A review of computer vision based approaches for physical rehabilitation and assessment

Bappaditya Debnath · **Mary O'brien** ·
Motonori Yamaguchi · **Ardhendu Behera**

Abstract Computer vision community has extensively researched the area of human motion analysis, which primarily focuses on pose estimation, activity recognition, pose or gesture recognition and so on. However, for many applications like monitoring of functional rehabilitation of patients with musculoskeletal or physical impairments, the requirement is to comparatively evaluate human motion. In this survey, we capture important literature on vision-based monitoring and physical rehabilitation that focuses on comparative evaluation of human motion during past two decades and discuss state of current research in this area. Unlike other reviews in this area, which are written from clinical objective, this article presents research in this area from computer vision application perspective. We propose our own taxonomy of computer vision-based rehabilitation and assessment researches which are further divided into sub-categories to capture novelties of each research. The review discusses the challenges of this domain due to wide ranging human motion abnormalities and difficulty in automatically assessing those abnormalities. Finally, suggestions on future direction of research are offered.

B. Debnath
Dept. of Computer Science
Edge Hill University, Ormskirk, UK
Tel.: +44-1695-1872
E-mail: debnathb@edgehill.ac.uk

1 Introduction

Computer Vision (CV)-based human motion modelling and analysis has been extensively researched by the community. But, most of the research can be categorised into pose estimation [160], human-object interaction [63, 98], activity/gesture recognition [31, 65, 113] or human-human interaction [53]. However, comparative analysis of human motion has received relatively less attention from the community. Comparative analysis of human motion is necessary for application areas like automated rehabilitation and/or assessment of stroke, Spinal Cord Injury (SCI), Parkinson's Disease (PD) or patients with other physical impairments. Patient's recovering from such impairments undergo extensive physical rehabilitation and are assessed by Clinicians (Physicians, Physiotherapists or Occupational Therapists) that require patients to spend time with their carer(s). **The process is both expensive, labour-intensive, time-consuming and subjected to human error. Statistics shows that informal care for rehabilitation is the reason behind 27% of the whole treatment cost. In case of stroke patients, this amounted to around 2.42 billion pounds a year in UK in 2016 [130]. Moreover, such assessments may suffer from inaccuracies as visual progress reporting scheme is prone to inconsistent perception. Inaccuracies may also arise from the subjectivity of these behavioral and clinical assessments [96]. In addition, integration of assessment based on kinematic parameters can be more robust and accurate as compared to visual assessment by clinicians alone [17].** Body-worn sensors or marker-based systems are expensive and can be very intrusive to a patient's day to day activities. Marker-less vision-based human motion modelling and subsequent comparison has the potential to provide home-based, inexpensive and unobtrusive monitoring. It also has potential applications in sports including, but not limited to diving and figure skating.

1.1 Scope of this review

This review includes relevant articles from last 20 years that is representative of research in the domain of vision-based physical rehabilitation and assessment. We have focused on articles where the data captured using CV methods has been used for comparative analysis i.e., where intelligent processing is involved. The article also included articles on virtual rehabilitation and serious games involving vision-based sensor. In virtual rehabilitation although the role of CV is largely limited to tracking, we have focused on articles having a secondary 'learning' objective. Activity recognition methods specific to rehabilitation exercises have been also included. Research not set in a clinical scenario but aimed towards assessment of physical impairments have been also covered. Existing research suggest that accurate body joint position estimation is vital for vision-based rehabilitation and assessment. However, human pose estimation has been extensively researched and covered in several surveys and

reviews [118, 160]. Similarly, human activity recognition also, has been widely explored by the CV community and covered in several surveys [65, 113]. Thus, this review does not aim to cover joint position estimation or human activity recognition methods. Also, it does not include inertial or other non-vision sensor-based research.

2 Domain characteristics
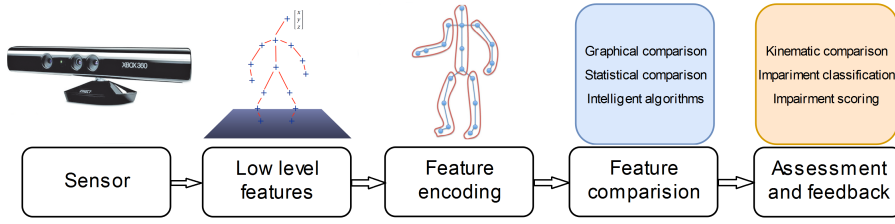
Fig. 1 A very high level illustration of general logical flow for a vision based physically impaired patient assessment system

There are many aspects to a vision-based research including but not limited to raw data, feature extraction, feature representation, feature comparison, statistical and stochastic modelling (DL). However, the general flow of a research in the domain of vision-based rehabilitation and monitoring can be broadly illustrated by the Fig. 1. The illustration highlights important characteristics of this domain. It includes, a vision-based sensor such as monocular RGB or depth camera for sensing the data. A low-level feature such as human joint positions. A feature encoding and representation method such as group of joint positions or combination of human kinematic parameters. Then, the encoded features are compared through simple graphical and statistical techniques or through intelligent algorithms. Finally, assessment is done in the form of kinematic parameter comparisons, pose recognition, automated clinical scoring, impairment classification and others. Rehabilitation systems usually have an exercise program and provide feedback. **These characteristics can be broadly described in three major parts: primary data, feature extraction and representation, and feature comparison. For application of CV to rehabilitation and assessment of physically impaired persons, we focus on the above mentioned aspects. The domain characteristics w.r.t these aspects are discussed next.**

2.1 Physical impairment data

In many other vision-based human motion modelling applications including, but not limited to human pose estimation and activity recognition, large-scale datasets are publicly available. Thus, collecting data is often outside the scope of research. But, for research in physical impairment domain, authors have often collected their own data. Human movements are multidimensional and so are its abnormalities. Musculo-skeletal impairments are exhibited differently in different patients over a period of time. Multitude of factors such as impairment involved, extent of injury, area affected, physiological characteristics and care provided lead to hugely varying manifestation of impairments across patients. This is in addition to the wide range of motion capabilities of human beings. Clinicians have specific tests and exercises designed for rehabilitation and assessment of different types of motor abnormalities. Therefore, researchers are also required to run specific experiment to capture data for the assessment of specific musculo-skeletal impairments. Thus, most authors have captured data catering to specific situations corresponding to their objective. Due to difficulty in accessing patients, ethical issues and other such issues, data is difficult to acquire and the datasets are often small. Researchers have used alternative strategies such as normal persons acting like patients, use of noise to create varied data and others. In this domain, there are a very few publicly available datasets (table 7) and even these are very small when compared to datasets available for other CV applications areas (e.g., image recognition, human activity recognition). In this article, we highlight the target abnormality, area of body affected and the corresponding data collected for each article reviewed.

2.2 Feature extraction and representation

The ultimate goal of musculo-skeletal patient monitoring is to provide an automated assessment. To assess the progress in physical rehabilitation in case of physically impaired persons, clinicians rely on physical characteristics such as extent of elbow flexion, shoulder abduction, speed of motion and others. To determine such characteristics researchers have almost exclusively relied on estimation of human body joint positions as primary low-level features. For automated assessment, it is often required to compare a patient's execution of Activities of Daily Living (ADL) or rehabilitation exercises with a regular healthy execution. Here also, the objective is to compare sequence of joint positions. For normal activity recognition, researchers use image-based features including VLAD [58], Bag of Visual Words (BoVW) [109], Dense trajectories [148] and others. These features incorporate valuable information such as context, optical flow and so on, which are not available when only joint positions are used as low-level features. So far, researchers have mostly used Kinect [161] for obtaining 3D joint positions which has its limitations as explained in [152]. Deep Convolutional Neural Networks (DCNN) have been very

successful in 2D human pose estimation [20, 50] and more recently, these networks are used for 3D pose estimation with much higher accuracy [107, 157]. But research in this domain is yet to fully explore the DCNN-based pose estimation.

Various kinematic features such as joint angle trajectory, relative joint position, speed and acceleration are used for establishing clinical condition of patients. Thus, joint position estimations have been encoded in various different forms for feature representation. Such encoding often comprises of simple human body kinematic features such as relative angles, velocities, body centric coordinates and others [127, 128]. This is useful when a specific type of impairment is in consideration. For example, for discriminating pathological gait, knee angle, step distance and other such parameters are considered. Another approach is to quantify the difference between patient activity and a perfect template consisting of regular healthy activity. For this, researches have used statistical representations such as Hidden Markov Model (HMM) [138] or Dynamic Time Warping (DTW) [11, 121]. The main aim of feature representation is to select and encode joint positions in a manner that improves the discriminatory power of comparative algorithms with regards to the given clinical condition.

2.3 Feature Comparison

One of the major goals of research in vision-based physical rehabilitation and monitoring is to provide an automated clinical assessment of a musculo-skeletal patient's physical condition. For many CV applications such as object detection or activity recognition the objective is well-defined (e.g., classification). However, for assessment of patients the goal varies widely and often depends on the clinical requirements. The requirements vary from, statistical analysis to methods for automatically establishing clinical scores such as Fugl-Meyer Assessment (FMA) [59], Unified PD Rating Scale (UPDRS) [103] and others. For some cases simple presentation and comparison of joint angle trajectories is enough, but for other cases such as automated clinical scoring, advanced comparison algorithms are often required. It needs to be emphasized that researchers have mostly relied only on joint positions as low-level features. Thus, rich vision-based feature representations (e.g., BoVW, MBH) that can provide contextual information are not available. Therefore, it is essential to develop techniques for comparative analysis of features based only on joint positions. Such comparisons can be done in many ways including, but not limited to simple graphical analysis, statistical analysis, sequence comparison, classification and regression. Methods such as graphical comparison are often simple and may not require large datasets. On the other hand, establishing automated clinical scoring requires advanced algorithms and large datasets to work reliably. As explained earlier, obtaining large-scale dataset for each type of abnormal motion is difficult. Therefore, the main challenge in this area is to maximize the applicability of advanced algorithms with limited data.

3 Surveys and Taxonomies

3.1 Surveys

Author	References	Journal	Comments
Zhou et al. 2004 [164]	184	Biomedical Signal Processing and Control	Highlights tracking methods
Webster et al. 2014 [152]	96	Journal of Neuroengineering and Rehabilitation	Focused on Kinect-based research, elderly care and stroke rehabilitation
Hondori et al. 2014 [96]	105	Journal of Medical Engineering	Focuses on Kinect-based research and highlights impact of Kinect
Da Gama et al. 2015 [40]	66	Games for health journal	Focuses on Kinect-based research
Sathyanarayana et al. 2018 [119]	192	Journal of Ambient Intelligence and Humanized Computing	Patient monitoring and algorithms
Ahad et al. 2019 [2]	79	CVPR workshop	Action understanding for assistive healthcare

Table 1: Existing reviews and surveys on vision-based physical rehabilitation and assessment research

Table 1 lists surveys and reviews aimed towards vision-based physical rehabilitation and assessment. Zhou et al. [164] surveyed human motion tracking for rehabilitation. It focuses mainly on various vision and sensor based tracking systems. It further discusses home-based and robot-aided rehabilitation systems. The article doesn't describe algorithms used for comparative evaluation or abnormal activity detection.

Webster and Celik [152] reviewed Kinect-based research and focused on formulation of rehabilitation exercises for monitoring. The authors discuss elderly care and stroke rehabilitation methods. Within elderly care, fall detection, fall risk reduction and Kinect-based gaming are discussed. Articles under stroke rehabilitation is categorised into evaluation of Kinect, rehabilitation methods and Kinect gaming. Similarly, Da Gama et al. [40] also reviewed Kinect based researches. The focus of this review is on formulation of rehabilitation experiments, subsequent monitoring of progress and analysis of various comparison techniques. Most of these techniques rely on basic methods including average angle flexion, Euclidean distance, mean error, correlation coefficient and others. The authors present taxonomy in terms of 'Evaluative', 'Applicability', 'Validation' and 'Improvement' category. The taxonomy is based on clinical perspective. Both [152] and [40] review articles from a clinical perspective where clinical progress made by patients is a major focus. Sathyanarayana et al. [119] reviewed articles from CV perspective and highlighted vision algorithms. Their taxonomy is based on clinical application and articles include areas such as ADL recognition or fall detection which does not always include abnormal or impaired physical motion. Moreover, the review does not include articles after 2014.

In the current study, existing research has been reviewed from CV application perspective. We highlight the musculo-skeletal impairment, visual sensor, feature extraction and comparison algorithms for each reviewed article. The discussion focuses on algorithms used for discriminating and assessing physically impaired activity in comparison to regular healthy activity.

3.2 Taxonomy

In this article we develop our own taxonomy, which is necessitated due to the lack of reviews in this area form a CV application perspective. The review both categorises and tabulates the articles for highlighting different aspects. **As discussed in Sec. 2, it focuses on the following three characteristics: 1) data collection, 2) feature extraction and representation, and 3) feature comparison**. Thus, the articles reviewed are tabulated to address these aspects. The columns headed **Target** and **Dataset** highlight the kind of impairment, area of body affected and briefly summarises the data collected. The columns headed **Sensor/Data** and **Feature** summarises the types of sensor data, feature extracted from the sensor and feature representation or encoding algorithm. We have also listed any non-vision hardware used along with vision sensors. The last column having heading **Objective** summarises the comparison method and the objective from the application perspective. Most of the reviews on other areas of vision-based research have focused on categorising the discussion in terms of algorithms or techniques used. Articles reviewed often have common goals such as activity recognition, pose estimation and they also use common datasets. Thus a readily available and fair comparison between the methods used can be drawn. But, due to the wide ranging goal of researches in vision-based rehabilitation domain, authors have used very different data, features and comparison methods. Thus, it is very difficult to categorise each research in terms of methods or algorithms used and compare them. Instead, we propose our taxonomy based on end user application. However, discussion on each of application types have been further broken in paragraphs based on similarity of methods used. **The *Author* column in each Table also indicates the sub-category an article is placed into.** Primarily applications are placed into two major categories, rehabilitation and assessment. These can be further sub-categorised as listed below:

1. Rehabilitation: Automated rehabilitation system
   (a) Virtual rehabilitation
   (b) Direct rehabilitation
2. Assessment: Point in time assessment
   (a) Comparison
   (b) Categorisation
   (c) Scoring

3.2.1 Rehabilitation

In rehabilitation systems, the primary goal is to provide an automated home or in clinic system for patients to undergo physical therapy, gesture therapy or other rehabilitation exercises. Such a system guides patients to perform their rehabilitation tasks. Rehabilitation may be fully automated and/or clinician mediated. Research in this category normally aim to improve the patient's physical condition. Most of the research in this category are of the type **Virtual Rehabilitation**. In virtual rehabilitation, a patient's performance in a virtual world is assessed rather than directly assessing patient's physical performance. This includes avatar performing tasks in virtual world and the use of serious games for rehabilitation. Here, subjects are required to perform activities in a virtual world through real world movements. In **Direct Rehabilitation** systems, users are guided by a web-based interface to perform rehabilitation exercises, while their movements are directly tracked through vision-based sensor. In this case, physical performance of patient is measured instead of their avatar's performance or their ability to complete tasks in virtual world. Patient assessment may be inbuilt or may require clinicians.

3.2.2 Assessment

In assessment applications, the goal is to provide a point in time assessment of a patient's quality of motion linked to one or more body parts. There is no rehabilitation system involved. Assessment may be carried out in a clinical or non-clinical setting. Assessment application can be further categorised into three types and based on the way a user would receive the end output. The first type is **Comparison** where a patient's data (e.g., kinematic parameters) are extracted for comparison but there is no decisive automated scoring system available. In such applications, there may be statistical comparison like Analysis of Variance (ANOVA) or simple graphical comparison of kinematics represented by trajectories of an ideal vs a patient's joint angle or position in time. Second, we have **Categorisation** type applications, which are more decisive, where the main goal is classification. Movements may be classified as correct-incorrect or may be classified into a few types of abnormalities. This includes both gesture/posture and activity recognition. In the third **Scoring** type applications, a decisive score is attached to patient movements to assess their quality of motion. This can be clinical scoring such as FMA [59] or author-proposed scoring. The score maybe for assessing the quality of movement or quantify the differences from an ideal motion. Next, we review various articles published in the domain of monitoring and rehabilitation of musculo-skeletal patients according to the taxonomy developed. We, present all the articles in tabular format and discuss more relevant articles in details.

4 Virtual Rehabilitation

Author	Target	Dataset	Sensor/data	Feature	Objective
Sucar et al. 2008 [132] (NS)	Stroke, upper limbs	1 patient, 6 therapy sessions	RGB camera/ skin colour	Colour based hand trajectory	Gesture therapy through hand tracking
Sucar et al. 2010 [133] (NS)	Stroke, upper limbs, face	42 stroke patients, 21 therapy sessions	RGB camera, pressure gripper/ colour ball, hand pressure	Hand trajectory based on colour ball detection	Gesture therapy through hand tracking and face detection
Cameirão et al. 2010 [19] (NS,S)	Stroke, upper limbs	10 control subjects, 12 patients, virtual tasks therapy	RGB camera, finger tracking gloves/color marker, finger positions	lower arm and finger based kinematic parameters	ANOVA analysis based patient assessment
Aviles et al. 2011 [7] (NS)	Stroke, upper limbs	4 healthy subjects, 3 min sessions	RGB camera, pressure gripper/ colour ball, hand pressure	Hand trajectory based on colour ball detection	Assessment based on POMDP
Schönauer et al. 2011 [122] (S)	Chronic pain, whole body	6 patients, 4 to 6 game therapy sessions	Kinect, iotracker [111]/skeleton data	FAAST [134] gesture, skeleton trajectory	Comparison of Kinect and marker based skeleton tracking
Kurillo et al. 2011 [76] (NS)	Stepping in place	12 healthy individuals, stepping exercise	RGB-D camera	segmentation algorithm based hip angle	correspondence algorithm, variance analysis
Da Gama et al. 2012 [39] (S)	General rehabilitation, upper limb	50 correct, 60 wrong movement	Kinect/ skeleton data	Shoulder and elbow angles	Correct Movement Recognition and visual guidance
Chang et al. 2012 [25] (S)	SCI, shoulder movements	2 subjects, 1 healthy, 1 patient	Kinect, OptiTrack/ OpenNI skeleton	Right hand, elbow and shoulder trajectory	Trajectory comparison between Kinect and OptiTrack
Fern'ndez-Baena et al. 2012 [52] (S)	Knee rehabilitation	1 subject, 9395 frames	Kinect, Vicon/ NITE skeleton	Joint angle trajectory	Comparison of ROM, ME, and MER
Antón et al. 2013 [3] (S)	General rehabilitation, whole body	5 subjects, 80 recordings	Kinect/ skeleton data	Body posture in frames	Posture recognition of 3D avatar
Parry et al. 2014 [106] (NS,S)	Burn injury, upper limbs rehabilitation	30 children video data	8 camera 3d motion analysis system/skeleton data	shoulder elbow angles, elevation time	Assessment of ROM, ANOVA analysis
Adams et al. 2015 [1] (S,A)	Stroke, upper limb	14 impaired arm stroke patients	Kinect/ Kinect SDK	Joint angle, rate and acceleration	Proposes VOTA metrics for clinical assessment
Pei et al. 2016 [108] (S)	Stroke, whole body rehabilitation	3 healthy adults, 7 exercises, 20 repetitions	Kinect/ skeleton data	Joint angle comparison	Joint angle statistical analysis
Desai et al. 2016 [41] (S)	Stroke, whole body rehabilitation	10 healthy subjects	Kinect/skeleton data	Joint angle based skeleton model	Skeleton model based game

Avola et al. 2018 [8] (S,A)	PD, whole body	20 healthy, therapy session, 3 rehabilitation exercises	Kinect, LMC, HMD/ Skeleton data	Kinect: position, speed, angle, LMC: pinch strength, fingertip speed; 3 whole body rehabilitation exercises	LSTM for providing impairment scoring

Table 2: Virtual rehabilitation: systems where users perform activities in virtual world for completing rehabilitation tasks. **NS: Non-skeleton based; S: Skeleton-based; A: Automated Assessment**

The objective in virtual reality and serious games-based rehabilitation application is to provide a set of virtual tasks that will require the user to perform therapeutic gestures, rehabilitative or cognitive exercises (Table 2). The movement of the user in the real world is tracked through devices like Kinect, or other sensors that can accurately reproduce an user's movement in the virtual world, often through an avatar. In virtual rehabilitation systems role of CV is largely limited to tracking. In this survey, we have focused on works with secondary objective related to CV such as gesture, pose recognition or simple graphical comparison of trajectories of the concerned body joint angle. The discussion is split into, non-skeleton, skeleton-based and automated assessment systems.

4.1 Non-skeleton based

Virtual rehabilitation existed before skeleton tracking became feasible. Early research in this area used indirect methods for tracking human limb movements such as colour detection, object detection and others. In 2008, Sucar et al. [132] used skin colour to track hands for gesture therapy. Colour markers-based skeleton tracking has been used as a cheap alternative to inertial sensor tracking. Sucar et al. [133], developed rehabilitation system for hand movement of stroke patients. A total of 42 patients went through the rehabilitation program. A green ball attached to a hand gripper is used for tracking as shown in Fig. 2. Participants are required to move their arm through a simulated environment. Stroke patients often compensate reduced hand movement through trunk. This trunk compensation is observed through face tracking. Face detection and tracking is implemented using Haar Cascade classifiers [147]. Authors have also attempted to use their own skeleton tracking algorithms for rehabilitation in virtual reality [101].

   Non-skeleton based methods are inherently limited in ability due to lack of joint positions. Mostly, such methods are able to track either one joint such as an arm [132]. This can be sometimes compensated by using vision-based feature extraction methods such as body tracking from silhouette [83, 100]. In [100], depth information has been used in a RANSAC-based plane fitting method to discriminate the subject plane from background. This, combined

Fig. 2 An example of virtual rehabilitation where performance in the virtual world is considered for assessment. Here, hand is tracked indirectly through the green ball [133].

with morphological operations enabled the users to select the human silhouette. In virtual rehabilitation, since most of the assessment is done to achieve the objective of completing the game, there is little scope for further statistical or other algorithmic comparison. However, to tackle complex decision processes, algorithms such as Partially Observable Markov's Decision Process (POMDP) can be applied as in [7].

4.2 Skeleton-based

With the introduction of Microsoft Kinect in 2010, skeleton tracking became feasible and readily accessible. Chang et al. [25] used Kinect to measure joint position and angle trajectories in their proposed game for shoulder rehabilitation. Each participant was required to perform 6 different shoulder exercises which were quantified as correct or wrong by OpenNI middleware. The authors also compared Kinect skeletal data with OptiTrack for establishing the ground truth. Fern et al. [52] used several common exercises of hip, knees and shoulders in form of a serious game called rehabtimals. Joint rotation data over time was used to calculate kinematic metrics such as Range of Motion (ROM), Mean Error (ME) and Mean Error Relative (MER) in ROM. Da Gama et al. [39] used joint angles calculated from Kinect skeleton data to detect correct exercise posture. A total of 3 physiotherapists, 4 adults and 3

elderly subjects were used to evaluate the prototype. The system was able to recognize the correct movements 100% of the time under controlled conditions.

Here, most authors have used their own small datasets and thus, it is difficult to ascertain their generalisability. Owing to availability of skeleton positions, kinematic parameters have been used for performing statistical comparisons like ANOVA analysis. Small datasets are not sufficient for the application of Deep Learning (DL) algorithms but other algorithms such as HMM, DTW could have been used for comparing temporal sequences. Joint angle comparison is good for posture recognition. However, time sequence comparison algorithms are essential for comparing joint angle and/or joint position trajectories.

4.3 Automated assessment

Some virtual rehabilitation systems also have an integrated automated assessment. Adams et al. [1] proposed to assess upper limbs motor function through practice of ADL in virtual reality. Motor function metrics, such as duration, normalized speed, Movement Arrest Period Ratio (MAPR) obtained from skeletal tracking via Kinect were used to calculate Wolf Motor Function Test (WMFT) [154]. This score was co-related to the proposed Virtual Occupational Therapy Assistant (VOTA) metrics and it was found that the proposed metrics can be used to assess patient's ability to perform ADL. With their affected arm, 14 hemiparetic stroke patients were asked to participate in a virtual meal preparation activity. The results indicated satisfactory correlation between proposed VOTA metrics and the standard WMFT metrics. VRehab [8] used Long Short-Term Memory (LSTM) networks for estimating the degree of patient impairment. For evaluation, 20 healthy subjects were filmed using Kinect and Leap Motion Controller (LMC). Kinect was used to provide joint positions, angles and speeds as features while LMC provided pinch strength, average speed of fingertips. Three different LSTM networks were trained for regressing impairment scores for three different exercises. The trial included 5 patients who were scored by the system and 5 physiotherapists. The proposed system provided score was shown to be very close to the average score by physiotherapists.

5 Direct rehabilitation systems

Author	Target	Dataset	Raw data	Feature	Objective
Ghali et al. 2003 [57] (V)	Stroke, upper limbs	1 stroke patient	RGB camera/colour	colour based object trajectory, colour histogram	Object recognition, Event detection for text feedback to patient
Tao et al. 2004 [139] (M)	Stroke, upper limbs	Target reaching motion	RGB Camera, Qualisys/Colour marker	Joint angles trajectory	Statistical comparison with Qualisys

|				
Zariffa et al. 2011 [159] (V)	Neurological injury of palm	10 healthy subjects, 3 types of grip	2 RGB cameras	Hu invariant and contour signature extracted from background subtraction	Grip classification through KNN
Huang et al. 2011 [67] (V)	Cerebral palsy and muscle atrophy, upper limbs	4 patient subjects	Kinect/ Skeleton data	Joint angles based posture	Correct exercise count through posture recognition
Chang et al. 2011 [26] (V)	Cerebral palsy and muscle atrophy, upper limbs	2 patients, 34 days therapy	Kinerehab [67]	Joint angles based posture	Correct exercise count through posture recognition
Frisoli et al. 2012 [55] (M)	Hemiparesis stroke, upper limb	3 healthy subjects, 4 chronic stroke patients	BCI, Robotic arm, Kinect, Eye tracker/Skeleton data	SURF feature, Eye-gaze distance, 3D object maps	Robot arm aided rehabilitation, SVM based BCI signal classification, Lucas-Kanade object tracking
Chang et al. 2013 [27] (V)	Cerebral palsy, upper limbs	2 patients 25 days therapy	Kinect/ SDK skeleton	Joint angles based posture	Correct exercise count through posture recognition
Lin et al. 2013 [89] (V)	Tai-Chi upper limb rehabilitation exercise	2 patients with bone motor impairment	Kinect/ skeleton data	Normalized skeleton trajectory from subject and database	Grading of posture through mean error
Exell et al. 2013 [49] (M)	Stroke, arm rehabilitation	3 patients, 18 therapy sessions	Kinect, stimulation glove/SDK skeleton	Joint angle trajectory	Graphical comparison of Joint angle trajectory
Galeano et al. 2014 [56] (M)	Balance training system	6 healthy subjects	Kinect, Wii balance board/ skeleton data	Mediolateral and antero-posterior sways	Rehab through FES, provide feedback through posturography
Su et al. 2014 [131] (V)	General shoulder rehabilitation exercises	320 vectors for training, 6 subjects for testing	Kinect/ skeleton data	Euclidean joint distance based DTW vector	Neural Fuzzy system for performance evaluation
Benettazzo et al. 2015 [15] (V)	Shoulder rehabilitation exercises	10 participants, 2 exercises	Kinect/ OpenNI skeleton	Joint position Euclidean Distance from reference	Audio feedback and ANN based posture recognition
Devanne et al. 2018 [42] (M)	Low back pain rehabilitation	1 patient, 1 clinician, 3 exercises	2 arm humanoid robot, Kinect/ skeleton	GP-LVM based modelling space	Model clinician movement adapted to patient
Devanne et al. 2018 [43] (V)	General rehabilitation	5 over 60 subjects	Kinect/ skeleton	GMM, Riemannian manifolds	Classification by temporal segmentation analysis
Baptista et al. [10] (V)	Stroke, lower, upper limbs	10 healthy subjects, two sessions	Kinect/ Depth image based skeleton [125]	Joint angle comparison, Euclidean distance	Real time feedback

Schez-Sobrino et al. 2019 [121] (V)	Stroke, upper limbs	1 patient, rehabilitation exercises	Kinect/ Skeleton data	OE-DTW [141]	Classification by time series comparison

Table 3: Physical rehabilitation: Instead of virtual performance subject's physical movements are tracked for guiding or assessing rehabilitation. **V:Pure vision-based; M: Multi-Modal**

In direct rehabilitation systems, there is usually an exercise regimen prescribed for patients and the purpose is to demonstrate their functional improvement. Patients may be guided through a web-based interface for performing tasks similar to virtual rehabilitation type applications. However, unlike virtual rehabilitation, a subject's physical performance in the physical world is considered for further assessment or feedback. The discussion can be split into two parts: First, where CV sensor is exclusively used to obtain primary data and second, where non-vision systems such as assistive robots are used.

5.1 Pure vision-based

Ghali et al. [57] used object detection techniques for tracking hand movement. A camera was placed above kitchen platform and movement and orientation of objects was used as a measure to track the hand movement. Sequence of hand movement is used to determine whether an activity such as 'making coffee' is successfully completed. Kinect does not track finger joint positions. Zariffa et al. [159] used Hu invariant and contour signature extracted from background subtraction as features for classification of hand grip variations. Two cameras, one for top view and one for side view, were used to film 10 subjects against a standard background. Several types of grips fundamental to ADL such as lateral key grip were filmed. KNN was employed for classification.
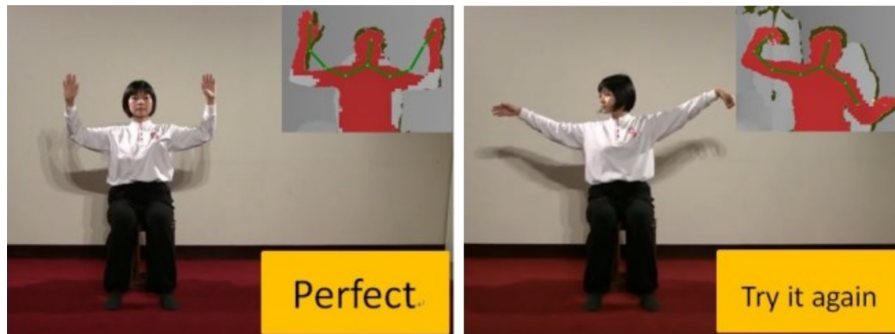
Fig. 3 An instance of a direct rehabilitation systems where a patient's performance is directly assessed through joint position tracking. In [89], Tai-Chi exercise pose is compared to a standard pose and feedback is provided.

The Kinect SDK provides advanced information such as kinematics and gesture recognition. Authors have used this to count the number of times correct posture was attained as a measure of rehabilitation progress [26, 27, 67]. Lin et al. [89] used 10 standing and 18 seated Tai-Chi regimen as rehabilitation exercises. Rehabilitation poses of two patients were compared to a perfect execution of Tai-Chi, for measuring progress over time. Patients were rehabilitated and monitored in two phases. First, with physiotherapists and then with video and Kinect. Posture attained by patients were compared through ME with target posture and subsequently graded. Feedback was provided to the user as shown in Fig. 3. Each time the system's assessment was compared to that of a physiotherapist for validation.

Su et al. [131] proposed a fully independent Kinect Enabled Home Rehabilitation (KEHR) system. The system provided four functions, 1) rehabilitation management software system, 2) reference exercises, 3) recording exercises performed at home using Kinect and 4) evaluation of performance. Performance was compared through DTW and Fuzzy Logic. Four different subjects were asked to perform different types of shoulder rehabilitation exercises in a controlled environment. Assessment was provided in form of messages like "right hand: good", "left hand: bad", "too slow" etc. Physicians and the KEHR system agreed 80% of the time.

5.2 Multi-Modal

In multi-modal applications, CV sensors (e.g, Kinect) are combined with other assistive technologies including, but not limited to assistive robots and electrical stimulation. Normally, the patients using the rehabilitation systems are guided via visual animation or clinicians. Galeano [56] used Functional Electro-Stimulation (FES) for assistance while providing visual feedback through posturography on skeletal data. Frisoli et al. [55] introduced a gaze independent, wearable Brain-Computer Interface (BCI) driven robotic exo-skeleton for upper limb rehabilitation in stroke patients. The first objective was to select real world object by estimating eye-gaze through a vision-based eye tracking system. Speeded Up Robust Features (SURF) [14] was used for object matching and Lucas-Kanade tracking algorithm [91] was applied to track objects using depth data from Kinect. Second objective was to assist patient arm movement for moving real world objects. To achieve this, signal from the BCI was fed to a Support Vector Machines (SVM) classifier to ascertain if subject intended to move his or her arm. Then, the signal was used to actuate robotic-arm. Devanne et al. [42] proposed a humanoid robot guidance system for rehabilitation from lower back pain. A Gaussian Process Based Latent Variable Model (GP-LVM) has been used to model exercise movements from a clinician. It then models the clinician's activity according to patient morphology to guide the rehabilitating patient.

In [26, 27, 67], the goal is to count correct postures by calculating the joint angles. This fails to tell us how close the patient is to getting the posture

correct. A slightly better way is to compare joint angle trajectories as in [49] or grading of error through ME as done by [88]. To judge if an exercise is executed correctly it is also essential to qualify the starting posture as correct [15], which is not the case in approaches mentioned above. These approaches are mostly primitive and lack analysis of the whole temporal sequence. Later approaches have taken advantage of time-sequence comparison algorithms such as DTW or variants of it like Open-ended DTW (OE-DTW) [121]. These have been combined with various grading methods for better understanding of patient's state. Clinicians mostly use their experience to judge a patient's state without taking into account kinematic parameters. Therefore, it may be beneficial to use kinematic parameters as training data and use clinician's score as labels to build a model that can present a true representative of clinicians. In automated rehabilitation, it is not always feasible to be guided via screen interface. In such scenarios, other assistive technologies like BCI and human motion imitating robots are very useful [51]. For assistive robots, it is important to work according to patient's morphology as demonstrated by [42]. The authors also show us very good implementation of latent model needed to transfer low dimensional latent space to high dimensional robot space through the probabilistic model GP-LVM.

6 Comparison

Author	Target	Dataset	Raw data	Feature	Objective
Goffredo et al. 2009 [60] (S)	Sit to stand	5 subjects and 5 healthy elderly	Joints marked in first RGB frame	Tracking by Gauss-Laguerre transform algorithm	Discriminatory analysis between young and old
Leu et al. 2011 [83] (K)	Gait abnormality	20 healthy, 10 patients	2 RGB camera system	skeleton extraction from silhouette	Knee angle based gait comparison
Stone et al. 2012 [129] (K)	Older adults, Gait abnormality	5 persons, 3 weeks walks in 4 homes	Kinect, Vicon/ skeleton data	gait parameters: speed, stride time and length	Graphical analysis
Scherer et al. 2012 [120] (K)	Stroke, brain activity	3 healthy subjects, hand closing and opening	Kinect, EEG/ NITE Skeleton data	Hand posture, EEG signals	Assessing (sub)cortical reorganization for hand movement
González et al. 2013 [61] (S)	Balance assessment	2 healthy subjects, 42 postures	Kinect, Wii balance board/ OpenNI skeleton	Center of position, statistically equivalent serial chain	Visual comparison of Center of Mass
Kurillo et al. 2013 [77] (K)	General upper limb reachable workspace	10 healthy subjects, 1 FHSD patient	Kinect/ skeleton information	Joint position, angle, ROM	ANOVA analysis, statistical comparison

Wang et al. 2013 [150] (S)	PD, upper and lower limbs	MSR-3D dataset [86], 1 healthy, 1 PD subject	Kinect/ SDK skeleton	Temporal skeleton trajectory	Action recognition through TASS for extracting clinically relevant features
Paiement et al. 2014 [104] (K)	Gait, knee injury	SPHERE-Staircase2014 dataset: 48 sequences, 12 healthy, 3 knee-injury subjects	Kinect/ SDK, OpenNI skeleton data	Joint positions, velocities, pairwise distance, angles	Abnormal gait detection from statistical modelling
Spasojević et al. 2015 [128] (K,D)	PD, Gait, shoulder	12 PD patients	Kinect/ skeleton data	9 MPIs on ROM, speed, rigidity, symmetry ratio	SVM, KNN, MLP based classification, LDA
Leightley et al. 2015 [82] (S,D)	Young and adult standard clinical tests	54 subjects, 13 rehabilitation exercises	Kinect/ skeleton data	K-means clustered poses based on Manhattan distances	Pose recognition through ANN, RF, GRBM, SVM, Kinematic comparison
Han et al. 2015 [64]	FHSD, upper body ROM	22 patients, 24 healthy subjects, standard ROM tests	Kinect/ skeleton data	Normalized body-centric 3D hand trajectory	Statistical comparison
Tao et al. 2016 [138] (S)	Stroke and Parkinson's, gait, sit and stand	SPHERE datasets [104] [138]	Kinect/ SDK, OpenNI, skeleton data	Joint positions, velocities, pairwise distance, angles	HMM model with discriminative classifier based online assessment
Antunes et al. 2016 [6] (S)	Stroke, General rehabilitation movements	ModifyAction, Weight&Balance, SPHERE-Walking2015 [104] datasets	Kinect/ Skeleton data	Normalized and temporally aligned skeleton sequence and standardized template	Human interpretable feedback to better match standard template
Vakanski et al. 2016 [144] (D)	General motion modelling	UTD-MHAD dataset [28]	Skeleton data	DTW aligned skeleton sequence, Mean log-likelihood	MDNN to predict mean log-likelihood as performance measure
Natarajan et al. 2017 [100] (K)	Gait analysis	20 healthy and 4 subjects with walking issues	RGB-D camera, Morphologically extracted skeleton from silhouette	Gait parameters, step-length, stride-length	Statistical analysis
Dolatabadi et al. 2017 [44] (K)	Stroke, upper body compensatory movement	10 healthy, 10 stroke patients	Haptic robot, Kinect/ SDK2 skeleton data	Joint angle trajectory	Comparison based on AUC
Spasojević et al. 2017 [127] (K, D)	PD, Gait,shoulder and Palm	30 PD patients, various stages	Kinect, sensor glove/ skeleton data	25 MPIs based on Kinematic parameters	SVM, MLP and KNN classification of disease stage, LDA

Baptista et al. 2017 [11] (S)	Physical activity assisting exercises	UTKinect dataset [156], stroke patient balance simulation	Kinect skeleton sequence	V2/se-	SS-DTW and TCD based temporal sequence alignment	Visual feedback for corrective action

Table 4: Comparison type assessment applications: Articles on patient monitoring applications that provide graphical or statistical comparison of patient action but do not provide a decisive patient assessment or score. **K: Kinematics-based modelling; S: Statistical modelling; D: Deep learning or Stochastic modelling**. EEG: Electroencephalograph

Table 4 summarises articles presenting comparative analysis of kinematic data obtained from vision sensors. In such systems, there is no rehabilitation program designed for patients. These articles are more important with regards to CV rather than clinical objectives. Authors have drawn comparison ranging from simple graphical visualisation, statistical techniques to more advanced Machine Learning (ML) algorithms. This discussion is split into three parts. First part discusses articles where kinematic data is directly used for comparison. Graphical and statistical comparison highlight differences between patient and healthy subjects' parameters. Second, applications where ML algorithms has been used for modelling kinematic data. The third part discusses use of DL algorithms for comparative analysis of patient motion.

6.1 Kinematics-based modelling

In this type of applications, researches directly use kinematic data for comparison. Before introduction of Kinect, authors have used other computer vision algorithms to extract skeleton. Leu et al. [83] used two cameras for filming 20 subjects against a standard background. Human silhouette was extracted through background subtraction and image segmentation. This data was compared to a standard stick figure model for extracting skeleton. For accuracy, the algorithm was tested against standard sensor-based marker. Simple graphical comparison showed visible difference between knee angle trajectories of regular and irregular gait. Natarajan et al. [100] also used their own tracking algorithm while introducing Reha@Home. The authors argued that detection on lower extremity joint in Kinect is not accurate enough. Reha@Home used depth information in combination with morphological operations to extract human silhouette. Four different subjects with varying conditions such as multiple sclerosis, were tested in a hospital setting both before and after treatment. The parameters for gait analysis are hip angle, knee angle, left and right foot step length and stride length. Performance of the system was evaluated through comparison with data from electrogoniometer. Graphical trajectories of gait parameters showed visible difference between be healthy subjects and patients.

The Toronto Rehab Stroke Pose Dataset (TRSP) [44] presents 3D joint positions consisting of upper arm movements for both stroke patients and health subjects. Kinect was used for tracking joint positions of 10 healthy subjects and 10 stroke survivors having restricted arm movements. Two experts

were recruited to annotate the dataset. The dataset was labelled into 3 different compensatory movements and one normal movement. Area Under Curve (AUC) values obtained from joint angle trajectory showed substantial measurable difference between regular healthy and physically impaired patients' examples.
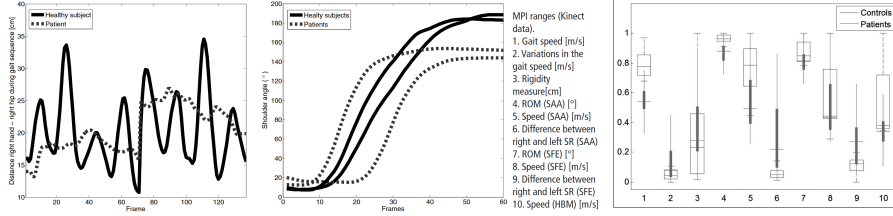
Fig. 4 Graphical comparison of patients and healthy subjects through kinematic parameters and joint angle trajectories [127].

Graphical and/or statistical comparison has also been used in situations where patients lack any specific impairment. Spasojevic et al. [128] used four different body movements and measurements, for discriminating PD patients from healthy subjects. Gait, Shoulder Abduction Adduction (SAA), Shoulder Flexion Extension (SFE) and Hand Boundary Movements (HBM) were considered for body movements. Speed, rigidity, ROM and symmetry ratio were used as measurement criteria. These wer combined to create a Movement Performance Indicator (MPI) vector of size 9. For example, only speed and rigidity was considered for gait movement. Experiments were conducted on 12 PD patients of stage 1, 2 and 3. Subjects were filmed from front at a distance of 1.5 meter. Ground truth was provided by physiotherapist. Graphical and statistical comparison based on kinematic parameters showed visible differences between patients and healthy subjects. Also, four ML classifiers SVM, K-Nearest Neighbour (KNN) and Multi-Layer Percepteron (MLP) were used for classification, among which SVM and MLP performed better. In 2017, Spasojevic et al. [127] added 16 more MPIs to the system described above. Data from finger tracking through sensory glove was used for 15 MPIs representing finger flexion, extension, tapping and hand rotation. For gait, another MPI was added making a vector of total 25 MPIs. Graphical comparison as illustrated in Fig. 4 showed visible differences between PD and healthy subjects. SVM, MLP and KNN were used to classify PD patient stages and healthy subjects. In this article, although ML algorithms have been used for classification, the research presents elaborate statistical comparison directly based on kinematic parameters.

6.2 Statistical modelling

Instead of directly comparing kinematic data, authors have also used ML algorithms for modelling human movement, which is subsequently compared statistically or graphically. Tao et al. [138] used HMM modelling, for online quality of motion assessment of gait on stairs, walking on flat surface, sitting and standing. For discriminating skeleton sequences using HMM, entire sequences has to be fed to a model. This was not possible in case of online assessment and thus, a variable window approach [99] was adapted to address the problem. Four different HMM models were used to extract features from skeleton data to classify abnormalities using SVM.

Wang et al. [150] devised a series of exercises for musculo-skeletal patients targeting PD patients. Activities include walking, walking with counting and sit to stand. Again, skeleton information was obtained through Kinect placed in front of patient and on top of table. Step size, postural swing level, arm swing level, stepping time were used as criteria to asses a patient's mobility level. The paper proposed a Temporal Alignment Spatial Summarisation (TASS) algorithm to isolate repetitive skeletal movements from video stream through Skeletal Action Unit. The SAU extracted clinically important kinematic parameters like arm swing level and stepping time for evaluation. This method was evaluated against the standard MSR-Action3D [86] action recognition dataset. For clinical validation, a single PD patient and a healthy subject were asked to perform walking and sit-to-stand experiment. Data from both the experiments showed difference between the PD patient and the healthy subject.

Antunes et al. [6] framed the assessment problem as feedback to be provided to a skeleton sequence to better match a standard execution sequence template. The system has been evaluated on three publicly available datasets. The first, ModifyAction used pairs of actions from UTKinect [156] and MSR-Action3D [86] dataset. The second dataset was SPHERE-Walking2015 [104] which contained normal walking and simulated stroke patient walking. The third dataset, called Weight&Balance was introduced in this paper and it presented simulated data of stroke affected arm mobility. Data normalization was used for spatial alignment and DTW was used for temporal alignment. The importance of this research resides in the feedback mechanism that was provided at each instant for better execution of human action.

Baptista et al. [11] also saw the problem as essentially finding the difference between two skeleton sequences. This allowed them to use publicly available UTKinect [156] dataset to address the problem without specifically using patient or simulated patient data. The authors used Sub-Sequence DTW (SS-DTW) [97] and TCD [33] algorithms to match user action to a specific template and provide feedback highlighting deviations from normal execution.

6.3 Introduction of stochastic methods

The area of vision-based rehabilitation and monitoring has not seen extensive application of DL methods. This is mainly due to lack of large scale datasets needed to train DL networks. In 2015, Leightley et al. [82] presented the Kinect 3D Active (K3D) dataset which captured motions based on common clinical assessments. used to determine altered patient movements. Fifty four subjects aged 18 to 81 were asked to perform 13 clinical tests including balance, open and closed eyes, jump, chair stand and others. Owing to the diverse age related conditions the subjects' movements varied widely for any given activity. Several algorithms were used for action classification out of which SVM and Artificial Neural Network (ANN) achieved the best accuracy. To assess clinical condition the activities were further analysed in terms of average time taken to complete an action.

In absence of large-scale publicly available dataset, simulating or generating data has been also considered. Vakanski et al. [144] trained their Mixture Density Neural Network (MDNN) on the standard action recognition UTD-MHAD dataset [28], to model human movement for each action. Mean log-likelihoods of observed sequences were used as the performance metric for evaluating the consistency of a subject's performance. Then, random noise was imparted to generate deviations from standard action and these deviations were measured. The proposed model was programmed to be usable with skeleton data captured through Kinect.

The articles presented above propose exclusive assessment type application and often do not include any rehabilitation method. They have used more robust approach for assessment in the sense that authors have compared more kinematic parameters, used more advanced statistical analysis and have used bigger datasets. For example, in rehabilitation type applications, many authors have chosen simple joint angle or joint angle trajectory comparison [49, 67]. In general, authors have used better statistical comparison including Linear Discriminant Analysis (LDA) [127, 128], TCD [11], likelihood [104, 144], ANOVA [77]. In comparison type applications, we also see the implementation of more robust kinematic parameters such as 25 different MPIs in [127, 128], normalized sequences [64], temporally aligned sequences [6]. As a result, such applications are able to carry out more complex comparison including gait analysis and compensatory movements, as opposed to simple gesture or posture recognition of a single or few joints. We also observe the introduction of publicly available datasets which paves the way for competitive evaluation of the proposed models [11, 104, 138]. However, statistical comparison does not provide a decisive scoring or classification of a patient's condition. The next two sections discuss applications that can classify or grade patient's quality of motion.

7 Categorisation

Author	Target	Dataset	Raw data	Feature	Objective
Cho et al. 2009 [32](R)	PD gait recognition	7 PD, 7 healthy subjects	RGB camera	PCA and LDA from binary silhouette	MDC based classification
Taati et al. 2012 [136](S)	Posture compensation	7 healthy subjects simulating compensation	Kinect/ skeleton data	3D oreintation of subset of joint lines	Posture classification based on HMM and SVM
Metcalf et al. 2013 [95](R)	Stroke, hand rehabilitation	2 hand modes, 76 videos, 1692 frames	Kinect/ depth data	Contour, kinematic hand model	Hand grip classification, joint angle comparison
Leightley et al. 2013 [79](S)	General rehabilitation activities	20 subjects, 200 activities, 60225 frames	Kinect/ SDK skeleton data	Position, velocity, energy	Activity recognition by PCA-SVM, RF
Kertész et al. 2013 [71](S)	General rehabilitation exercise, whole body	7 subjects, 8 exercises 654 samples	Asus Xtion/ NITE skeleton data	skeleton trajectory based reference model	SVM, Numerical model based Posture recognition
Jun et al. 2013 [69](S)	Knee Osteoarthritis	5 healthy subjects, 2 sets * 4 trials	Kinect/ skeleton data	Normalized and PCA reduced skeleton information	Subject classification using KNN, automation for individualization of exercise regimen
Liu et al. 2013 [90](S)	General rehabilitation exercise, arm	10 subjects, 3200 postures	Kinect/ OpenNI skeleton	45 normalized 3D coordinates	Posture recognition, Action recognition through SVM
Kargar et al. 2014 [70](S)	Gait severity, Go and get up test	12 elderly subjects, 50 samples	Kinect/ skeleton data	Gait features: step distance, duration; skeletal features: joint distance, angle	Fall risk classification through SVM and BoW
Cary et al. 2014 [24](S)	Stroke, upper limbs	10 subjects, 5 poses, 5 repetitions	Kinect, skelton information	17D vector, 4 joint angles and body inclination angle	ANN Gesture recognition
González-Ortega et al. 2014 [62](R)	Cognitive psychomotor exercises	10 healthy, 3 frontal lobe injury, 2 dementia, 5 tests, 14 exercises	Kinect/ OpenNI skeleton, depth image	Joint position, Mean and Gaussian curvature from depth image	Skeleton, Face eye, ears, nose detection for posture recognition using HK classification [16]
Palma et al. 2016 [105](S)	General rehabilitation exercises, upper, lower limbs	14 subjects, 10 exercise, 100 incorrect	Kinect/ Skeleton data	Quantized joint angles	Error recognition using HMM and MD-DTW
Capecci et al. 2016 [21](S)	General motor disabilities, whole body	19 healthy, 14 patients, 5 Lower back pain exercises	Kinect V2/ skeleton data	Relative joint angle, velocity, constraints	HSMM for discriminating improper execution

Richter et al. 2107 [115](S)	Errors in hip abduction	3 patients, 3 scenarios	Kinect/ skeleton data	Local and hierarchical coordinates based I-DTW	SVM based error classification, live feedback
Leightley et al. 2017 [81](S)	Motion instability identification	K3D dataset [82]	Kinect/ skeleton data	Joint groups comprising body centric coordinate positions, angles	Unstable motion classification using DCNN and ML algorithms
Richter et al. 2017 [114](R)	Errors in hip abduction	1 patient	Kinect/ skeleton data	automatically determined class specific joint combinations	Weight-based hip abduction error classification, visual feedback
Leightley et al. 2017 [80](R)	Standard clinical tests, whole body	K3Da Dataset [82]	Kinect/ skeleton data	group of joint angle, distance, body lean angle	Mobility classification through ML algorithms
Pogrzeba et al. [112](R)	Stroke, PD, repetitive hand movements	10 healthy, 20 patient subjects, drum beats 122-126 per minutes	Kinect Kinect SDK/ skeleton data	PCA reduced skeleton trajectories, mean, standard deviations	Logistic regression based correct speed, consistency, variability classification
Khan et al. 2018 [72](S)	Infants with motor disabilities	10 infants, supine and prone position video	Kinect/ RGB, depth data	size, area, position of bounding box through Body segmentation	Infant movement classification through SVM
Rivas et al. 2018 [116](S)	Stroke patient engagement	5 patients, 10 sessions, gesture therapy	Camera pressure gripper/ colour marker	Kinematic features, grip pressure features	MSNB classifier for patient states: anxiety, pain, engagement and tiredness
Chen et al. 2018 [30](S)	General upper limb rehabilitation exercises	20 samples, 11134 frames	Kinect/ RGB, Depth	Skeleton extraction algorithm, skin color detection for face orientation	DTW based correct exercise recognition
Zhi et al. 2018 [163](S)	Stroke, upper body compensatory movement	TRSP dataset [44]	Kinect/ skeleton data	Noise reduced and body centric joint positions	Classification through SVM and LSTM
Antunes et al. 2018 [5](S)	Senior lower body fitness	11 young, 10 elderly subjects, 4 exercises	Kinect/ RGB, skeleton data	RGB video, skeleton trajectory	ANN, LSTM activity recognition, Kolmogorov-Smirnov test
Li et al. 2018 [84](S)	General rehabilitation exercises	UI-PRMD dataset [145]	Kinect/ skeleton data	Scaled and mean shifted joint angle trajectory, RMS based soft-labels	Modelling and evaluation of movement through GAN

Table 5: Categorisation type assessment applications: Articles that discriminate patient movement as correct-incorrect or provide a discrete rating. **R: Rule-based; S: Statistical or Stochastic Algorithms-based.** MDC: Minimum distance classification, MSNB: Multi-Resolution Semi-Naive Bayesian

In this section, we review articles where the primary goal of the research is to categorise a patient activity into discrete categories including, but not limited to correct/incorrect posture and good/bad movement. In contrast to comparative analysis, articles reviewed in this section are more decisive in terms of providing patient assessment. Technically, most of the articles in this section have the goal of posture or action recognition where discrimination is done between improper and proper execution of activities. But, it also includes disease severity classifications, determination of a patient's cognitive abilities and so on. The discussion is split into two parts: 1) hand-crafted or rule-based and 2) statistical algorithms based.

7.1 Rule-based

In some cases, final posture is important and simple hand-crafted algorithms are sufficient for correct posture recognition. Metcalf et al. [95] used depth frames for measuring hand (finger) kinematics. A Kinect device was placed 80 cm above a table where subjects were filmed. Binary image of palm was extracted from depth and RGB data. Palm contour was then fitted to a geometrical kinematic model to determine joint angles. Based on sequence of joint key-points (finger tip, finger spaces), a grip classification algorithm was developed. Gonzalez-Ortega et al. [62] used computer vision for assessment of patient's cognitive motor abilities. Here, the goal was to see whether a patient can understand verbal instructions and perform simple motor tasks. A group of 10 subjects were used to provide healthy reference while 3 subjects with frontal lobe injury and 2 with dementia were used for rehabilitation using the proposed system. The subjects were asked to perform 14 different type of movements such as "touch right eye with right hand" in a controlled environment. Facial expression was detected by combining skeleton data and depth image from Kinect with AdaBoost-based face detector. Eyes and nose were detected using HK classification [16], which is based on curvature obtained from depth image. In psycho-motor exercises, the final posture is important to judge whether the subject understood the instructions. The proximity of 3D hand position to eyes, ear and nose helped in determining successful exercise execution. The result provided by system was compared with physicians and overall successful monitoring rate was 96.2%. Leightley et al. [80] used the K3D dataset [82] for automated human mobility analysis. K-means clustering was used to create clinically relevant joint groups for each action. The joint groups containing relevant joint trajectories were classified for recognizing the action. Discrimination between well-performed and poorly performed action was done on basis of the standard deviation method proposed in [13].

7.2 Statistical and Stochastic Algorithms-based
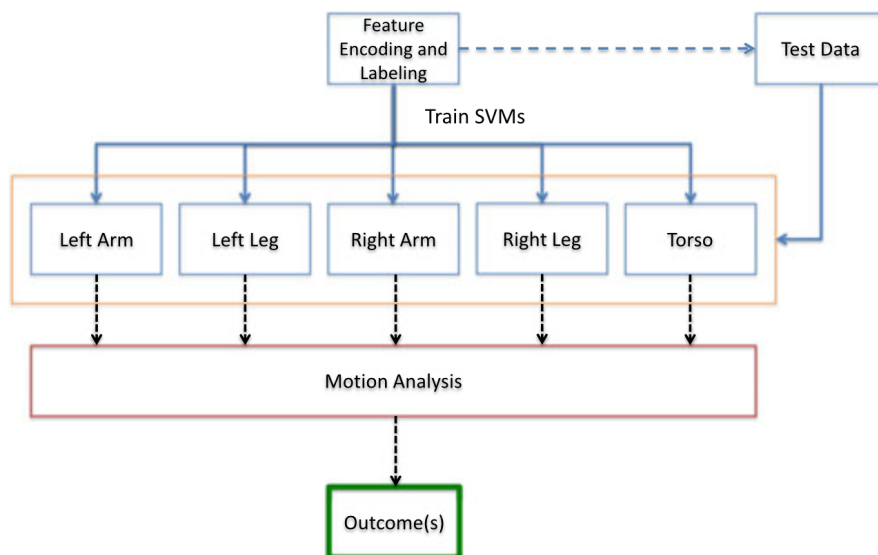
Fig. 5 An example of categorisation type system. Group of joints are used as encoded features for SVM. Patient's are classified as mobile or immobile. [80].

Researchers have extensively used advanced Machine learning (ML) algorithms for categorisation type systems. Taati et al. [136] developed a interactive system where subjects interacted with robots for posture correction. Again, skeleton data was obtained from a Kinect device placed 90 cm behind the subject and 60 cm above the subject. Seven healthy subjects were asked to simulate a series of compensated mobility. Such movements include shoulder hike, trunk rotation compensation, lean forward and slouch postures. For posture classification, a combined HMM and SVM-based algorithm was used. An active learning strategy, which used a combination of manual and automatic labelling, was employed to label the data for classification. The overall accuracy is 86%. Palma et al. [105] presented a method for detecting deviations from normal movements using HMM and Multiple-Dimension DTW (MD-DTW) [140]. The authors created a dataset of 10 different upper and lower limb movements such as hip abduction, elbow flexion and so on with 14 healthy subjects. Then, a cohort of 10 subjects were asked to perform the same movements incorrectly with specified errors. For analysis, the activities were divided into two parts: 1) The limb moved away from the body and 2) the limb moved towards body. HMM was found to be more accurate for detecting error in movements when compared to MD-DTW.

In recent times, Generative Adversarial Networks (GANs) have been used to generate synthetic data including but not limited to human faces and human poses. Li et al. [84] used the UI-PMRD dataset [145] to generate synthetic dataset of incorrect human activities. Four different GANs models were trained, which included two Deep Convolutional GANs (DCGAN), a Wasserstein GAN and a Recurrent GAN. A 1D Convolutional Neural Network (CNN) was trained as discriminator with the GANs and soft-metric based on absolute differences was used for evaluating the performance of GANs. Modelling or replicating kinematic data through GAN is a major contribution of this article, although it aims to classify physical movements.

In categorisation type applications, authors have used techniques ranging from very basic rule-based classification to state-of-the-art GANs. In [62, 95], authors have used hand-crafted algorithms to classify patient's stages which are very specific conditions whereas in [80], simple standard deviation was used to classify a patient's state. It is very difficult to ascertain the generalisability of these applications in abesence of comparison through publicly available datasets. **Authors have extensively used ML algorithms such as SVM in categorisation type applications [136, 163].** As primary data, authors have mostly used only skeleton data, with exception of [95], who have used depth data only. Some authors have relied on kinematic parameters as extracted features [70, 71, 81, 105]. However, others have introduced statistical techniques for feature extraction. For example, Junet al. [69] used PCA (Principal Component Analysis) reduced kinematics and Zhi et al. [163] used noise reduced kinematics. Then, these features have been used for classification through standard algorithms such as SVM, CNN and LSTM. Use of classification algorithms have enabled authors to grade patient's state rather than presenting a simple visual, graphical or statistical comparison.

8 Scoring

Author	Target	Dataset	Sensor	Feature	Objective
Venugopalan et al. 2013 [146](A)	Brain injury, upper limbs	16 videos, 6 different poses, 4 subjects	Kinect/ skeleton data	normalized position, velocity, acceleration	Template matching and DTW for calculating similarity score
Hsiao et al. 2013 [66](A)	Stroke, box and block test, hand	6 trials, 100 boxes moved	Kinect/ OpenNI depth	Contour, largest circle in contour	Hand and box detection, calculate number of box transferred
Cuellar et al. 2014 [38](A)	General rehabilitation exercises, lower and upper body	10 healthy subjects, 3 motion and 2 posture holding exercise	Kinect/ skeleton data	Transformed 3D vector from quaternion rotation of joint angles	Performance scoring, Template based posture matching, DTW based skeleton sequence matching

Khan et al. 2014 [73](C)	PD, rapid finger tapping	13 Parkinson's patient, 387 clips	RGB camera/colour marker	Motion analysis of fingers by motion-template gradient algorithm	Severity classification through SVM based on UPDRS attributes
Olesh et al. [102](A)	Stroke hemiparesis, upper limbs	9 patients, 10 arm movements, 5 to 28 repetitions	Kinect/ skeleton data	Joint angle trajectory	Automated scoring based on averaged temporal profiles
Wang et al. 2014 [149](C)	Stroke, upper limbs	24 stroke patients, Shoudler and elbow movements	Kinect V2/ skeleton data	9 kinematic parameters composed of velocity, energy and angle	Automated clinical scoring (FMA) through SVR
Dyshel et al. 2015 [46](C)	LID, upper limb	9 PD with LID patients 24 recording following AIMS protocol	Kinect/ SDK skeleton data	Single Number deducted from most discriminatory joint motion segment	Automated AIMS grading through soft SVM based algorithm
Ciabattoni et al. 2016 [36](A)	General rehabilitation exercises	6 subjects, 5 exercises	RGB-D camera	quaternion based posed distance, Virtual joint angles	DTW algorithm based assessment scores
Ciabattoni et al. 2016 [35](A)	Lower back pain exercises	5 subjects, 5 exercises	RGB-D camera/ skeleton data	joint angle, distance, torso surface	Exercise performance scoring system based on fuzzy logic
Soran et al. [126](C)	Infants, SMA	15 SMA patients, 72 minute film	Kinect/ Colour marker	limb trajectory	CNN based clinical disease progression sore
Capecci et al. [22](C)	General rehabilitation exercises, whole body	22 healthy, 19 neurological disability, 5 exercises	Kinect/ skeleton data	joint angles, velocity, constraint angles	HSMM and DTW model for automated clinical scoring
Li et al. 2018 [85](C)	Idiopathic PD, whole body	9 PD patients, 134 footage for 4 tasks	RGB camera/ CPM skeleton data	Subset of joint positions	Classification of PD type, Regression for clinical assessment score
Eichler et al. 2018 [47](C)	Stroke, upper limb	12 stroke, 10 healthy, FMA mmovements	2 Kinect V1/ skeleton data	FMA related kinematic parameters	Automated clinical scoring (FMA) through SVM and RF
Einarsson et al. 2018 [48](C)	PD, whole body	33 healthy, 30 patient's,	Kinect/ Skeleton data	Normalized, scaled joint coordinates	Automated UPDRS scoring based on Spare Ordinal classification
Capecci et al. [23](A)	Stroke, PD, back-pain rehabilitation exercise	44 healthy 34 patients	Kinect / skeleton data	joint angles	rule, template based algorithm, Spearman Correlation

Liao et al. 2019 [87](A)	General rehabilitation exercises	UI-PRMD dataset [145]	Kinect/ Skeleton data	Auto-encoder reduced skeleton sequence, GMM log-likelihood based performance metric	Automated assessment based on LSTM regressed scoring

Table 6: Scoring type assessment system: articles that provide a clinical or author proposed scoring of patient movements. **A: Author-proposed scoring; C: Clinical Scoring**

In this section, we review articles which aim to provide automated assessment of a patient's state. This include both clinical (e.g., FMA, UPDRS) and author proposed (non-clinical) scoring. For musculo-skeletal diseases, there are often multitude of factors that describe a patient's state or condition. Simple movements such as hip abduction or individual exercises may be classified into correct or incorrect. But, to describe a patient's state, clinicians often use standard scoring systems including FMA, UPDRS and others. Scoring may be discrete or continuous. In technical terms, authors have used both classification and regression for scoring. This discussion can be split into two parts: 1) author proposed and 2) clinical scoring.

8.1 Author proposed scoring

PReSenS, developed by Cuellar et al. [38] is a rehabilitation exercise systems where physiotherapists can remotely upload exercise templates to be followed by patients at home. A complete exercise program was developed consisting of two major types of exercises, posture holding and motion. Posture was compared to single exercise template whereas, motion was compared by the time series matching algorithm DTW. Features such as joint angle and joint rotation was used with DTW for action comparison. For experiment, data from 10 healthy participants were collected. They were asked to do diagnostic exercises, such as arms up, arm extension and flexion, leg-up, flamenco and cross arm. These exercises are are widely used in physical therapy . All motion signals were summarised using Piece-Wise Aggregation Approximation for scoring the performance.

Khan et al. [73] used rapid finger tapping test for clinical evaluation of PD patients. A total of 387 video footage were used from patients with advanced PD. Severity was rated by physiotherapist on a scale of 0 to 3. A group of 84 healthy subjects were clinically evaluated in the same way. Subjects were asked to tap their hands besides their face and above shoulders. For assessment, first, region of interest was selected as rectangles besides face. Face detection was achieved by Haar Cascade classifier [147] and motion-template gradient algorithm [18] was used to detect hand movements. Kinematic parameters for calculating UPDRS features were extracted and classified using SVM.

A major limitation of many assessment systems is that they require users to sit in front of camera and perform exercises. It may be difficult for musculo-

skeletal patients to operate the system and perform exercises in a highly constrained setting. Compliance may be poor in such cases for actual patients. Venugopalan et al. [146] proposed a rehabilitation system for traumatic brain injury where patients can be monitored in real time. In the experiments, two Kinect cameras and a near infra-red motion sensor were used to film patients at home. Real-time patient data from the system was compared with data from observation in clinical setting to compute similarity score. The score was calculated through template matching based on DTW. For evaluation, 16 videos were captured which covered six different movements that were performed by four different volunteers.

Liao et al. [87] proposed a log-likelihood based performance metric to train their DL framework for assessment of rehabilitation exercises. Low level skeleton data was represented through a deep Auto Encoder (AE) network to initially train a Gaussian mixture model for calculating log-likelihood. Using the UI-PRMD dataset [145], the authors then trained and compared the performances of CNN, RNN and Hierarchical Neural Network (HNN) [45], where the log-likelihood based performance metric was used as label to regress the network for predicting deviations from normal actions.

8.2 Clinical scoring

The ultimate goal of any assessment system is to assess the state of a patient in terms of clinical scoring. This is a very difficult task considering the multitude of factors involved in assessment. Eichler et al. [47] proposed a two Kinect camera based system for automated FMA. The two cameras were placed at 45 degree angle with respect to the subject. Temporal synchronization was done through a network time protocol server. From Kinect skeleton data, kinematic features relevant to FMA were calculated. The features included sequence time length, minimum and maximum of each measure, average variance of each measure, difference between start and end values of each sequence, variation of average speed and acceleration of each measure. SVM was used for classification. A cohort of 22 participants took part, including 12 stroke and 10 healthy subjects. The proposed system is able to successfully predict scores on for the two standard motions "Salute" and "Hand lift". An ideal automated FMA system would be able to assess full range of impairments in both upper and lower extremity.

Performance of patients also vary from time to time during the day. Following the Abnormal Involuntary Movement Scale (AIMS) protocol, Dyshel et al. [46] recorded 9 PD patients with varying severity of Levodopa-Induced Dyskinesia (LID). The subjects performed two motor tasks normally used for UPDRS assessment. After motion segmentation and noise reduction, discriminative features are extracted. For each joint motion, chunks are extracted and put into distributions represented by two 30-bin histograms. One histogram represents normal and other represents dyskinetic state. Earth Mover Distance (EMD) is calculated and 10 motion chunks representing highest discrimination
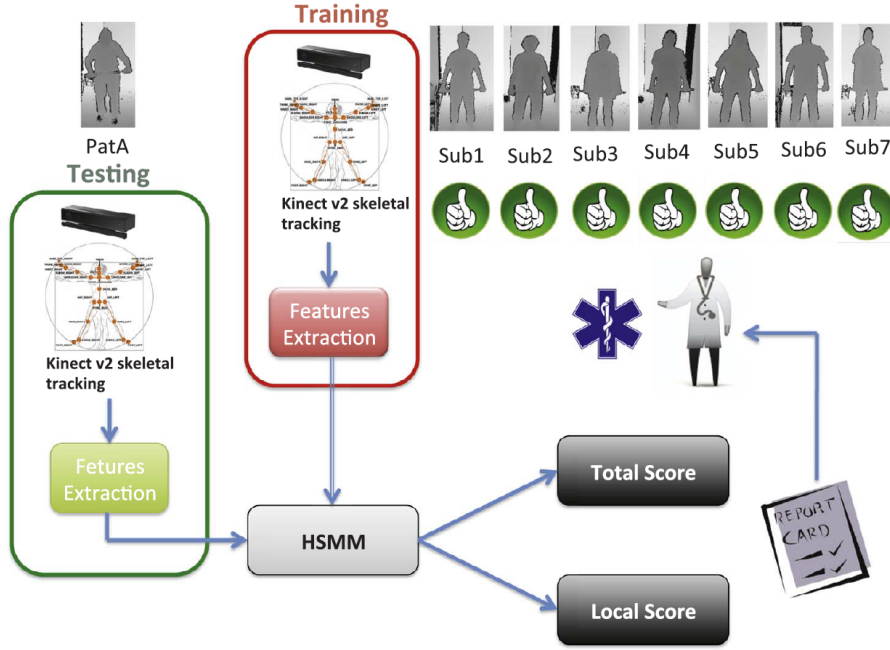
Fig. 6 An illustration of scoring type systems. Extracted features from patient are compared to a pre-trained HSMM for automated clinical scoring [22]. Total score reflect the overall score of the whole body whereas local score include features that assist clinicians to localise movement errors. PatA: Patient A

were selected. Each 10 dimensional vector was then reduced to a single number using one of the three methods: average motion length, average motion speed, distribution of quantized motion lengths. Soft-margin SVM-based algorithm was used to calculate AIMS score.

Since the introduction of DeepPose [142] in 2014, CNN-based human pose estimation has achieved very high accuracy. Li et al. [85] used the well-known Convolutional Pose Machines (CPM) [153] for extracting skeleton data for analysing LID. Levodopa is used to treat PD but its prolonged use causes motor complications (Dyskinesia). The study involved creating a publicly available dataset involving 9 participants having LID. The skeleton data extracted using CPM was used to generate 15 kinematic features. These features helped to score the participants based on the UPDRS.

In above-mentioned articles, authors have relied mostly on skeleton data obtained from Kinect, with the exception of Li et al. [85] who have used CNN for pose extraction from RGB data. Authors have not combined RGB and skeleton data, which may result in improve accuracy. Mostly authors have used kinematic parameters directly as primary features. In [38] and [36] quaternion-based pose distance and has been used as primary features. Quaternions can help in catpuring rotation in 3D and is better than the normally used Eu-

clidean distance. Liao et al. [87] have used Auto-encoders for dimensionality reduction of skeleton sequences. In most applications we have seen little application of dimensionality reduction techniques applied to kinematic data. On applications requiring continuous score Support Vector Regression (SVR) and LSTM has been used. In order to measure similarity score, temporal sequence matching algorithms such as HMM and DTW has been used. In applications requiring discrete scoring such as UPDRS, SVM has been mainly used. It remains to be seen how modern DL algorithms would perform classification in such cases where large-scale datasets are not available.

9 Datasets

Author	Impairment	Details	Sensor/Data
SPHERE-Staircase2014 [104]	Walking-up stairs	48 sequences, 12 subjects, normal and abnormal gait	Kinect/ Open NI skeleton
SPHERE-Walking2015 [138]	Walking	40 sequences, 10 subjects, normal and abnormal gait	Kinect/ Kinect SDK, OpenNI SDK skeleton
SPHERE-SitStand2015 [138]	Sit to stand	109 sequences, 10 individuals, restricted knee, hip, freezing	Kinect/ Kinect SDK, OpenNI SDK skeleton
TRSP [44]	Stroke, compensatory movement	10 healthy, 10 stroke 4 compensatory movements	Kinect, Haptic robot/ Kinect SDK skeleton
Parkinson's pose estimation [85]	PD, LID, UPDRS assessment tasks	526 sequence, PD, LID patients, 4 UPDRS assessment tasks	RGB Camera/ CPM [153] skeleton
UI-PRMD [145]	General rehabilitation exercises	10 subjects, 10 exercises, 10 repetitions	Kinect Vicon/ Kinect SDK skeleton
KIMORE Dataset [23]	Stroke, PD, back pain exercises	44 healthy, 34 patient subjects, 5 exercises 5 repetitions	Kinect/ RGB, depth, skeleton
AHA-3D Dataset [5]	Senior lower body fitness	11 young, 10 elderly subjects, 4 exercises	Kinect/ RGB, depth, skeleton

Table 7: Publicly available datasets that include physically impaired patient motion

Table 7 summarises publicly available datasets that are captured through vision-based methods. SPHERE is a series of datasets that presents normal and physically impaired movements for walking, walking-up stairs and sit to stand movements. Vakanski et al. [145] introduced the UI-PRMD dataset consisting of 10 different physical activities commonly performed in physical rehabilitation or therapy scenarios. The dataset provides skeleton data obtained through Kinect along with joint angles. Mean Square Error (MSE) on joint angles has been used by authors to calculate variability between each subject, which has also provided a benchmark for establishing incorrect movements. Unlike other areas of CV, most of the research is based on relatively small datasets and are not available publicly.

10 Discussion

In this section, the methods used in articles reviewed are discussed in terms of their usage, drawbacks and disadvantages. Research in this area is very different from objectives like activity recognition where the common goal is to explore machine learning and pattern recognition techniques to recognise various activities. Also, often the datasets used to evaluate the models are same and thus a direct comparison between various methods employed by authors is useful. However, due to the widely varying goals, datasets used and types of physical impairments, such comparison in this domain is difficult. Instead, for the benefit of readers, we chose to compare the general techniques and algorithms employed to achieve the goals. Following section 2, we split the discussion into data, feature encoding and feature comparison.

10.1 Physical impairment data

In articles discussed in this review, authors have mostly used Kinect-based skeleton data. The main advantage is that Kinect provides RGB videos, depth videos and 3D joint positions as well as posture through a very cheap and easy to use hardware/software system. Thus, authors from domains other than CV can take advantage of it. However, Kinect system is not very accurate [152] and today's DL-based solution outperforms the Kinect system both in-terms of 2D [20, 50] and 3D pose estimation [107, 157]. Due to the lack of direct comparison, it is difficult to gauge the scope of improvement in the articles reviewed with DL-based methods instead of Kinect. Unlike other areas of CV application such as activity recognition, authors have not used RGB or depth data in combination with skeleton information. RGB data lacks the precise joint positions whereas skeleton data lack information such as optical flow, curves, edges and others. Modern neural networks are very good at learning such information. Combining skeleton data with RGB information guides the DL model to focus on RGB features on the human body. This has lead to increased accuracy in activity recognition models [12,143] and thus, researches in this domain can also benefit from the same. Authors have also used colour-based tracking, including, tracking hand while holding a coloured ball and skin colour tracking. These methods were in use before the introduction of Kinect, but some of them are still in use today. They have several limitations such as tracking only one part of body and are subject to noise, background interference. It also needs to be noted that Kinect is no longer in production and researchers will need to switch to other devices such the Orbec Astra [37]. In [37], authors discuss the interchangeability and accuracy of Orbec Astra and the Kinect device. So it is worth taking the time and effort to switch to new devices and techniques. Authors have also used other non-vision based devices such as BCI, LMC which when used with vision-based devices expand the domain of physical rehabilitation to other areas such as BCI to support physical rehabilitation [55].

10.2 Feature encoding

Method	Usage	Drawbacks	Alternatives
colour trajectory	track body part through coloured object	limited skeleton tracking, prone to background interference	skeleton tracking
skeleton trajectory	tracking body parts	do not quantify physical characteristics	kinematic parameters
Kinematic parameters	indicates physical ability	very specific to type of impairment(s)	None
contour signature	mark hand boundaries for grip classification	cannot handle noisy, blurry images	DL based segmentation [9]
Hu invariant	Image boundary descriptor for grip classification	cannot handle noisy, blurry images	DL based segmentation [9]
AUC	For comparing kinematic trajectories	AUC can be same for different curves	Statistical analyses, KL divergence
Performance metrics (ME, MER, RMSE, N-RMSE)	encoding patient sequence error w.r.t standard template	over-fitting, difficult to generalize	learnable encoding methods
Log likelihood	Probabilistic encoding of skeleton sequence	specific formula needed for calculate likelihood, non trivial estimation	KL-divergence, Cross Entropy
SURF	encodes local RGB features	less accurate than SIFT although faster, clutterd keypoints	SIFT, ORB
Depth maps	body part segmentation, skeleton detection	missing colour, texture, skeleton information	use with RGB and skeleton data
GP-LVM	dimensionality reduction of skeleton sequence	assumes independent distributions, needs strong prior	PCA, LDA, Autoencoders
Gaussian mixture model	Encoding skeleton sequence for performance metric	expensive for high dimensional data, need to set number of clusters	Spectral clustering, Manifold learning
Gauss Laguerre transform (GLT)	encoding video squence in GLT domain	needs manual marking to select area for transform	SIFT, SURF, ORB etc. for keypoint descriptors
Human body sillhoute	human body segmentation	cannot handle noisy, blurry images	DL based semantic segmentation [9]
Pairwise skeleton trajectory	enables relative trajectory encoding	overftting, cannot learn the general trend	learnable encoding methods
K-means clustering	encoding kinematic parameters	No of clusters needs to be manually set	GMM

Distances (Manhattan, Euclidean)	encoding patient sequence distance wrt standard template	overftting, cannot learn the general trend	learnable encoding methods
PCA reduced sequence	dimensionality reduction of skeleton sequence	mean and covariance does not always describe distribution	LDA, autoencoder
Colour segmentation	track body part through coloured object	prone to background noise, interference	DL based semantic segmentation [9]
GAN generated sequences	generation of artificial data	hard to train and converge	Different types of GANs [68]
Quaternion sequences	represent orientation and rotation of skeleton sequence in 3D	contains only rotation but no scaling and translation	Affine transformation matrices
Motion template gradient	human motion encoding, through successive frame sillhoute	pixel based approaches prone to background noise	Optical flow based approaches, graph-cut algorithm
Autoencoder	dimensionality reduction of sequence	requires more data, not generally used for dimensionality reduction	PCA, LDA
LDA	dimensionality reduction of skeleton sequence	needs labelled data, lot of tunable parameters	PCA, Autoencoders

Table 8: A summary of feature encoding methods used, their drawbacks and alternatives that can be used

Table 8 highlights the various feature encoding methods used by authors. It also outlines their drawbacks and suggests alternatives. Many authors have used skeleton trajectories or kinematic parameters derived from these trajectories directly as features for comparison. While such parameters are useful for purposes such as posture recognition, joint mobility determination, these are highly specific to the physical impairment, thus not generalisable and may suffer from over-fitting. Instead of encoding kinematic parameters, relationship between parameters such as performance metrics, distances, pairwise relations and others, can be used for encoding. Although these methods can produce better results, they also suffer from the same drawbacks such as inability to learn, over-fitting and so on. A better alternative would be to learn from the data instead of comparing kinematic parameters numerically or graphically. Thus, more recently authors have used techniques including, but not limited to DTW, HMM and TASS to build temporal models that can help to discriminate differences between patient and ideal pose sequences. Authors have also attempted to encode feature from RGB videos for goals such as activity recognition. Feature encoding techniques include, Hu moments, colour-based segmentation, motion template gradients and so on. Mostly, these are pixel-based techniques suffer from noise interference and do not work in case of blurry im-

ages. Modern alternatives include use of generalised local feature descriptors such as Scale-Invariant Feature Transform (SIFT), SURF, Oriented FAST and rotated BRIEF (ORB) or image descriptors such as Bag of Words (BoW), Histogram of Oriented Gradients (HoG) and others. Modern techniques also involve DL-based algorithms for semantic segmentation [9] which have produced state-of-the-art results but again, these require large-scale datasets. In absence of large datasets, using GANs for modelling artificial patient data can be very useful as shown by [84]. There are many variants of GANs, each of which have their own domain of applicability and limitations. In [68], authors present a quantitative comparisons of various GAN types. Instead of manually selecting joints for recognizing abnormal motions [80] one can use attention-mechanism [94] to learn the importance of joints for a particular impairment.

10.3 Feature comparison

Methods	Usage	Description	Benefits	Drawbacks
numerical comparison	numerically comparing kinematic parameters, skeleton position	joint movement, posture comparison	lacks generalization, lacks statistical significance	statistical comparison
graphical comparison	graphically comparing kinematic parameters, skeleton sequence	comparison of normal vs patient trajectory	lacks generalization, lacks statistical significance	statistical analysis (ANOVA, Chi-Squared etc.)
Statistical analysis	tests to see and compare different results	comparison of kinematic parameters	unable to categorise or grade patients	time series comparison algorithms such as HMM
POMDP	MDP where underlying process is not directly observable	virtual game assessment	Intractable, assumes convex value functions	Reinforcement learning
KNN	non parametric classification algorithm	classification of kinematic parameters	need to determine K value, high computation cost	SVM
SVM	classifies by finding discriminating hyperplanes	classification of kinematic parameters	choosing appropriate kernel, memory intensive	Ensamble, RF methods, Neural networks
Neural fuzzy system	learns fuzzy parameters with neural networks	build model based on clinician experience instead of rules	difficult to interpret results and generalize	
ANN	stochastic learning networks that learns from examples	classification of kinematic parameters	prone to overfitting, not good with images or sequential data	DCNN, LSTM
GRBM	RBM that uses binary latent variables to model hidden states	pose recognition	difficult to train with contrastive divergences, requires sampling like GS, MCMC etc.	MRF, CRF, Autoencoders
MDNN	Combination of ANN and Mixture density models	classify through log-likelihood based performance metrics	Number of mixture modes need to be fixed manually	various adaptations of MDNN [92, 151, 158]

HMM, HSMM	Represents probability distributions over sequences	Discriminate between patient and model skeleton sequences	limited by markov property, cannot capture higher order dependencies	CRF, Bayesian Networks
DTW, MD-DTW, SS-DTW	Measures similarity between two temporal sequences	Calculate similarity between patient and model skeleton sequences	Quadratic complexity, works with only smaller templates	HMM
DCNN	Networks that learns grid like topolgy like images	classifying poses from RGB images, videos	resource consuming, requires big datasets	Capsule Networks [117]
MSNB	Bayesian classifier variable independence assumption	classification of kinematic parameters	needs explicit modelling of interdependence between variables	Fuzzy logic, different types of SNB [162]
LSTM	RNN architecture, learns temporal sequences	learning skeleton sequence features	slow, cannot memorize long temporal sequences	TCN [78], ODE Networks [29]

Table 9: A summary of feature comparison methods used, their drawbacks and alternatives that can be used. MRF: Markov Random Fields

In Table 9, various feature encoding methods used by authors is highlighted along with drawbacks and possible alternatives. Most basic methods used by researchers are simple numerical and graphical comparison of skeleton trajectories, joint angles or other kinematic parameters. The results are hard to generalise beyond the examples presented and may lack statistical significance. A better alternative to use some statistical test such as ANOVA analysis, Chi-squared tests etc. Graph trajectories can be compared with methods such as KL-divergence which could provide statistically significant results. In general, authors have used temporal sequence comparison algorithms like HMM, DTW, LSTM and their variants such as HSMM (Hidden Semi-Markov Model) MD-DTW, Incremental DTW (I-DTW), SS-DTW and others. Note that these algorithms can be used for sequence encoding as well as sequence comparison. Some authors have used classification algorithms such as K-means, SVM to compare encoded sequence generated by HMM or DTW. The same could also be done with techniques such as Conditional Random Fields (CRF) and Bayesian networks. Other techniques involve use of generative models such as Restricted Boltzmann Machine (RBM), Gaussian RBM (GRBM), Semi-Naive Bayes (SNB) for classification. These have largely been replaced by DL-based algorithms such as DCNN, LSTM and TCN. To compensate for lack of datasets, one can also look at learning from single images [155]. When comparing sequential data with DL, LSTM is the most popular type of architecture that has been used. But, recently Temporal Convolutional Networks (TCN) [78] and Ordinary Differential Equation (ODE) networks [29] have shown very competitive results and these two architectures are being actively pursued by researchers. DCNNs have been almost exclusively used for processing image and video data but researchers are now exploring the use of capsule networks [117] for the same.

11 Conclusion

In this review, we have collected, summarised and analysed major computer vision-based research in the area of rehabilitation and assessment of patients having physical impairments. In this article, we present our own taxonomy. To the best of our knowledge, this is the only article to date, that has covered the latest advances in this application area, and presented them from a CV application point of view. It particularly focuses on comparison and assessment of abnormal human motions. This is especially significant due to the wide-ranging and hugely varying manifestations of abnormal or impaired human movements. We have seen simple graphical comparison of joint angle trajectories to application of complex algorithms such as GANs. The absence of image, video-based DL algorithms is quite contrasting to other areas such as pose estimation and action/activity recognition where DL algorithms have been almost exclusively used. This could be down to unavailability of large-scale datasets. Also, in this domain, most articles are exclusively focused on the use of skeletal information as raw data. This means low-level image/video features and high-level contextual cues (e.g. body-objects interaction) are not a part of the intelligent processing. **Movement information deduced form skeleton information is sparse in nature, whilst image-based dense optical flow form video information is richer in contextual information. Thus, research in this domain may benefit from meaningful combination of skeleton and spatio-temporal information linked to video data.** In case of scoring type applications, DL-based scoring may not be easy to adapt as its often more complicated for patients and clinicians to understand. More recently, researchers have attempted to fit existing scoring methods while training DL-based models. CV will play a significant role in rehabilitation and assessment, which is a sub-field of health and social care. But owing to several factors such as difficulty in obtaining patient data, ethical issues and so on, this area is yet to be extensively explored by the CV community. We conclude the discussion with recommendation for future research:

Datasets: **The lack of DL-based methods compared to other applications of CV could be due to unavailability of large-scale publicly available datasets demonstrating physically impaired patients' activities. The publicly available datasets mentioned in Table 7 are relatively small as compared to modern dataset targeting DL. For example, NTU-RGB dataset [123] targeted towards DL-based activity recognition contains 60K samples and is much larger than the datasets presented in Table 7. Therefore, research in this domain needs publicly available large-scale datasets to take advantage of modern DL methods. Although datasets are indispensable, the problem can be mitigated to a certain extent by using GANs. Data Augmentation GAN (DAGAN) has been purpose-built for the augmenting data [4]. It is based on conditional GAN and is capable of generating unseen within-class data samples. This is different to traditional data augmentation techniques, where images/videos are ro-**

tated or translated to augment the data. Likewise, balancing GANs can be used to mitigate class imbalance problems [93]. In absence of real data, authors have recently created fully synthetic data from GANs. For example, Li et al. [84] generated synthetic data for incorrect human activity from four different types of GANs. Frid-Avar et al. [54] used GAN to create synthetic data for liver-lesion classification. Besides GAN, single-shot or few-shots learning has the potential to learn from small amount of data [135]. These are often presented as Siamese networks to discriminate or tell deviations from reference sample [34]. In [34], the authors used a two-stream convolutional Siamese network to for person re-identification. A similar approach could be adapted for assessing physically impaired persons where deviation from regular healthy activity could be measured through a single-shot or few-shot learning.

Statistically significant results: Simple graphical or numerical comparison of skeleton-trajectory is not suitable to produce statistically significant results. In such cases, authors can use statistical tests including, but not limited to ANOVA and Chi-Square. However, such statistical approaches are model-less. These approaches lack generalisation and are not scalable. Moreover, they are often inferior to the model-based techniques. Venugopalan et al. (2013) and Liao et al. (2019) show that model-based approaches such as DTW, Gaussian Mixture Model (GMM) log-likelihood works better than non-model approaches such as Euclidean distance, Mahalanobis distance and Cross-Correlation. Thus, instead of directly comparing kinematic parameters we recommend modelling the data with algorithms including, but not limited to HMM, DTW and TASS. Moreover, a combination of classification/regression algorithm such as SVM, Random Forest (RF) are even better than modelling the data alone. Taati et al. [136] model the data using HMM and use SVM to classify the data and show that the combination of HMM and SVM works better than using SVM alone.

Modern DL techniques: Researchers in this domain have begun to use DL-based techniques such as CNNs and LSTMs. However, authors have used very basic and obsolete architectures that fail to demonstrate the true potential of these algorithms. CNNs used by Zhi et al. [163] and Leightley et al. [81] are very basic in nature. Their recommendation is to use modern pre-trained CNN architectures including, but not limited to EfficientNets [137] and NasNet [110]. Authors have introduced TCN [74, 78] as an effective and faster alternative to LSTMs. TCN-based networks are suitable for human activity recognition as demonstrated by Kim et al. [74]. Similarly, ODENets [29] are being extensively researched for processing of temporal information. DL is a rapidly evolving field and the introduction of better and efficient techniques is a regular occurrence. For example, Graph Neural Networks first introduced in [75] has

been extensively used by authors for recent state-of-the-art activity recognition models [124] and authors in this domain can potentially benefit from the same.

12 Acknowledgement

References

1. Richard J Adams, Matthew D Lichter, Eileen T Krepkovich, Allison Ellington, Marga White, and Paul T Diamond. Assessing upper extremity motor function in practice of virtual activities of daily living. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23(2):287–296, 2015.
2. Md Atiqur Rahman Ahad, Anindya Das Antar, and Omar Shahid. Vision-based action understanding for assistive healthcare: A short review. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops 2019*, pages 1–11, 2019.
3. David Antón, Alfredo Goñi, Arantza Illarramendi, Juan José Torres-Unda, and Jesús Seco. Kires: A kinect-based telerehabilitation system. In *e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on*, pages 444–448. IEEE, 2013.
4. Antreas Antoniou, Amos Storkey, and Harrison Edwards. Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340*, 2017.
5. João Antunes, Alexandre Bernardino, Asim Smailagic, and Daniel P Siewiorek. Aha-3d: A labelled dataset for senior fitness exercise recognition and segmentation from 3d skeletal data. In *BMVC*, page 332, 2018.
6. Michel Antunes, Renato Baptista, Girum Demisse, Djamila Aouada, and Björn Ottersten. Visual and human-interpretable feedback for assisting physical activity. In *ECCV*, pages 115–129. Springer, 2016.
7. Héctor Avilés, Roger Luis, Juan Oropeza, Felipe Orihuela-Espina, Ronald Leder, Jorge Hernández-Franco, and Enrique Sucar. Gesture therapy 2.0: Adapting the rehabilitation therapy to the patient progress. *Probabilistic Problem Solving in BioMedicine*, page 3, 2011.
8. Danilo Avola, Luigi Cinque, Gian Luca Foresti, Marco Raoul Marini, and Daniele Pannone. Vrheab: a fully immersive motor rehabilitation system based on recurrent neural network. *Multimedia Tools and Applications*, pages 1–28, 2018.
9. Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.
10. Renato Baptista, Enjie Ghorbel, Abd El Rahman Shabayek, Florent Moissenet, Djamila Aouada, Alice Douchet, Mathilde André, Julien Pager, and Stéphane Bouilland. Home self-training: Visual feedback for assisting physical activity for stroke survivors. *Computer methods and programs in biomedicine*, 176:111–120, 2019.
11. Renato Baptista, Michel Goncalves Almeida Antunes, Djamila Aouada, and Björn Ottersten. Video-based feedback for assisting physical activity. In *VISAPP*, 2017.
12. Fabien Baradel, Christian Wolf, and Julien Mille. Human activity recognition with pose-driven attention to rgb. In *BMVC 2018-29th British Machine Vision Conference*, pages 1–14, 2018.
13. Richard N Baumgartner, Kathleen M Koehler, Dympna Gallagher, Linda Romero, Steven B Heymsfield, Robert R Ross, Philip J Garry, and Robert D Lindeman. Epidemiology of sarcopenia among the elderly in new mexico. *American journal of epidemiology*, 147(8):755–763, 1998.

14. Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *ECCV*, pages 404–417. Springer, 2006.

15. Flavia Benettazzo, Sabrina Iarlori, Francesco Ferracuti, Andrea Giantomassi, Davide Ortenzi, Alessandro Freddi, Andrea Monteriù, Silvia Innocenzi, Marianna Capecci, Maria Gabriella Ceravolo, et al. Low cost rgb-d vision based system to support motor disabilities rehabilitation at home. In *Ambient Assisted Living*, pages 449–461. Springer, 2015.

16. Paul J Besl and Ramesh C Jain. Invariant surface characteristics for 3-d object recognition in range images. *Computer Vision, Graphics, and Image Processing*, 31(3):400, 1985.

17. Matteo Bigoni, Silvia Baudo, Veronica Cimolin, Nicola Cau, Manuela Galli, Lucia Pianta, Elena Tacchini, Paolo Capodaglio, and Alessandro Mauro. Does kinematics add meaningful information to clinical assessment in post-stroke upper limb rehabilitation? a case report. *Journal of physical therapy science*, 28(8):2408–2413, 2016.

18. Gary R Bradski and James W Davis. Motion segmentation and pose recognition with motion history gradients. *Machine Vision and Applications*, 13(3):174–184, 2002.

19. Mónica S Cameirão, Sergi Bermúdez i Badia, Esther Duarte Oller, and Paul FMJ Verschure. Neurorehabilitation using the virtual reality based rehabilitation gaming system: methodology, design, psychometrics, usability and validation. *Journal of neuroengineering and rehabilitation*, 7(1):48, 2010.

20. Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. nov 2016.

21. Marianna Capecci, M Gabriella Ceravolo, Francesco Ferracuti, Sabrina Iarlori, Ville Kyrki, Sauro Longhi, Luca Romeo, and Federica Verdini. Physical rehabilitation exercises assessment based on hidden semi-markov model by kinect v2. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 256–259. IEEE, 2016.

22. Marianna Capecci, Maria Gabriella Ceravolo, Francesco Ferracuti, Sabrina Iarlori, Ville Kyrki, Andrea Monteriù, Luca Romeo, and Federica Verdini. A hidden semi-markov model based approach for rehabilitation exercise assessment. *Journal of biomedical informatics*, 78:1–11, 2018.

23. Marianna Capecci, Maria Gabriella Ceravolo, Francesco Ferracuti, Sabrina Iarlori, Andrea Monteriù, Luca Romeo, and Federica Verdini. The kimore dataset: Kinematic assessment of movement and clinical scores for remote monitoring of physical rehabilitation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(7):1436–1448, 2019.

24. Francisco Cary, Octavian Postolache, and Pedro Silva Girao. Kinect based system and artificial neural networks classifiers for physiotherapy assessment. In *Medical Measurements and Applications (MeMeA), 2014 IEEE International Symposium on*, pages 1–6. IEEE, 2014.

25. Chien-Yen Chang, Belinda Lange, Mi Zhang, Sebastian Koenig, Phil Requejo, Noom Somboon, Alexander A Sawchuk, Albert A Rizzo, et al. Towards pervasive physical rehabilitation using microsoft kinect. In *PervasiveHealth*, pages 159–162, 2012.

26. Yao-Jen Chang, Shu-Fang Chen, and Jun-Da Huang. A kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Research in developmental disabilities*, 32(6):2566–2570, 2011.

27. Yao-Jen Chang, Wen-Ying Han, and Yu-Chi Tsai. A kinect-based upper limb rehabilitation system to assist people with cerebral palsy. *Research in developmental disabilities*, 34(11):3654–3659, 2013.

28. Chen Chen, Roozbeh Jafari, and Nasser Kehtarnavaz. Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In *ICIP*, pages 168–172. IEEE, 2015.

29. Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in neural information processing systems*, pages 6571–6583, 2018.

30. Yen-Lin Chen, Chin-Hsuan Liu, Chao-Wei Yu, Posen Lee, and Yao-Wen Kuo. An upper extremity rehabilitation system using efficient vision-based action identification techniques. *Applied Sciences*, 8(7):1161, 2018.

31. Yixin Chen, Siyuan Huang, Tao Yuan, Siyuan Qi, Yixin Zhu, and Song-Chun Zhu. Holistic++ scene understanding: Single-view 3d holistic scene parsing and human pose estimation with human-object interaction and physical commonsense. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 8648–8657, 2019.

32. Chien-Wen Cho, Wen-Hung Chao, Sheng-Huang Lin, and You-Yin Chen. A vision-based analysis system for gait recognition in patients with parkinson's disease. *Expert Systems with applications*, 36(3):7033–7039, 2009.

33. Wen-Sheng Chu, Feng Zhou, and Fernando De la Torre. Unsupervised temporal commonality discovery. In *ECCV*, pages 373–387. Springer, 2012.

34. Dahjung Chung, Khalid Tahboub, and Edward J Delp. A two stream siamese convolutional neural network for person re-identification. In *Proceedings of the IEEE international conference on computer vision*, pages 1983–1991, 2017.

35. Lucio Ciabattoni, Francesco Ferracuti, Sabrina Iarlori, Sauro Longhi, and Luca Romeo. A novel computer vision based e-rehabilitation system: From gaming to therapy support. In *Consumer Electronics (ICCE), 2016 IEEE International Conference on*, pages 43–44. IEEE, 2016.

36. Lucio Ciabattoni, Francesco Ferracuti, Giuseppe Lazzaro, Luca Romeo, and Federica Verdini. Serious gaming approach for physical activity monitoring: A visual feedback based on quantitative evaluation. In *International Conference on Consumer Electronics*, pages 209–213. IEEE, 2016.

37. Alina Delia Călin Adriana Coroiu and Adriana Coroiu. Interchangeability of kinect and orbbec sensors for gesture recognition. In *2018 IEEE 14th international conference on intelligent computer communication and processing (ICCP)*, pages 309–315. IEEE, 2018.

38. Manuel P Cuellar, Maria Ros, Maria J Martin-Bautista, Y Le Borgne, and Gianluca Bontempi. An approach for the evaluation of human activities in physical therapy scenarios. In *International Conference on Mobile Networks and Management*, pages 401–414. Springer, 2014.

39. Alana Da Gama, Thiago Chaves, Lucas Figueiredo, and Veronica Teichrieb. Guidance and movement correction based on therapeutics movements for motor rehabilitation support systems. In *2012 14th Symposium on Virtual and Augmented Reality*, pages 191–200. IEEE, 2012.

40. Alana Da Gama, Pascal Fallavollita, Veronica Teichrieb, and Nassir Navab. Motor rehabilitation using kinect: a systematic review. *Games for health journal*, 4(2):123–135, 2015.

41. Kevin Desai, Kanchan Bahirat, Sudhir Ramalingam, Balakrishnan Prabhakaran, Thiru Annaswamy, and Una E Makris. Augmented reality-based exergames for rehabilitation. In *Proceedings of the 7th International Conference on Multimedia Systems*, page 22. ACM, 2016.

42. Maxime Devanne et al. Generating shared latent variables for robots to imitate human movements and understand their physical limitations. In *ECCV*, pages 190–197. Springer, 2018.

43. Maxime Devanne, Olivier Remy-Neris, Beatrice Le Gals-Garnett, Gilles Kermarrec, Andre Thepaut, et al. A co-design approach for a rehabilitation robot coach for physical rehabilitation based on the error classification of motion errors. In *2018 Second IEEE International Conference on Robotic Computing (IRC)*, pages 352–357. IEEE, 2018.

44. Elham Dolatabadi, Ying Xuan Zhi, Bing Ye, Marge Coahran, Giorgia Lupinacci, Alex Mihailidis, Rosalie Wang, and Babak 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*, pages 375–381. ACM, 2017.

45. Yong Du, Wei Wang, and Liang Wang. Hierarchical recurrent neural network for skeleton based action recognition. In *CVPR*, pages 1110–1118, 2015.

46. Maria Dyshel, David Arkadir, Hagai Bergman, and Daphna Weinshall. Quantifying levodopa-induced dyskinesia using depth camera. In *ICCV Workshops*, pages 119–126, 2015.

47. Nadav Eichler, Hagit Hel-Or, Ilan Shmishoni, Dorit Itah, Bella Gross, and Shmuel Raz. Non-invasive motion analysis for stroke rehabilitation using off the shelf 3d sensors. In

2018 International Joint Conference on Neural Networks (IJCNN), pages 1–8. IEEE, 2018.

48. Gudmundur Einarsson, Line KH Clemmensen, Ditte Rudǎ, Anders Fink-Jensen, Jannik B Nielsen, Anne Katrine Pagsberg, Kristian Winge, and Rasmus R Paulsen. Computer aided identification of motion disturbances related to parkinson's disease. In *International Workshop on PRedictive Intelligence In MEdicine*, pages 1–8. Springer, 2018.

49. Timothy Exell, Christopher Freeman, Katie Meadmore, Mustafa Kutlu, Eric Rogers, Ann-Marie Hughes, Emma Hallewell, and Jane Burridge. Goal orientated stroke rehabilitation utilising electrical stimulation, iterative learning and microsoft kinect. In *Rehabilitation robotics (icorr), 2013 ieee international conference on*, pages 1–6. IEEE, 2013.

50. Haoshu Fang, Shuqin Xie, Yu-Wing Tai, and Cewu Lu. Rmpe: Regional multi-person pose estimation. In *ICCV*, volume 2, 2017.

51. Juan Fasola and Maja J Matarić. A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction*, 2(2):3–32, 2013.

52. Adso Fern'ndez-Baena, Antonio Susín, and Xavier Lligadas. Biomechanical validation of upper-body and lower-body joint movements of kinect motion capture data for rehabilitation treatments. In *Intelligent networking and collaborative systems (INCoS), 2012 4th international conference on*, pages 656–661. IEEE, 2012.

53. Nuno Ferreira Duarte, Mirko Rakovic, Jorge Marques, and José Santos-Victor. Action alignment from gaze cues in human-human and human-robot interaction. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018.

54. Maayan Frid-Adar, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Synthetic data augmentation using gan for improved liver lesion classification. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 289–293. IEEE, 2018.

55. Antonio Frisoli, Claudio Loconsole, Daniele Leonardis, Filippo Banno, Michele Barsotti, Carmelo Chisari, and Massimo Bergamasco. A new gaze-bci-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6):1169–1179, 2012.

56. D Galeano, F Brunetti, D Torricelli, S Piazza, and Jose L Pons. A tool for balance control training using muscle synergies and multimodal interfaces. *BioMed research international*, 2014, 2014.

57. Ahmed Ghali, Andrew S Cunningham, and Tony P Pridmore. Object and event recognition for stroke rehabilitation. In *Visual Communications and Image Processing 2003*, volume 5150, pages 980–990. International Society for Optics and Photonics, 2003.

58. Rohit Girdhar, Deva Ramanan, Abhinav Gupta, Josef Sivic, and Bryan Russell. Actionvlad: Learning spatio-temporal aggregation for action classification. In *CVPR*, pages 971–980, 2017.

59. David J Gladstone, Cynthia J Danells, and Sandra E Black. The fugl-meyer assessment of motor recovery after stroke: a critical review of its measurement properties. *Neurorehabilitation and neural repair*, 16(3):232–240, 2002.

60. Michela Goffredo, Maurizio Schmid, Silvia Conforto, Marco Carli, Alessandro Neri, and Tommaso D'Alessio. Markerless human motion analysis in gauss–laguerre transform domain: An application to sit-to-stand in young and elderly people. *IEEE Transactions on Information Technology in Biomedicine*, 13(2):207–216, 2009.

61. Alejandro González, Mitsuhiro Hayashibe, and Philippe Fraisse. Three dimensional visualization of the statically equivalent serial chain from kinect recording. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 4843–4846. IEEE, 2012.

62. D González-Ortega, FJ Díaz-Pernas, Mario Martínez-Zarzuela, and Miriam Antón-Rodríguez. A kinect-based system for cognitive rehabilitation exercises monitoring. *Computer methods and programs in biomedicine*, 113(2):620–631, 2014.

63. Tanmay Gupta, Alexander Schwing, and Derek Hoiem. No-frills human-object interaction detection: Factorization, layout encodings, and training techniques. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9677–9685, 2019.

64. Jay J Han, Gregorij Kurillo, Richard T Abresch, Evan de Bie, Alina Nicorici, and Ruzena Bajcsy. Reachable workspace in facioscapulohumeral muscular dystrophy (fshd) by kinect. *Muscle & nerve*, 51(2):168–175, 2015.
65. Samitha Herath, Mehrtash Harandi, and Fatih Porikli. Going deeper into action recognition: A survey. *Image and vision computing*, 60:4–21, 2017.
66. Chih-Pin Hsiao, Chen Zhao, and Ellen Yi-Leun Do. The digital box and block test automating traditional post-stroke rehabilitation assessment. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on*, pages 360–363. IEEE, 2013.
67. Jun-Da Huang. Kinerehab: a kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, pages 319–320. ACM, 2011.
68. Daniel Jiwoong Im, He Ma, Graham Taylor, and Kristin Branson. Quantitatively evaluating gans with divergences proposed for training. *arXiv preprint arXiv:1803.01045*, 2018.
69. Seung-kook Jun, Suren Kumar, Xiaobo Zhou, Daniel K Ramsey, and Venkat N Krovi. Automation for individualization of kinect-based quantitative progressive exercise regimen. In *Automation Science and Engineering (CASE), 2013 IEEE International Conference on*, pages 243–248. IEEE, 2013.
70. B Amir H Kargar, Ali Mollahosseini, Taylor Struemph, Wilson Pace, Rodney D Nielsen, and Mohammad H Mahoor. Automatic measurement of physical mobility in get-up-and-go test using kinect sensor. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3492–3495. IEEE, 2014.
71. Csaba Kertész. Physiotherapy exercises recognition based on rgb-d human skeleton models. In *Modelling Symposium (EMS), 2013 European*, pages 21–29. IEEE, 2013.
72. Muhammad Hassan Khan, Julien Helsper, Muhammad Shahid Farid, and Marcin Grzegorzek. A computer vision-based system for monitoring vojta therapy. *International journal of medical informatics*, 113:85–95, 2018.
73. Taha Khan, Dag Nyholm, Jerker Westin, and Mark Dougherty. A computer vision framework for finger-tapping evaluation in parkinson's disease. *Artificial intelligence in medicine*, 60(1):27–40, 2014.
74. Tae Soo Kim and Austin Reiter. Interpretable 3d human action analysis with temporal convolutional networks. In *CVPR Workshop*, pages 1623–1631. IEEE, 2017.
75. Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
76. G KURILLO et al. Real-time 3d avatars for tele-rehabilitation in virtual reality. *Medicine Meets Virtual Reality 18: NextMed*, 163:290, 2011.
77. Gregorij Kurillo, Alic Chen, Ruzena Bajcsy, and Jay J Han. Evaluation of upper extremity reachable workspace using kinect camera. *Technology and Health Care*, 21(6):641–656, 2013.
78. Colin Lea, Michael D Flynn, Rene Vidal, Austin Reiter, and Gregory D Hager. Temporal convolutional networks for action segmentation and detection. In *CVPR*, pages 156–165, 2017.
79. Daniel Leightley, John Darby, Baihua Li, Jamie S McPhee, and Moi Hoon Yap. Human activity recognition for physical rehabilitation. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, pages 261–266. IEEE, 2013.
80. Daniel Leightley, Jamie S McPhee, and Moi Hoon Yap. Automated analysis and quantification of human mobility using a depth sensor. *IEEE journal of biomedical and health informatics*, 21(4):939–948, 2017.
81. Daniel Leightley, Subhas C Mukhopadhyay, Hemant Ghayvat, and Moi Hoon Yap. Deep convolutional neural networks for motion instability identification using kinect. In *Machine Vision Applications (MVA), 2017 Fifteenth IAPR International Conference on*, pages 310–313. IEEE, 2017.
82. Daniel Leightley, Moi Hoon Yap, Jessica Coulson, Yoann Barnouin, and Jamie S McPhee. Benchmarking human motion analysis using kinect one: An open source dataset. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2015 Asia-Pacific*, pages 1–7. IEEE, 2015.

83. Adrian Leu, D Ristić-Durrant, and Axel Gräser. A robust markerless vision-based human gait analysis system. In *Applied Computational Intelligence and Informatics (SACI), 2011 6th IEEE International Symposium on*, pages 415–420. IEEE, 2011.

84. Longze Li and Aleksandar Vakanski. Generative adversarial networks for generation and classification of physical rehabilitation movement episodes. *International journal of machine learning and computing*, 8(5):428, 2018.

85. Michael H Li, Tiago A Mestre, Susan H Fox, and Babak Taati. Vision-based assessment of parkinsonism and levodopa-induced dyskinesia with pose estimation. *Journal of neuroengineering and rehabilitation*, 15(1):97, 2018.

86. Wanqing Li, Zhengyou Zhang, and Zicheng Liu. Action recognition based on a bag of 3d points. In *CVPR Workshops (CVPRW)*, pages 9–14. IEEE, 2010.

87. Y Liao, A Vakanski, and M Xian. A deep learning framework for assessment of quality of rehabilitation exercises. *arXiv preprint arXiv:1901.10435*, 2019.

88. Ting Yang Lin, Chung Hung Hsieh, and Jiann Der Lee. A kinect-based system for physical rehabilitation: Utilizing Tai Chi exercises to improve movement disorders in patients with balance ability. In *Proceedings - Asia Modelling Symposium 2013: 7th Asia International Conference on Mathematical Modelling and Computer Simulation, AMS 2013*, pages 149–153. IEEE, jul 2013.

89. Ting-Yang Lin, Chung-Hung Hsieh, and Jiann-Der Lee. A kinect-based system for physical rehabilitation: Utilizing tai chi exercises to improve movement disorders in patients with balance ability. In *Modelling Symposium (AMS), 2013 7th Asia*, pages 149–153. IEEE, 2013.

90. Tsun Te Liu, Ching Tang Hsieh, Ruei Chi Chung, and Yuan Sheng Wang. Physical rehabilitation assistant system based on kinect. In *Applied Mechanics and Materials*, volume 284, pages 1686–1690. Trans Tech Publ, 2013.

91. Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. pages 674–679, 1981.

92. Osama Makansi, Eddy Ilg, Ozgun Cicek, and Thomas Brox. Overcoming limitations of mixture density networks: A sampling and fitting framework for multimodal future prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7144–7153, 2019.

93. Giovanni Mariani, Florian Scheidegger, Roxana Istrate, Costas Bekas, and Cristiano Malossi. Bagan: Data augmentation with balancing gan. *arXiv preprint arXiv:1803.09655*, 2018.

94. Kenji Matsuo, Kentaro Yamada, Satoshi Ueno, and Sei Naito. An attention-based activity recognition for egocentric video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 551–556, 2014.

95. Cheryl D. Metcalf, Rebecca Robinson, Adam J. Malpass, Tristan P. Bogle, Thomas A. Dell, Chris Harris, and Sara H. Demain. Markerless motion capture and measurement of hand kinematics: Validation and application to home-based upper limb rehabilitation. *IEEE Transactions on Biomedical Engineering*, 60(8):2184–2192, aug 2013.

96. Hossein Mousavi Hondori and Maryam Khademi. A review on technical and clinical impact of microsoft kinect on physical therapy and rehabilitation. *Journal of medical engineering*, 2014, 2014.

97. Meinard Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

98. Tushar Nagarajan, Christoph Feichtenhofer, and Kristen Grauman. Grounded human-object interaction hotspots from video. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 8688–8697, 2019.

99. Mukund Narasimhan, Paul Viola, and Michael Shilman. Online decoding of markov models under latency constraints. In *Proceedings of the 23rd international conference on Machine learning*, pages 657–664. ACM, 2006.

100. Saravana K Natarajan, Xingchen Wang, Matthias Spranger, and Axel Gräser. Reha@ home-a vision based markerless gait analysis system for rehabilitation at home. In *Biomedical Engineering (BioMed), 2017 13th IASTED International Conference on*, pages 32–41. IEEE, 2017.

101. S Obdržálek, Gregorij Kurillo, Jay Han, Ted Abresch, Ruzena Bajcsy, et al. Real-time human pose detection and tracking for tele-rehabilitation in virtual reality. *Studies in health technology and informatics*, 173:320–324, 2012.

102. Erienne V Olesh, Sergiy Yakovenko, and Valeriya Gritsenko. Automated assessment of upper extremity movement impairment due to stroke. *PloS one*, 9(8):e104487, 2014.
103. Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease. The unified parkinson's disease rating scale (updrs): status and recommendations. *Movement Disorders*, 18(7):738–750, 2003.
104. Adeline Paiement, Lili Tao, Sion Hannuna, Massimo Camplani, Dima Damen, and Majid Mirmehdi. Online quality assessment of human movement from skeleton data. In *BMVA*, pages 153–166. BMVA press, 2014.
105. Carlos Palma, Augusto Salazar, and Francisco Vargas. Hmm and dtw for evaluation of therapeutical gestures using kinect. *arXiv preprint arXiv:1602.03742*, 2016.
106. Ingrid Parry, Clarissa Carbullido, Jason Kawada, Anita Bagley, Soman Sen, David Greenhalgh, and Tina Palmieri. Keeping up with video game technology: Objective analysis of xbox kinect™ and playstation 3 move™ for use in burn rehabilitation. *Burns*, 40(5):852–859, 2014.
107. Georgios Pavlakos, Xiaowei Zhou, and Kostas Daniilidis. Ordinal depth supervision for 3d human pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7307–7316, 2018.
108. Wei Pei, Guanghua Xu, Min Li, Hui Ding, Sicong Zhang, and Ailing Luo. A motion rehabilitation self-training and evaluation system using kinect. In *Ubiquitous Robots and Ambient Intelligence (URAI), 2016 13th International Conference on*, pages 353–357. IEEE, 2016.
109. Xiaojiang Peng, Limin Wang, Xingxing Wang, and Yu Qiao. Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice. *CVIU*, 150:109–125, 2016.
110. Hieu Pham, Melody Guan, Barret Zoph, Quoc Le, and Jeff Dean. Efficient neural architecture search via parameters sharing. In *International Conference on Machine Learning*, pages 4095–4104. PMLR, 2018.
111. Thomas Pintaric and Hannes Kaufmann. Affordable infrared-optical pose-tracking for virtual and augmented reality. In *Proceedings of Trends and Issues in Tracking for Virtual Environments Workshop, IEEE VR*, pages 44–51, 2007.
112. Loreen Pogrzeba, Thomas Neumann, Markus Wacker, and Bernhard Jung. Analysis and quantification of repetitive motion in long-term rehabilitation. *IEEE journal of biomedical and health informatics*, 23(3):1075–1085, 2018.
113. Ronald Poppe. A survey on vision-based human action recognition. *Image and vision computing*, 28(6):976–990, 2010.
114. Julia Richter, Christian Wiede, Lars Lehmann, and Gangolf Hirtz. Motion evaluation by means of joint filtering for assisted physical therapy. In *Consumer Electronics-Berlin (ICCE-Berlin), 2017 IEEE 7th International Conference on*, pages 10–14. IEEE, 2017.
115. Julia Richter, Christian Wiede, Bharat Shinde, and Gangolf Hirtz. Motion error classification for assisted physical therapy - a novel approach using incremental dynamic time warping and normalised hierarchical skeleton joint data. In *ICPRAM 2017*, 2017.
116. Jesús Joel Rivas, Felipe Orihuela-Espina, Lorena Palafox, Nadia Berthouze, María del Carmen Lara, Jorge Hernández-Franco, and Enrique Sucar. Unobtrusive inference of affective states in virtual rehabilitation from upper limb motions: A feasibility study. *IEEE Transactions on Affective Computing*, 2018.
117. Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules. In *Advances in neural information processing systems*, pages 3856–3866, 2017.
118. Nikolaos Sarafianos, Bogdan Boteanu, Bogdan Ionescu, and Ioannis A Kakadiaris. 3d human pose estimation: A review of the literature and analysis of covariates. *CVIU*, 152:1–20, 2016.
119. Supriya Sathyanarayana, Ravi Kumar Satzoda, Suchitra Sathyanarayana, and Srikanthan Thambipillai. Vision-based patient monitoring: a comprehensive review of algorithms and technologies. *Journal of Ambient Intelligence and Humanized Computing*, 9(2):225–251, Apr 2018.
120. Reinhold Scherer, Johanna Wagner, Günter Moitzi, and Gernot Müller-Putz. Kinect-based detection of self-paced hand movements: enhancing functional brain mapping paradigms. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 4748–4751. IEEE, 2012.

121. Santiago Schez-Sobrino, Dorothy N Monekosso, Paolo Remagnino, David Vallejo, and Carlos 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)*, pages 1–6. IEEE, 2019.

122. Christian Schönauer, Thomas Pintaric, Hannes Kaufmann, Stephanie Jansen-Kosterink, and Miriam Vollenbroek-Hutten. Chronic pain rehabilitation with a serious game using multimodal input. In *Virtual Rehabilitation (ICVR), 2011 International Conference on*, pages 1–8. IEEE, 2011.

123. Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. Ntu rgb+ d: A large scale dataset for 3d human activity analysis. In *CVPR*, pages 1010–1019, 2016.

124. Lei Shi, Yifan Zhang, Jian Cheng, and Hanqing Lu. Skeleton-based action recognition with directed graph neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7912–7921, 2019.

125. Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. Real-time human pose recognition in parts from single depth images. In *CVPR 2011*, pages 1297–1304. Ieee, 2011.

126. Bilge Soran, Linda Lowes, and Katherine M Steele. Evaluation of infants with spinal muscular atrophy type-i using convolutional neural networks. In *ECCV*, pages 495–507. Springer, 2016.

127. Sofija Spasojević, Tihomir V Ilić, Slađan Milanović, Veljko Potkonjak, Aleksandar Rodić, and José Santos-Victor. Combined vision and wearable sensors-based system for movement analysis in rehabilitation. *Methods of information in medicine*, 56(02):95–111, 2017.

128. Sofija Spasojević, José Santos-Victor, Tihomir Ilić, Slađan Milanović, Veljko Potkonjak, and Aleksandar Rodić. A vision-based system for movement analysis in medical applications: the example of parkinson disease. In *International Conference on Computer Vision Systems*, pages 424–434. Springer, 2015.

129. Erik E Stone and Marjorie Skubic. Capturing habitual, in-home gait parameter trends using an inexpensive depth camera. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 5106–5109. IEEE, 2012.

130. Stroke Association UK. State of the nation: Stroke statistics. *https://www.stroke.org.uk/*, 2017.

131. Chuan-Jun Su, Chang-Yu Chiang, and Jing-Yan Huang. Kinect-enabled home-based rehabilitation system using dynamic time warping and fuzzy logic. *Applied Soft Computing*, 22:652–666, 2014.

132. L Enrique Sucar, Gildardo Azcárate, Ron S Leder, David Reinkensmeyer, Jorge Hernández, Israel Sanchez, and Pedro Saucedo. Gesture therapy: A vision-based system for arm rehabilitation after stroke. In *International Joint Conference on Biomedical Engineering Systems and Technologies*, pages 531–540. Springer, 2008.

133. L Enrique Sucar, Roger Luis, Ron Leder, Jorge Hernández, and Israel Sánchez. Gesture therapy: A vision-based system for upper extremity stroke rehabilitation. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 3690–3693. IEEE, 2010.

134. Evan A Suma, Belinda Lange, Albert Skip Rizzo, David M Krum, and Mark Bolas. Faast: The flexible action and articulated skeleton toolkit. In *Virtual Reality Conference (VR), 2011 IEEE*, pages 247–248. IEEE, 2011.

135. Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1199–1208, 2018.

136. Babak Taati, Rosalie Wang, Rajibul Huq, Jasper Snoek, and Alex Mihailidis. Vision-based posture assessment to detect and categorize compensation during robotic rehabilitation therapy. In *Biomedical Robotics and Biomechatronics (BioRob), 2012 4th IEEE RAS & EMBS International Conference on*, pages 1607–1613. IEEE, 2012.

137. Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019.

138. Lili Tao, Adeline Paiement, Dima Damen, Majid Mirmehdi, Sion Hannuna, Massimo Camplani, Tilo Burghardt, and Ian Craddock. A comparative study of pose representation and dynamics modelling for online motion quality assessment. *CVIU*, 148:136–152, 2016.

139. Yaqin Tao and Huosheng Hu. Colour based human motion tracking for home-based rehabilitation. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 1, pages 773–778. IEEE, 2004.

140. Gineke A ten Holt, Marcel JT Reinders, and EA Hendriks. Multi-dimensional dynamic time warping for gesture recognition. In *Thirteenth annual conference of the Advanced School for Computing and Imaging*, volume 300, page 1, 2007.

141. Paolo Tormene, Toni Giorgino, Silvana Quaglini, and Mario Stefanelli. Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation. *Artificial intelligence in medicine*, 45(1):11–34, 2009.

142. Alexander Toshev and Christian Szegedy. Deeppose: Human pose estimation via deep neural networks. In *CVPR*, pages 1653–1660, 2014.

143. Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *ICCV*, pages 4489–4497, 2015.

144. A Vakanski, JM Ferguson, and S Lee. Mathematical modeling and evaluation of human motions in physical therapy using mixture density neural networks. *Journal of physiotherapy & physical rehabilitation*, 1(4), 2016.

145. Aleksandar Vakanski, Hyung-pil Jun, David Paul, and Russell Baker. A data set of human body movements for physical rehabilitation exercises. *Data*, 3(1):2, 2018.

146. Janani Venugopalan, Chihwen Cheng, Todd H Stokes, and May D Wang. Kinect-based rehabilitation system for patients with traumatic brain injury. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 4625–4628. IEEE, 2013.

147. Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

148. Heng Wang, Alexander Kläser, Cordelia Schmid, and Cheng-Lin Liu. Dense trajectories and motion boundary descriptors for action recognition. *International journal of computer vision*, 103(1):60–79, 2013.

149. Jingli Wang, Lei Yu, Jiping Wang, Liquan Guo, Xudong Gu, and Qiang Fang. Automated fugl-meyer assessment using svr model. In *2014 IEEE International Symposium on Bioelectronics and Bioinformatics (IEEE ISBB 2014)*, pages 1–4. IEEE, 2014.

150. Ruizhe Wang, Gérard Medioni, Carolee Winstein, and Cesar Blanco. Home monitoring musculo-skeletal disorders with a single 3d sensor. In *CVPR Workshops*, pages 521–528, 2013.

151. Tian Wang, Kyunghyun Cho, and Musen Wen. Attention-based mixture density recurrent networks for history-based recommendation. In *Proceedings of the 1st International Workshop on Deep Learning Practice for High-Dimensional Sparse Data*, pages 1–9, 2019.

152. David Webster and Ozkan Celik. Systematic review of kinect applications in elderly care and stroke rehabilitation. *Journal of neuroengineering and rehabilitation*, 11(1):108, 2014.

153. Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional pose machines. In *CVPR*, pages 4724–4732, 2016.

154. Steven L Wolf, Pamela A Catlin, Michael Ellis, Audrey Link Archer, Bryn Morgan, and Aimee Piacentino. Assessing wolf motor function test as outcome measure for research in patients after stroke. *Stroke*, 32(7):1635–1639, 2001.

155. Jiajun Wu, Tianfan Xue, Joseph J Lim, Yuandong Tian, Joshua B Tenenbaum, Antonio Torralba, and William T Freeman. Single image 3d interpreter network. In *European Conference on Computer Vision*, pages 365–382. Springer, 2016.

156. Lu Xia, Chia-Chih Chen, and Jake K Aggarwal. View invariant human action recognition using histograms of 3d joints. In *CVPR Workshops*, pages 20–27. IEEE, 2012.

157. Wei Yang, Wanli Ouyang, Xiaolong Wang, Jimmy Ren, Hongsheng Li, and Xiaogang Wang. 3d human pose estimation in the wild by adversarial learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5255–5264, 2018.

158. Qi Ye and Tae-Kyun Kim. Occlusion-aware hand pose estimation using hierarchical mixture density network. In *ECCV*, pages 801–817, 2018.
159. José Zariffa and John D Steeves. Computer vision-based classification of hand grip variations in neurorehabilitation. In *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*, pages 1–4. IEEE, 2011.
160. Hong-Bo Zhang, Qing Lei, Bi-Neng Zhong, Ji-Xiang Du, and JiaLin Peng. A survey on human pose estimation. *Intelligent Automation & Soft Computing*, 22(3):483–489, 2016.
161. Zhengyou Zhang. Microsoft kinect sensor and its effect. *IEEE multimedia*, 19(2):4–10, 2012.
162. Fei Zheng and Geoffrey I. Webb. A comparative study of semi-naive bayes methods in classification learning. 2005.
163. Ying Xuan Zhi, Michelle Lukasik, Michael H Li, Elham Dolatabadi, Rosalie H Wang, and Babak Taati. Automatic detection of compensation during robotic stroke rehabilitation therapy. *IEEE journal of translational engineering in health and medicine*, 6:1–7, 2018.
164. Huiyu Zhou and Huosheng Hu. Human motion tracking for rehabilitation—a survey. *Biomedical Signal Processing and Control*, 3(1):1–18, 2008.