the temporal action from the starting to the ending point
MAE of joint angles	[95,96]	Providing a better understanding of the action compared to the discrete scores	Not taking the whole performance of the temporal action from the starting to the ending point/ not suitable for time series data
Euclidean Distance of skeleton data	[98,99,102,128]	Providing a better understanding of the action compared to the discrete scores	Not taking the whole performance of the temporal action from the starting to the ending point/ not suitable for time series data
DTW and its variations	[74,97,102,131]	Suitable for time series data with variable length	Probably not as accurate and generalized as model-based methods
GMM or HMM	[66,102,132]	Model based method of scoring (accurate and generalized)	–

Markov Model (HMM) and GMM are some of the well-known model-based methodologies for scoring the actions based on the probabilistic density functions [66,102,132].

8.2. Evaluation methods for HAE system

The second step in the evaluation process is the evaluation and comparison of the HAE systems based on some standard classification and/or regression metrics. In other words, a very essential step in

conducting research on designing an AI technique for action evaluation is to explore the existing performance evaluation criteria for validating the proposed HAE system. However, according to Lei et al. [11], the evaluation criteria vary in different studies performed in rehabilitation exercise assessment because of its non-uniformity in formulating the data collection. Most of the studies in this scope use their own dataset (which are non-public because of ethical issues and intellectual property restrictions), with different configurations, and evaluation

criteria, which makes it harder to compare different DL/ML-based methodologies applied on them.

As mentioned in the previous section, many papers used MAD, MAP, RMSE, Spearman correlation coefficient, and maybe other methods for evaluating regression models. However, there is no uniformity and coherency in using these criteria to make them comparable with future works. One interesting limitation of previous related work related to the evaluation criteria is not paying attention to the variations in actions cross-subject and cross-view and therefore did not provide a cross-subject and cross-view train-test split and score. For example, in the Shahroudy's et al. [46] study they used one cross-subject evaluation in which they split the data into two sets of train-set and test-set based on the subjects only. In cross-view evaluation, only the data collected by the two front cameras were for training and the data from camera 1 is used for the testing.

In addition to the previously mentioned limitations, it is worth mentioning that all of the studies including designing an HAE system for rehabilitation problem provided the general scores for the actions [35,45,80,105,107]. However, a general score for each action diminishes the explainability of the activity feedback in which the patient will not be able to interpret the score and decide which body part to improve. In an attempt to create interpretable feedback, Deb et al. [35] utilized the attention map to illustrate the problematic body part movements. However, to the best of our knowledge, the use of separate scores for each body part needs to be studied further in the future.

9. Summary of the detected limitations of previous studies

In this section, we briefly discuss the detected challenges in the previous related studies. The studies on developing HAE systems for rehabilitation exercises have the following gaps:

- The previous related public datasets have many limitations such as limited data, single-view data capturing, targeting a specific population, low-resolution capturing devices, and discrete labeling of the activities. This raises the need for new data collection to cover all of these gaps.
- The studies conducted on developing AI-based methods for HAE are very limited and few in number, which shows the potential of this area to be explored further. They have used different datasets for different targets (for activity recognition, or scoring the action based on correct/incorrectness, or scoring actions with a continuous label). Since, providing a continuous label can demonstrate the improvement of the action better, developing a more accurate HAE system for this aim is necessary.
- The accuracy of the scoring system plays an important role in effective treatment. Due to the fact that very limited studies in the literature have been detected, further studies on promoting scoring accuracy should be conducted.
- The related methodologies provide feedback in a way that the patient and expert are provided with either label for actions as correct/incorrect or continuous scores. However, one future study direction can be to use interpretable scores including visual, audible, or tactile tangible feedback. This feedback system can either be used as a reminder (of the incorrect posture or action of the patient) or guidance (of the correct performance of the activity) method for the patients. That can play a key role in a successful rehabilitation procedure.

10. Conclusion

Physical activities have been widely used by physiotherapists as the most adequate prescription for the physical rehabilitation of different disabilities. With the advent and combination of computer vision methods and high-resolution sensors, many studies proposed different

ML/DL-based activity recognition and evaluation assistant systems to help medical experts with decision-making and prescriptions. This paper comprehensively reviews the different stages of designing a system for such a task. Thus, the current review contributes significantly to the literature on automated assessment of physical activity and exercise. First, we discussed about different data-capturing technologies, physical activities to be captured, and the challenges of data collection for physical rehabilitation. Then, we explored the recent ML/DL-based methodologies proposed by different studies for the HAR/HAE task based on the skeleton modality, together with their evaluation methods and the limitations and related gaps.

As mentioned above, the focus of this work is exploring the HAE systems built based on skeleton data for the rehabilitation problem. This decision is made to constrain the research domain in order to make conducting this systematic review feasible. Thus, it is worthwhile to suggest the exploration of different modalities (such as radar, audio, wearable, and Wi-Fi) utilized for the same purpose in future studies to examine their computational cost and accuracy. This will pave the way for future researchers in activity type selection for the specific modality that they are using as the input data. Another future work that we could offer is a comprehensive analysis of designing HAE systems for general applications (including rehabilitation actions, sports, and daily activities) for a better comparison of different techniques' performance (especially DL-based methods).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Coventry University and Deakin University for jointly funding this PhD project titled "Activity Recognition Using Digital Frame Streams for Monitoring Rehab Period".

References

- [1] Australian Demographic Statistics, Canberra: Australian bureau of statistics, Commonwealth Aust. (2017).
- [2] A. Storey, Living longer: how our population is changing and why it matters, 2018, Office for National Statistics: London, UK.
- [3] I.D. Cameron, S.E. Kurrle, 1: Rehabilitation and older people, *Med. J. Aust.* 177 (7) (2002) 387–391.
- [4] A. Reichert, R. Jacobs, The impact of waiting time on patient outcomes: Evidence from early intervention in psychosis services in England, *Health Econ.* 27 (11) (2018) 1772–1787.
- [5] Q. Jiang, Y.H. Kim, J.K. Han, The effect of center-based versus home-based training for rehabilitation of chronic ankle instability in recreational athletes, *J. Men's Health* 18 (1) (2022) 18.
- [6] P. Frigerio, L. Del Monte, A. Sotgiu, C. De Giacomo, A. Vignoli, Parents' satisfaction of tele-rehabilitation for children with neurodevelopmental disabilities during the COVID-19 pandemic, *BMC Primary Care* 23 (1) (2022) 1–10.
- [7] S.C. Cramer, L. Dodakian, V. Le, A. McKenzie, J. See, R. Augsburg, R.J. Zhou, S.M. Raefsky, T. Nguyen, B. Vanderschelden, et al., A feasibility study of expanded home-based telerehabilitation after stroke, *Front. Neurol.* 11 (2021) 611453.
- [8] S. Stephenson, R. Wiles, Advantages and disadvantages of the home setting for therapy: views of patients and therapists, *Br. J. Occupat. Therapy* 63 (2) (2000) 59–64.
- [9] A.Y. Gelaw, B. Janakiraman, B.F. Gebremeskel, H. Ravichandran, Effectiveness of home-based rehabilitation in improving physical function of persons with stroke and other physical disability: A systematic review of randomized controlled trials, *J. Stroke Cerebrovasc. Dis.* 29 (6) (2020) 104800.
- [10] H. Manjunatha, S. Pareek, S.S. Jujavarapu, M. Ghobadi, T. Kesavadas, E.T. Esfahani, Upper limb home-based robotic rehabilitation during COVID-19 outbreak, *Front. Robot. AI* (2021) 75.
- [11] Q. Lei, J.-X. Du, H.-B. Zhang, S. Ye, D.-S. Chen, A survey of vision-based human action evaluation methods, *Sensors* 19 (19) (2019) 4129.
- [12] M. Vrigkas, C. Nikou, I.A. Kakadiaris, A review of human activity recognition methods, *Front. Robot. AI* 2 (2015) 28.

- [13] D. Leightley, J. Darby, B. Li, J.S. McPhee, M.H. Yap, Human activity recognition for physical rehabilitation, in: 2013 IEEE International Conference on Systems, Man, and Cybernetics, IEEE, 2013, pp. 261–266.
- [14] L. Schrader, A. Vargas Toro, S. Konietzny, S. Rüping, B. Schäpers, M. Steinböck, C. Krewer, F. Müller, J. Güttler, T. Bock, Advanced sensing and human activity recognition in early intervention and rehabilitation of elderly people, *J. Popul. Ageing* 13 (2) (2020) 139–165.
- [15] M. Babiker, O.O. Khalifa, K.K. Htike, A. Hassan, M. Zaharadeen, Automated daily human activity recognition for video surveillance using neural network, in: 2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application, ICSIMA, IEEE, 2017, pp. 1–5.
- [16] R.K. Tripathi, A.S. Jalal, S.C. Agrawal, Suspicious human activity recognition: a review, *Artif. Intell. Rev.* 50 (2) (2018) 283–339.
- [17] M. Tammvee, G. Anbarjafari, Human activity recognition-based path planning for autonomous vehicles, *Signal Image Video Process.* 15 (4) (2021) 809–816.
- [18] D.D. Heikoop, M. Hagenzieker, G. Mecacci, S. Calvert, F. Santoni De Sio, B. van Arem, Human behaviour with automated driving systems: a quantitative framework for meaningful human control, *Theor. Issues Ergon. Sci.* 20 (6) (2019) 711–730.
- [19] C. Pham, S. Nguyen-Thai, H. Tran-Quang, S. Tran, H. Vu, T.-H. Tran, T.-L. Le, SensCapsNet: deep neural network for non-obtrusive sensing based human activity recognition, *IEEE Access* 8 (2020) 86934–86946.
- [20] T.M. Do, S.W. Loke, F. Liu, Healthylife: An activity recognition system with smartphone using logic-based stream reasoning, in: International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services, Springer, 2012, pp. 188–199.
- [21] A. Coimbra, C. Neto, D. Ferreira, J. Duarte, D. Oliveira, F. Hak, F. Gonçalves, J. Fonseca, N. Lori, A. Abella, et al., Review of trends in automatic human activity recognition in vehicle based in synthetic data, in: International Conference on Intelligent Data Engineering and Automated Learning, Springer, 2020, pp. 368–376.
- [22] S. Jindal, M. Sachdeva, A.K.S. Kushwaha, A systematic analysis of the human activity recognition systems for video surveillance, in: IoT and Analytics for Sensor Networks, Springer, 2022, pp. 345–354.
- [23] M. Rashmi, R.M.R. Guddeti, Skeleton based human action recognition for smart city application using deep learning, in: 2020 International Conference on COMMunication Systems & NETWORKS, COMSNETS, IEEE, 2020, pp. 756–761.
- [24] A. Khan, S. Khan, B. Hassan, Z. Zheng, CNN-based smoker classification and detection in smart city application, *Sensors* 22 (3) (2022) 892.
- [25] A.B. Abkenar, S.W. Loke, A. Zaslavsky, IoT-enabled group activity recognition services using a modeling language approach, in: 2018 3rd International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), IEEE, 2018, pp. 1–6.
- [26] X. Hu, J. Dai, M. Li, C. Peng, Y. Li, S. Du, Online human action detection and anticipation in videos: A survey, *Neurocomputing* 491 (2022) 395–413.
- [27] P. Ni, S. Lv, X. Zhu, Q. Cao, W. Zhang, A light-weight on-line action detection with hand trajectories for industrial surveillance, *Digit. Commun. Netw.* 7 (1) (2021) 157–166.
- [28] Y. Kong, Y. Fu, Human action recognition and prediction: A survey, *Int. J. Comput. Vis.* 130 (5) (2022) 1366–1401.
- [29] A. Rudenko, L. Palmieri, M. Herman, K.M. Kitani, D.M. Gavrila, K.O. Arras, Human motion trajectory prediction: A survey, *Int. J. Robot. Res.* 39 (8) (2020) 895–935.
- [30] A. Zia, Y. Sharma, V. Bettadapura, E.L. Sarin, M.A. Clements, I. Essa, Automated assessment of surgical skills using frequency analysis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 430–438.
- [31] K. Lam, J. Chen, Z. Wang, F.M. Iqbal, A. Darzi, B. Lo, S. Purkayastha, J.M. Kinross, Machine learning for technical skill assessment in surgery: a systematic review, *NPJ Digital Med.* 5 (1) (2022) 1–16.
- [32] A. Ahmadi, E. Mitchell, C. Richter, F. Destelle, M. Gowing, N.E. O'Connor, K. Moran, Toward automatic activity classification and movement assessment during a sports training session, *IEEE Internet Things J.* 2 (1) (2014) 23–32.
- [33] V. Camomilla, E. Bergamini, S. Fantozzi, G. Vannozzi, Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: A systematic review, *Sensors* 18 (3) (2018) 873.
- [34] P. Parmar, B.T. Morris, Measuring the quality of exercises, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2016, pp. 2241–2244.
- [35] S. Deb, M.F. Islam, S. Rahman, S. Rahman, Graph convolutional networks for assessment of physical rehabilitation exercises, *IEEE Trans. Neural Syst. Rehabil. Eng.* 30 (2022) 410–419.
- [36] S.K. Yadav, K. Tiwari, H.M. Pandey, S.A. Akbar, A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions, *Knowl.-Based Syst.* 223 (2021) 106970.
- [37] Z. Sun, Q. Ke, H. Rahmani, M. Bennamoun, G. Wang, J. Liu, Human action recognition from various data modalities: A review, *IEEE Trans. Pattern Anal. Mach. Intell.* (2022).
- [38] Y. Karayaneva, S. Sharifzadeh, Y. Jing, K. Chetty, B. Tan, Sparse feature extraction for activity detection using low-resolution IR streams, in: 2019 18th IEEE International Conference on Machine Learning and Applications, ICMLA, IEEE, 2019, pp. 1837–1843.
- [39] Y. Wang, S. Cang, H. Yu, A survey on wearable sensor modality centred human activity recognition in health care, *Expert Syst. Appl.* 137 (2019) 167–190.
- [40] X. Li, Y. He, X. Jing, A survey of deep learning-based human activity recognition in radar, *Remote Sens.* 11 (9) (2019) 1068.
- [41] Y. Karayaneva, S. Sharifzadeh, W. Li, Y. Jing, B. Tan, Unsupervised Doppler radar based activity recognition for e-healthcare, *IEEE Access* 9 (2021) 62984–63001.
- [42] H.M. Do, K.C. Welch, W. Sheng, Soham: A sound-based human activity monitoring framework for home service robots, *IEEE Trans. Autom. Sci. Eng.* (2021).
- [43] H. Yan, Y. Zhang, Y. Wang, K. Xu, WiAct: A passive WiFi-based human activity recognition system, *IEEE Sens. J.* 20 (1) (2019) 296–305.
- [44] L. Shi, Y. Zhang, J. Cheng, H. Lu, Skeleton-based action recognition with directed graph neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 7912–7921.
- [45] Y. Du, W. Wang, L. Wang, Hierarchical recurrent neural network for skeleton based action recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1110–1118.
- [46] A. Shahroudy, J. Liu, T.-T. Ng, G. Wang, Ntu rgb+d: A large scale dataset for 3d human activity analysis, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1010–1019.
- [47] S. Sardari, B. Nakisa, M.N. Rastgoo, P. Eklund, Audio based depression detection using Convolutional Autoencoder, *Expert Syst. Appl.* 189 (2022) 116076.
- [48] M.B. Shaikh, D. Chai, RGB-D data-based action recognition: a review, *Sensors* 21 (12) (2021) 4246.
- [49] C. Chen, R. Jafari, N. Kehtarnavaz, Action recognition from depth sequences using depth motion maps-based local binary patterns, in: 2015 IEEE Winter Conference on Applications of Computer Vision, IEEE, 2015, pp. 1092–1099.
- [50] L. Xia, C.-C. Chen, J.K. Aggarwal, View invariant human action recognition using histograms of 3d joints, in: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, IEEE, 2012, pp. 20–27.
- [51] Y. Xie, J. Zhao, B. Qiang, L. Mi, C. Tang, L. Li, Attention mechanism-based CNN-LSTM model for wind turbine fault prediction using SSN ontology annotation, *Wirel. Commun. Mob. Comput.* 2021 (2021).
- [52] L. Mou, C. Zhou, P. Zhao, B. Nakisa, M.N. Rastgoo, R. Jain, W. Gao, Driver stress detection via multimodal fusion using attention-based CNN-LSTM, *Expert Syst. Appl.* 173 (2021) 114693.
- [53] H. Zhou, H. Hu, Human motion tracking for rehabilitation—A survey, *Biomed. Signal Process. Control* 3 (1) (2008) 1–18.
- [54] D. Webster, O. Celik, Systematic review of kinect applications in elderly care and stroke rehabilitation, *J. Neuroeng. Rehabil.* 11 (1) (2014) 1–24.
- [55] A. Da Gama, P. Fallavollita, V. Teichrieb, N. Navab, Motor rehabilitation using Kinect: a systematic review, *Games Health J.* 4 (2) (2015) 123–135.
- [56] B. Debnath, M. O'Brien, M. Yamaguchi, A. Behera, A review of computer vision-based approaches for physical rehabilitation and assessment, *Multimedia Syst.* (2021) 1–31.
- [57] S. Sathyanarayana, R.K. Satzoda, S. Sathyanarayana, S. Thambipillai, Vision-based patient monitoring: a comprehensive review of algorithms and technologies, *J. Ambient Intell. Humaniz. Comput.* 9 (2) (2018) 225–251.
- [58] M.A.R. Ahad, A.D. Antar, O. Shahid, Vision-based action understanding for assistive healthcare: A short review, in: CVPR Workshops, 2019, pp. 1–11.
- [59] A.C. Tricco, E. Lillie, W. Zarin, K.K. O'Brien, H. Colquhoun, D. Levac, D. Moher, M.D. Peters, T. Horsley, L. Weeks, et al., PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation, *Ann. Internal Med.* 169 (7) (2018) 467–473.
- [60] D.R. Walega, Determination of disability, in: Essentials of Pain Medicine, Elsevier, 2018, pp. 99–102.
- [61] S.F. Atashzar, J. Carriere, M. Tavakoli, How can intelligent robots and smart mechatronic modules facilitate remote assessment, assistance, and rehabilitation for isolated adults with neuro-musculoskeletal conditions? *Front. Robot. AI* 8 (2021) 610529.
- [62] K. Carroll, J. Adams, K. de Valle, R. Forbes, R.A. Kennedy, A.J. Kornberg, M. Vandeleur, D. Villano, I.R. Woodcock, E.M. Yiu, et al., Delivering multidisciplinary neuromuscular care for children via telehealth, *Muscle Nerve* (2022).
- [63] G. Fan, X. Zhang, Video-based human motion estimation by part-whole gait manifold learning, in: Machine Learning for Vision-Based Motion Analysis, Springer, 2011, pp. 215–261.
- [64] N.K. Mangal, A.K. Tiwari, A review of the evolution of scientific literature on technology-assisted approaches using RGB-D sensors for musculoskeletal health monitoring, *Comput. Biol. Med.* 132 (2021) 104316.
- [65] M.H. Lee, D.P. Siewiorek, A. Smailagic, A. Bernardino, S.B.i. Badia, Learning to assess the quality of stroke rehabilitation exercises, in: Proceedings of the 24th International Conference on Intelligent User Interfaces, 2019, pp. 218–228.
- [66] Y. Liao, A. Vakanski, M. Xian, A deep learning framework for assessing physical rehabilitation exercises, *IEEE Trans. Neural Syst. Rehabil. Eng.* 28 (2) (2020) 468–477.

- [67] Free and simple exercise prescription, 2022, <https://www.physiotherapyexercises.com/>, (Accessed 12 August 2022).
- [68] J. Glinisky, L. Harvey, C. Sherrington, O. Katalinic, W. www.physiotherapyexercises.com—new exercises and features to help physiotherapists prescribe home exercise programs, *Physiotherapy* 101 (2015) e1381.
- [69] A. Vakanski, H.-p. Jun, D. Paul, R. Baker, A data set of human body movements for physical rehabilitation exercises, *Data* 3 (1) (2018) 2.
- [70] Shoulder surgery exercise guide, 2022, <https://orthoinfo.aaos.org/en/recovery/shoulder-surgery-exercise-guide/>, (Accessed 12 August 2022).
- [71] A. Miron, N. Sadawi, W. Ismail, H. Hussain, C. Grosan, IntelliRehabDS (IRDS)—A dataset of physical rehabilitation movements, *Data* 6 (5) (2021) 46.
- [72] L. Tao, A. Paiement, D. Damen, M. Mirmehdi, S. Hannuna, M. Camplani, T. Burghardt, I. Craddock, A comparative study of pose representation and dynamics modelling for online motion quality assessment, *Comput. Vis. Image Underst.* 148 (2016) 136–152.
- [73] A. Paiement, L. Tao, S. Hannuna, M. Camplani, D. Damen, M. Mirmehdi, Online quality assessment of human movement from skeleton data, in: *British Machine Vision Conference, BMVA press*, 2014, pp. 153–166.
- [74] M. Capecci, M.G. Ceravolo, F. Ferracuti, S. Iarlori, V. Kyrki, A. Monteriu, L. Romeo, F. Verdini, A hidden semi-Markov model based approach for rehabilitation exercise assessment, *J. Biomed. Inform.* 78 (2018) 1–11.
- [75] D. Facca, M.J. Smith, J. Shelley, D. Lizotte, L. Donelle, Exploring the ethical issues in research using digital data collection strategies with minors: A scoping review, *PLoS One* 15 (8) (2020) e0237875.
- [76] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, Y. Yagi, Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition, *IPSN Trans. Comput. Vis. Appl.* 10 (1) (2018) 1–14.
- [77] C. Xu, Y. Makihara, R. Liao, H. Niitsuma, X. Li, Y. Yagi, J. Lu, Real-time gait-based age estimation and gender classification from a single image, in: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 3460–3470.
- [78] S. Nirjon, C. Greenwood, C. Torres, S. Zhou, J.A. Stankovic, H.J. Yoon, H.-K. Ra, C. Basaran, T. Park, S.H. Son, Kintense: A robust, accurate, real-time and evolving system for detecting aggressive actions from streaming 3D skeleton data, in: *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, IEEE, 2014, pp. 2–10.
- [79] W. Wang, Y.-D. Zhang, A short survey on deep learning for skeleton-based action recognition, in: *Proceedings of the 14th IEEE/ACM International Conference on Utility and Cloud Computing Companion*, 2021, pp. 1–6.
- [80] C. Du, S. Graham, C. Depp, T. Nguyen, Assessing physical rehabilitation exercises using graph convolutional network with self-supervised regularization, in: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE*, 2021, pp. 281–285.
- [81] C.-Y. Chang, B. Lange, M. Zhang, S. Koenig, P. Requejo, N. Somboon, A.A. Sawchuk, A.A. Rizzo, Towards pervasive physical rehabilitation using microsoft kinect, in: *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, IEEE, 2012, pp. 159–162.
- [82] N.M.F. Franslin, G.W. Ng, Vision-based dynamic hand gesture recognition techniques and applications: A review, in: *Proceedings of the 8th International Conference on Computational Science and Technology*, Springer, 2022, pp. 125–138.
- [83] J. Spooner, V. Palade, M. Cheah, S. Kanarachos, A. Daneshkhah, Generation of pedestrian crossing scenarios using ped-cross generative adversarial network, *Appl. Sci.* 11 (2) (2021) 471.
- [84] K.J. Swaroop, J. Channegowda, S. Mishra, A complete overview of sign language recognition and translation systems, *Bioinform. Med. Appl. Big Data using Deep Learn. Algorithms* (2022) 289–313.
- [85] Y. Wang, F. Chang, Y. Wu, Z. Hu, L. Li, P. Li, P. Lang, S. Yao, Multi-Kinects fusion for full-body tracking in virtual reality-aided assembly simulation, *Int. J. Distrib. Sens. Netw.* 18 (5) (2022) 15501329221097591.
- [86] M. Tölgyessy, M. Dekan, L. Chovanec, P. Hubinský, Evaluation of the azure Kinect and its comparison to Kinect V1 and Kinect V2, *Sensors* 21 (2) (2021) 413.
- [87] H. Sarbolandi, D. Lefloch, A. Kolb, Kinect range sensing: Structured-light versus Time-of-Flight Kinect, *Comput. Vis. Image Underst.* 139 (2015) 1–20.
- [88] C.S. Bamji, S. Mehta, B. Thompson, T. Elkhatab, S. Wurster, O. Akkaya, A. Payne, J. Godbaz, M. Fenton, V. Rajasekaran, et al., IMpixel 65nm BSI 320MHz demodulated TOF image sensor with 3μm global shutter pixels and analog binning, in: *2018 IEEE International Solid-State Circuits Conference*, IEEE, 2018, pp. 94–96.
- [89] I. Ar, Y.S. Akgul, A computerized recognition system for the home-based physiotherapy exercises using an RGBD camera, *IEEE Trans. Neural Syst. Rehabil. Eng.* 22 (6) (2014) 1160–1171.
- [90] M. Capecci, M.G. Ceravolo, F. Ferracuti, S. Iarlori, A. Monteriu, L. Romeo, F. Verdini, The KIMORE dataset: Kinematic assessment of movement and clinical scores for remote monitoring of physical rehabilitation, *IEEE Trans. Neural Syst. Rehabil. Eng.* 27 (7) (2019) 1436–1448.
- [91] J. Antunes, A. Bernardino, A. Smailagic, D.P. Siewiorek, AHA-3D: A labelled dataset for senior fitness exercise recognition and segmentation from 3D skeletal data, in: *BMVC*, 2018, p. 332.
- [92] E. Dolatabadi, Y.X. Zhi, B. Ye, M. Coahran, G. Lupinacci, A. Mihailidis, R. Wang, B. Taati, The toronto rehab stroke pose dataset to detect compensation during stroke rehabilitation therapy, in: *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 2017, pp. 375–381.
- [93] Y.-J. Chang, S.-F. Chen, J.-D. Huang, A kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities, *Res. Dev. Disabil.* 32 (6) (2011) 2566–2570.
- [94] Y.-J. Chang, W.-Y. Han, Y.-C. Tsai, A kinect-based upper limb rehabilitation system to assist people with cerebral palsy, *Res. Dev. Disabil.* 34 (11) (2013) 3654–3659.
- [95] T.-Y. Lin, C.-H. Hsieh, J.-D. Lee, A kinect-based system for physical rehabilitation: Utilizing tai chi exercises to improve movement disorders in patients with balance ability, in: *2013 7th Asia Modelling Symposium, IEEE*, 2013, pp. 149–153.
- [96] T. Exell, C. Freeman, K. Meadmore, M. Kutlu, E. Rogers, A.-M. Hughes, E. Halleswell, J. Burridge, Goal orientated stroke rehabilitation utilising electrical stimulation, iterative learning and microsoft kinect, in: *2013 IEEE 13th International Conference on Rehabilitation Robotics, ICORR, IEEE*, 2013, pp. 1–6.
- [97] C.-J. Su, C.-Y. Chiang, J.-Y. Huang, Kinect-enabled home-based rehabilitation system using Dynamic Time Warping and fuzzy logic, *Appl. Soft Comput.* 22 (2014) 652–666.
- [98] F. Benettazzo, S. Iarlori, F. Ferracuti, A. Giantomassi, D. Ortenzi, A. Freddi, A. Monteriu, S. Innocenzi, M. Capecci, M.G. Ceravolo, et al., Low cost RGB-D vision based system to support motor disabilities rehabilitation at home, in: *Ambient Assisted Living, Springer*, 2015, pp. 449–461.
- [99] M. Antunes, R. Baptista, G. Demisse, D. Aouada, B. Ottersten, Visual and human-interpretable feedback for assisting physical activity, in: *European Conference on Computer Vision, Springer*, 2016, pp. 115–129.
- [100] N. Eichler, H. Hel-Or, I. Shmishoni, D. Itah, B. Gross, S. Raz, Non-invasive motion analysis for stroke rehabilitation using off the shelf 3d sensors, in: *2018 International Joint Conference on Neural Networks, IJCNN, IEEE*, 2018, pp. 1–8.
- [101] L. Li, A. Vakanski, Generative adversarial networks for generation and classification of physical rehabilitation movement episodes, *Int. J. Mach. Learn. Comput.* 8 (5) (2018) 428.
- [102] C. Williams, A. Vakanski, S. Lee, D. Paul, Assessment of physical rehabilitation movements through dimensionality reduction and statistical modeling, *Med. Eng. Phys.* 74 (2019) 13–22.
- [103] J.-K. Kim, K.B. Lee, J.-C. Kim, S.G. Hong, Patient identification based on physical rehabilitation movements using skeleton data, in: *2021 International Conference on Information and Communication Technology Convergence, ICTC, IEEE*, 2021, pp. 1572–1574.
- [104] S.H. Chowdhury, M. Al Amin, A.M. Rahman, M.A. Amin, A.A. Ali, Assessment of rehabilitation exercises from depth sensor data, in: *2021 24th International Conference on Computer and Information Technology, ICCIT, IEEE*, 2021, pp. 1–7.
- [105] J. Albert, P. Glöckner, B. Pfitzner, B. Arnrich, Data augmentation of kinematic time-series from rehabilitation exercises using GANs, in: *2021 IEEE International Conference on Omni-Layer Intelligent Systems, COINS, IEEE*, 2021, pp. 1–6.
- [106] M.J. Raihan, M.A.R. Ahad, A.-A. Nahid, Automated rehabilitation exercise assessment by genetic algorithm-optimized CNN, in: *2021 Joint 10th International Conference on Informatics, Electronics & Vision (ICIEV) and 2021 5th International Conference on Imaging, Vision & Pattern Recognition (IcIVPR)*, IEEE, 2021, pp. 1–6.
- [107] E. Mottaghi, M.-R. Akbarzadeh-T, Automatic evaluation of motor rehabilitation exercises based on deep mixture density neural networks, *J. Biomed. Inform.* 130 (2022) 104077.
- [108] B. Fernando, E. Gavves, J.M. Oramas, A. Ghodrati, T. Tuytelaars, Modeling video evolution for action recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5378–5387.
- [109] C. Esteban, S.L. Hyland, G. Rätsch, Real-valued (medical) time series generation with recurrent conditional gans, 2017, *arXiv preprint arXiv:1706.02633*.
- [110] S. Li, W. Li, C. Cook, C. Zhu, Y. Gao, Independently recurrent neural network (indrn): Building a longer and deeper rnn, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5457–5466.
- [111] M. Liu, H. Liu, C. Chen, Enhanced skeleton visualization for view invariant human action recognition, *Pattern Recognit.* 68 (2017) 346–362.
- [112] C. Li, Q. Zhong, D. Xie, S. Pu, Skeleton-based action recognition with convolutional neural networks, in: *2017 IEEE International Conference on Multimedia & Expo Workshops, ICMEW, IEEE*, 2017, pp. 597–600.
- [113] B. Li, Y. Dai, X. Cheng, H. Chen, Y. Lin, M. He, Skeleton based action recognition using translation-scale invariant image mapping and multi-scale deep CNN, in: *2017 IEEE International Conference on Multimedia & Expo Workshops, ICMEW, IEEE*, 2017, pp. 601–604.
- [114] Z. Huang, X. Shen, X. Tian, H. Li, J. Huang, X.-S. Hua, Spatio-temporal inception graph convolutional networks for skeleton-based action recognition, in: *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2122–2130.

- [115] B. Andayeshgar, F. Abdali-Mohammadi, M. Sepahvand, A. Daneshkhah, A. Almasi, N. Salari, Developing graph convolutional networks and mutual information for arrhythmic diagnosis based on multichannel ECG signals, *Int. J. Environ. Res. Public Health* 19 (17) (2022) 10707.
- [116] S. Yan, Y. Xiong, D. Lin, Spatial temporal graph convolutional networks for skeleton-based action recognition, in: *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [117] H. Yang, D. Yan, L. Zhang, Y. Sun, D. Li, S.J. Maybank, Feedback graph convolutional network for skeleton-based action recognition, *IEEE Trans. Image Process.* 31 (2021) 164–175.
- [118] J. Li, X. Xie, Z. Zhao, Y. Cao, Q. Pan, G. Shi, Temporal graph modeling for skeleton-based action recognition, 2020, *arXiv preprint arXiv:2012.08804*.
- [119] Y.-F. Song, Z. Zhang, C. Shan, L. Wang, Richly activated graph convolutional network for robust skeleton-based action recognition, *IEEE Trans. Circuits Syst. Video Technol.* 31 (5) (2020) 1915–1925.
- [120] P. Zhang, C. Lan, W. Zeng, J. Xing, J. Xue, N. Zheng, Semantics-guided neural networks for efficient skeleton-based human action recognition, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1112–1121.
- [121] C. Li, Q. Zhong, D. Xie, S. Pu, Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation, 2018, *arXiv preprint arXiv:1804.06055*.
- [122] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, 2015, *arXiv preprint arXiv:1511.06434*.
- [123] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, in: *International Conference on Machine Learning*, PMLR, 2017, pp. 214–223.
- [124] D.J. Gladstone, C.J. Danells, S.E. Black, The fugl-meyer assessment of motor recovery after stroke: a critical review of its measurement properties, *Neurorehabilitation Neural Repair* 16 (3) (2002) 232–240.
- [125] M.D.S.T.F. on Rating Scales for Parkinson's Disease, The unified Parkinson's disease rating scale (UPDRS): status and recommendations, *Movement Disorders* 18 (7) (2003) 738–750.
- [126] M. Dyschel, D. Arkadir, H. Bergman, D. Weinshall, Quantifying levodopa-induced dyskinesia using depth camera, in: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 119–126.
- [127] M. Capecchi, M.G. Ceravolo, F. Ferracuti, M. Grugnetti, S. Iarlori, S. Longhi, L. Romeo, F. Verdini, An instrumental approach for monitoring physical exercises in a visual markerless scenario: A proof of concept, *J. Biomech.* 69 (2018) 70–80.
- [128] R. Baptista, E. Ghorbel, F. Moissenet, D. Aouada, A. Douchet, M. André, J. Pager, S. Bouilland, et al., Home self-training: Visual feedback for assisting physical activity for stroke survivors, *Comput. Methods Programs Biomed.* 176 (2019) 111–120.
- [129] M. Dvornik, I. Hadji, K.G. Derpanis, A. Garg, A. Jepsen, Drop-dtw: Aligning common signal between sequences while dropping outliers, *Adv. Neural Inf. Process. Syst.* 34 (2021) 13782–13793.
- [130] F. Zhou, F. De la Torre, Generalized time warping for multi-modal alignment of human motion, in: *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2012, pp. 1282–1289.
- [131] S. Schetz-Sobrino, D.N. Monekosso, P. Remagnino, D. Vallejo, C. Glez-Morcillo, Automatic recognition of physical exercises performed by stroke survivors to improve remote rehabilitation, in: *2019 International Conference on Multimedia Analysis and Pattern Recognition, MAPR*, IEEE, 2019, pp. 1–6.
- [132] G.A. Ten Holt, M.J. Reinders, E.A. Hendriks, Multi-dimensional dynamic time warping for gesture recognition, in: *Thirteenth Annual Conference of the Advanced School for Computing and Imaging*, vol. 300, 2007, p. 1.