Pattern Recognition 119 (2021) 108095

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Skeleton-based human action evaluation using graph convolutional network for monitoring Alzheimer's progression

Bruce X.B. Yu[a], Yan Liu[a], Keith C.C. Chan[a], Qintai Yang[b], Xiaoying Wang[b,*]

[a] Department of Computing, The Hong Kong Polytechnic University, Hong Kong
[b] Third Affiliated Hospital of Sun Yat-sen University, China

ARTICLE INFO

Article history:
Received 15 August 2020
Revised 5 May 2021
Accepted 3 June 2021
Available online 9 June 2021

Keywords:
Human action evaluation
Alzheimer's disease
Graph neural network
Abnormality detection

ABSTRACT

Human action evaluation (HAE) involves judgments about the abnormality and quality of human actions. If performed effectively, HAE based on skeleton data can be used to monitor the outcomes of behavioral therapies for Alzheimer's disease (AD). In this paper, we propose a two-task graph convolutional network (2T-GCN) to represent skeleton data for HAE tasks involving abnormality detection and quality evaluation. The network is first evaluated using the UI-PRMD dataset and demonstrates accurate abnormality detection. Regarding quality evaluation, in addition to laboratory-collected UI-PRMD data, we test the network on a set of real exercise data collected from patients with AD. A numerical score indicating the degree to which actions deviate from normal is taken to reflect the severity of AD; thus, we apply 2T-GCN to determine such scores. Experimental results show that numerical scores for certain exercises performed by patients with AD are consistent with their AD severity level as identified by clinical staff. This corroboration highlights the potential of our approach for monitoring AD and other neurodegenerative diseases.

1. Introduction

Human action evaluation (HAE) involves the determination of computational models to automatically detect abnormalities and assess the quality of human motions performed for specific purposes. HAE differs from human action recognition (HAR), which focuses on classifying different actions [1]; however, the two methods share similar sensor technologies. HAE has many useful applications in areas such as physical rehabilitation, assisted living, skills training, and sports activity scoring [2]. Scholars have recently attempted to investigate skeleton-based HAE [3,4]. Current HAE approaches that use deep learning (DL) methods model action assessment as a regression problem based on the supervision of either arbitrary function scores [5] or subjective human labels [6]. However, it remains challenging for models to simultaneously handle HAE tasks of abnormality detection and quality assessment. To advance existing HAE methods, we propose a two-task graph convolutional network (2T-GCN) approach to tackle relevant tasks. Given a set of temporally related skeleton frames, we construct a skeleton graph to represent skeleton data. The abnormality detection task is modeled by training the proposed 2T-GCN as a binary classifier; for quality evaluation purposes, a numerical action evaluation score is then retrieved from the trained 2T-GCN model. The proposed method is tested on a public dataset called URI-PRMD [7]. Our approach achieves encouraging results and demonstrates the better HAE ability of Kinect v2.

Alzheimer's disease (AD) is especially prevalent compared with other non-communicable diseases that collectively account for nearly 70% of deaths worldwide [8,9]. To improve the health of the elderly, it has become increasingly important to develop effective AD treatments [10,11]. Although various biomarkers have been developed for early AD diagnosis, accurate diagnosis partially relies on clinical criteria that often require at least 6 months for symptoms to appear [12]. Even if symptoms manifest to the point that a diagnosis of AD can be confirmed, the disease is currently incurable. Per the Global Deterioration Scale [13], AD symptoms such as increased forgetfulness, decreased work performance, and higher frequency of getting lost become noticeable to family members during the mild cognitive impairment (MCI) stage. Because MCI represents a relatively long dysfunction process that can persist for roughly 7 years [14], clinicians have striven to develop criteria to facilitate early AD diagnosis and thus decelerate or even prevent disease progression [10–13]. For example, a growing body of literature suggests that systematic exercise training can improve pa-

* Corresponding author.
E-mail addresses: csyliu@comp.polyu.edu.hk (Y. Liu), wxying@mail.sysu.edu.cn (X. Wang).

https://doi.org/10.1016/j.patcog.2021.108095
0031-3203/

tients cardiovascular functioning; increase flexibility, balance, and strength; and prevent cognitive dysfunction [15–17].

However, clinical diagnosis and behavioral therapies for AD usually require a number of sessions, which can be unaffordable for patients families. Yet HAE tasks, if tackled effectively, can support the diagnosis and treatment of AD [11,18]. Given this observation, we note that most AD symptoms manifest in cognitive domains such as praxis, executive functions, language, complex visual processing, and gnosis. Praxis impairment can include impaired cognitive functioning involving gesture-based imitation, production, or recognition [19]. Therefore, in this study, we decided to conduct experiments to investigate the praxis domain of cognitive changes. Specifically, we employed a Kinect v2 sensor to monitor morning exercises among normal and AD subjects in an elderly home. To determine their praxis health conditions, we gathered and analyzed a dataset using the proposed 2T-GCN to generate numerical evaluation scores for the collected exercises. Results showed that the scores of many exercises corresponded with AD severity scores assigned by clinical staff, suggesting that our proposed method could benefit AD diagnosis and related behavioral interventions.

2. Related work

In this section, we review state-of-the-art research on skeleton-based HAE from the perspectives of skeleton retrieval, skeleton representation methods, and datasets and their evaluation criteria.

2.1. Skeleton retrieval

A summary of existing HAE methods in [2] indicated motion detection as the first step in data retrieval. Thus, we introduce skeleton detection methods in this section. Vision devices that support 2D/3D skeleton retrieval include three types: motion capture (Mocap) systems, depth cameras, and RGB cameras. Mocap system companies provide these systems in domains such as biomechanics, sports, engineering, and entertainment. Mocap systems can provide highly accurate skeleton data but carry high costs and low flexibility for commercialization purposes; off-the-shelf commercial depth cameras such as Kinect and Intel RealSense can retrieve skeleton data more affordably than Mocap systems. RGB cameras can retrieve 2D skeleton [20] or 3D skeleton [21,22] data but come with higher computational costs. Kinect sensors, which can also efficiently retrieve skeleton data, are often used in relevant research. A list of public benchmark datasets identified in a survey of human body skeleton representation [23] revealed that 29 out of 41 datasets were collected with Kinect; Mocap systems and RGB cameras were the second and least popular approaches, respectively. The accuracy of Kinect v1 when measuring movements in people with Parkinson's disease was evaluated in [24]. Results showed that the device could accurately measure the timing and gross spatial characteristics of clinically relevant movements.

Given the popularity and affordability of Kinect sensors, we used Kinect v2 to collect a dataset in this study. To further validate the sensor's suitability, we performed extensive experiments on UI-PRMD [7] using Kinect v2 and Vicon Mocap.

2.2. Skeleton representation

Effective skeleton-based HAE methods rely on proper representations of skeleton data. Available skeleton representations can be classified into two main approaches: handcrafted feature representations and deep feature representations [2]. Handcrafted feature representations rely on constructing effective geometric features from skeleton data. Based on geometric features, traditional algorithms such as the hidden Markov model (HMM) [6], support vector machine [25], and k-nearest neighbor [26] approaches are commonly used for HAE. The training process in [6] was supervised by the abnormality degree (on a scale of 1 to 5) as evaluated by a professional physiatrist. Various HMM models were compared in [27], and DL models have outperformed them when applied to various datasets [28]. Common DL models such as convolutional neural networks and long short-term memory (LSTM) were adopted in [29] for gesture correctness estimation. Additionally, a DL framework using LSTM was proposed to encode skeleton data from the UI-PRMD dataset, which was supervised by a quality score function [5]. Advanced LSTM representation models, such as ST-LSTM [30,31], were initially proposed for skeleton-based HAR tasks. More recently, GCN models [32–34] have demonstrated encouraging performance in HAR tasks; however, these models [32–34] have seldom been applied to HAE. In this paper, we adopt a basic GCN model in [32] to represent skeleton data and adapt the model for two HAE tasks (i.e., inferring the abnormality and quality of an action). As far as we know, our work represents a pioneering attempt to use a GCN model with skeleton-based HAE.

In terms of HAE, existing methods are either supervised by human labels [6] or an arbitrary score function [5], which might not satisfy the requirements for AD diagnosis. On one hand, training a regression model based on a subjective clinical label alone can make it difficult to evaluate symptom severity in AD patients. This approach is therefore not adopted in our study. On the other hand, supervising the training process with a score function leads to redundancy: the results could already be obtained through the evaluation function. Our method differs from available approaches. First, we use a graph representation to train a GCN model supervised with binary labels (i.e., normal or abnormal). We then retrieve the evaluation score from the probability distribution before the model's SoftMax classifier, as the evaluation score contains more information than the classification result [35].

2.3. Datasets and evaluation

Few studies have reviewed HAE benchmarks. Several healthcare-related HAE datasets were outlined in [36], but most surveyed datasets focused on action recognition rather than evaluation. To our best knowledge, no research has yet investigated the standard evaluation method of HAE algorithms. Given the popularity of Kinect sensors for human action analysis, we have examined the evaluation methods of representative benchmarks that used Kinect to retrieve skeleton data (see Table 1). In [37], Kinect v1 was adopted to monitor several psychomotor exercises (e.g., "touch the right eye with the right hand", "touch the left eye with the right hand", and "raise the right hand"). The system presented in [37] could detect 14 psychomotor exercises but did not involve action assessment. The SPHERE dataset [27] included three sub-datasets Staircase2014, Walking2015, and SitStand2015 but simply provided data on the center of the body instead of all skeleton joints. The UI-PRMD dataset [7] represents a fitness dataset collected for HAE algorithm evaluation. Because this dataset [7] did not provide a standard evaluation method, the methods proposed in [5] and [38] could not be properly compared. Furthermore, different from the 10 incorrect exercises in UI-PRMD that are simulated by subjects who perform the other correct motion sequences, exercises in the AHA-3D dataset [39] are performed by older and younger people. However, as of this writing, the AHA-3D dataset is not publicly accessible. The evaluation method in [39] presented a per-frame view rather than assessing the entire action sequence. EJMQA [6] includes four simple actions and is similar to SPHERE [27].

Unlike existing datasets that were performed by young subjects, we collected an elderly home exercise (EHE) dataset from subjects with AD.

Table 1
Benchmark human activity evaluation datasets.

Dataset	Year	Sensor	Disease	NS	NA
Ortega et al. [37]	2014	Kinect v1	CD	15	14
SPHERE [27]	2016	Kinect v2	Stroke, PD	12/10/10	3
UI-PRMD [7]	2018	Kinect v2, Vicon	Rehabilitation	20	10
AHA-3D [39]	2018	Kinect v2	Fitness	23	4
EJMQA [6]	2020	Kinect v2	NMD	32	4
Our EHE	2020	Kinect v2	AD	25	6

Note: NS, number of subjects; NA, number of actions; CD, cognitive deterioration; PD, Parkinson's disease; NMD, neuromuscular disorder; AD, Alzheimer's disease.

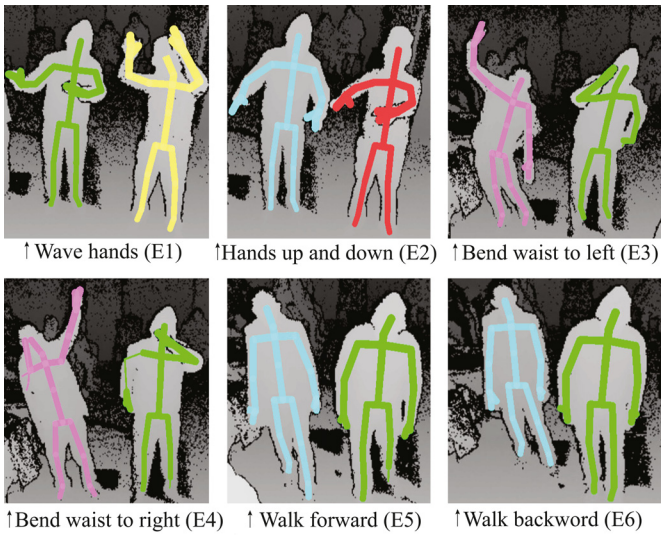

Fig. 1. Examples of morning exercises in the elderly home.

3. Elderly home exercise dataset

In this study, we provide valuable insight for the HAE field by collecting an EHE dataset from an elderly home based on real elderly patients with AD. Collecting routine exercise data is time-consuming and accompanied by intensive labeling costs. Moreover, as HAE has rarely been applied to real-world disease monitoring, we chose to gather sample data to validate effective HAE methods that could then be expanded on a larger validation scale. Our dataset consists of several actions from morning exercises that patients complete daily in the elderly home. In this section, we introduce our data collection method along with the structure of our EHE dataset. In the selected elderly home, residents complete daily morning exercises led by clinical staff. The staff member demonstrates physical exercises to a group of residents who follow the model to imitate the exercises. We did not provide novel exercises for the elderly residents, as doing so would have disrupted their daily schedule and presented challenges for clinical staff and residents. To maintain a natural setting in the elderly home, all exercises in this study were identical to those that residents normally complete each day. Similar to sensor usage in prior research, six exercises (see Fig. 1) were collected for our EHE dataset via Kinect v2.

The demographic information for 25 subjects who participated in data collection is listed in Table 2. Subjects average age was 68.4 years with a standard deviation of 10.82 years. Our dataset therefore differed from existing datasets performed by young subjects. The relatively younger residents in our dataset, such as Subjects 16, 18, and 20, were caregivers in the elderly home. Ten subjects had been diagnosed with AD of varying severity as indicated in the "AD" column of Table 2. Clinical staff assessed subjects severity level based on a scale of 0 to 10, where 0 represents no AD and 10 represents the last stage of AD. Table 2 also indicates the number of exercise repetitions retrieved from raw Kinect v2 data. Some exercises had 0 repetitions due to failed skeleton detection.

Table 2
Demographic information and number of repetitions per exercise in EHE dataset.

Subject ID	Demographic Information					No. of Repetitions					
	A	G	W	H	AD	E1	E2	E3	E4	E5	E6
1	76	M	70	168	0	7	5	4	9	2	3
2	72	M	66	180	7	7	5	6	6	3	3
3	60	M	54	172	8	6	5	6	8	3	3
4	68	M	60	160	0	6	5	5	6	2	2
5	62	M	70	165	5	0	5	8	6	2	2
6	72	F	51	157	10	7	5	8	7	1	1
7	68	F	54	158	0	7	7	6	1	4	3
8	92	M	55	165	0	8	6	6	3	3	3
9	86	F	55	163	0	11	5	6	4	2	2
10	54	M	60	162	10	8	5	6	7	2	3
11	67	M	85	185	5	9	6	10	9	3	3
12	83	M	65	170	6	9	5	12	9	3	2
13	81	F	48	151	4	9	6	13	0	3	3
14	64	M	65	172	8	7	6	6	5	3	2
15	67	M	70	170	6	8	6	6	6	3	2
16	57	F	69	156	0	8	6	7	6	4	4
17	70	M	94	182	0	8	6	14	3	4	4
18	56	F	64	158	0	7	6	12	4	4	3
19	84	M	61	175	0	8	5	15	2	4	4
20	55	F	55	160	0	8	6	13	13	4	4
21	77	F	47	161	0	8	7	14	15	3	3
22	60	F	58	163	0	11	7	13	15	3	3
23	55	F	58	162	0	9	7	0	12	3	3
24	66	F	65	163	0	11	6	0	15	3	3
25	58	F	66	161	0	11	6	0	15	3	3

Note: A, age; G, gender; W, weight in kilograms; H, height in centimeters *(weight and height are rounded to the nearest integer)*; AD, severity of Alzheimer's disease on a scale of 1 to 10; M, male; F, female; E1 to E6, Exercise 1 to Exercise 6.

4. Our 2T-GCN method

In this section, we introduce the skeleton data notation and convolutional operations proposed in our 2T-GCN method for the HAE tasks of abnormality detection and exercise quality evaluation.

4.1. Data structure and notation

We denote N repetitions of an exercise performed by all subjects as $S = \left\{S^{(i)} \mid i = 1, ..., N\right\}$, where $S^{(i)}$ is a sequence of skeleton frames that characterize the exercise. For a particular exercise, we recorded a sequence of skeleton frames corresponding to that exercise. As Fig. 2 depicts, each skeleton frame consisted of a set of skeleton joints labeled as HEAD, NECK, FOOTLEFT, and so on. A skeleton frame with J skeleton joints observed at time t is denoted as $S_t^{(i)} = \left(S_{t1}^{(i)}, ..., S_{tj}^{(i)}, ..., S_{tJ}^{(i)}\right)$, where $S_{tj}^{(i)}$ contains eight attributes corresponding to a joint's position and orientation features. The position of joint $S_{tj}^{(i)}$ has four attribute features, including 3-D

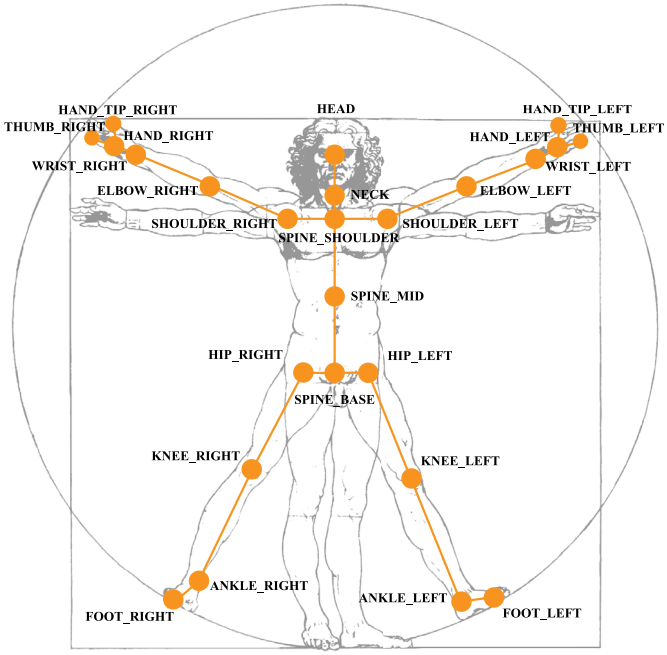

Fig. 2. Skeleton joints of Kinect v2 sensor.

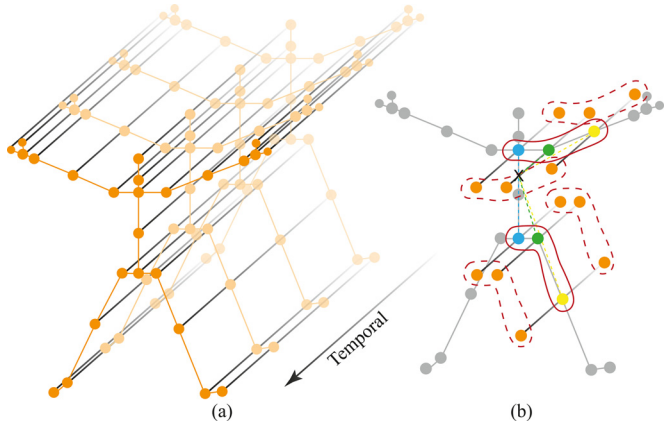

Fig. 3. (a). Illustration of the constructed skeleton graph. (b). Illustration of the convolutional sampling strategy for green nodes. Different colors denote different subsets. × represents the skeleton's center of gravity.

Cartesian coordinates (x, y, z) and its height h from the floor. Comparatively, the orientation of joint $S_{tj}^{(i)}$ is represented by a quaternion that has a set of values (X, Y, Z, W), where (X, Y, Z) can be transformed into angular orientation values $(yaw, roll, pitch)$ and W is used to calculate the ground plain of the environment. Hence, we can use $S_{tj}^{(i)} = (x, y, z, h, X, Y, Z, W)$ to denote all attributes of a skeleton joint. We considered multiple attribute combinations in our experiments.

4.2. Graph convolutional network

4.2.1. Graph construction

Each skeleton frame includes a number of skeleton joints. Raw skeleton joint data were streamed as an ordered list of vectors. Each vector included the position and orientation features of the corresponding joint. We propose adopting a GCN model to represent structured information among these joints along spatial and temporal dimensions. Figure 3(a) illustrates the constructed skeleton graph. On the spatial dimension, skeleton body joints are represented as vertices; their natural connections, indicating bones, are spatially connected as edges (the orange lines in Fig. 3[a]). On the temporal dimension, the corresponding joints between two adjacent frames are temporally connected by edges (the black lines in Fig. 3[a]). To symbolize the skeleton graph at time t, we denote it as $\vartheta_t = \{\upsilon_t, \varepsilon_t\}$, where υ_t represents the skeleton joints and ε_t represents the skeleton bones, respectively. In this skeleton graph, we denote all joints of a complete exercise repetition as a node set $\upsilon_t = \left\{\upsilon_{tj} | \upsilon_{tj} = S_{tj}^{(i)}, j = 1, \ldots, J\right\}$. The attributes of a vertex on the graph are the corresponding position and orientation features of a skeleton joint.

4.2.2. Graph convolutional operation

Similar to 2D convolution, the convolutional operation of GCN requires a sampling area. As in [32], the sampling area of a node υ_{tj} is defined as a spatial and temporal neighbor set. Figure 3(b) illustrates this strategy, where the spatial sampling area $B_S(\upsilon_{tj})$ is enclosed by the red curve. Empirically, as Fig. 3(b) shows, the sampling strategy uses 3 spatial subsets: the vertex itself; the centripetal subset that contains neighboring vertices closer to the center of gravity; and the centrifugal subset that contains neighboring vertices farther from the center of gravity, respectively illustrated by green, blue, and yellow circles. Suppose the whole sampling area $B_S(\upsilon_{tj})$ has a fixed number of L subsets; every neighbor set could thus be numerically labeled with a mapping $h_{tj} : B_S(\upsilon_{tj}) \rightarrow \{0, \ldots, L-1\}$. Then the graph convolution could be computed as

$$f_{out}(\upsilon_{tj}) = \sum_{\upsilon_{tk} \in B_S \upsilon_{tj}} \frac{1}{Z_{tj}(\upsilon_{tk})} f_{in}(\upsilon_{tk}) w(h_{tj}(\upsilon_{tk})) \quad (1)$$

where $f_{in} : \upsilon_{tk} \rightarrow R^c$ is the mapping that obtains the attribute features of joint node υ_{tk}, $Z_{tj}(\upsilon_{tk}) = |\{\upsilon_{tm} | h_{tj}(\upsilon_{tm}) = h_{tj}(\upsilon_{tk})\}|$ is a normalization term equal to the cardinality of the corresponding subset. $w(h_{tj}(\upsilon_{tk}))$ is a weight function $w(\upsilon_{tj}, \upsilon_{tk}) : B_S(\upsilon_{tj}) \rightarrow R^c$ implemented by indexing a tensor with a size of (c, L).

Temporally, the neighborhood concept is extended to sequentially connected joints as $B_T(\upsilon_{tj}) = \left\{\upsilon_{qj} | |q - t| \leq \Gamma/2\right\}$, where Γ is the temporal kernel size that controls the temporal range of the neighbor set. The dotted red curve Fig. 3(b) indicates the temporal sampling area expanded from the spatial sampling area.

4.2.3. GCN block

The implementation of graph-based convolution is not as straightforward as 2D convolution and relies instead on a defined sampling strategy. The network input can be represented as a $C \times T \times J$ tensor, where C is the number of attributes on a joint vertex. The determined spatial sampling strategy can be represented by a $J \times J$ adjacency matrix $\mathbf{A}$ whose elements reflect whether a vertex υ_{tj} is included in the neighbor set. With L spatial sampling strategies $\sum_{k=1}^{L} \mathbf{A}_l$, a spatial GCN (Convs) layer is implemented by multiplying the resulting tensor of a $L \times 1$ classical 2D convolution with the normalized adjacency matrix $\mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{A} \mathbf{\Lambda}^{-\frac{1}{2}}$. Hence, Eq. (1) can be transformed into

$$f_{out}(S^{(i)}) = \sum_{l=1}^{L} \mathbf{\Lambda}_l^{-\frac{1}{2}} \mathbf{A}_l \mathbf{\Lambda}_l^{-\frac{1}{2}} f_{in} \mathbf{W}_l \odot \mathbf{M}_l \quad (2)$$

where $\mathbf{\Lambda}_l^{jj} = \sum_k^J (\mathbf{A}_l^{jk}) + \alpha$ is a diagonal matrix with α set to 0.001 to prevent empty rows. $\mathbf{W}_l$ denotes the weighting function of Eq. (1), which is a weight tensor of the 1×1 convolutional operation. $\mathbf{M}_l$ represents an attention map indicating the importance of graph nodes, which has the same size of $\mathbf{A}_l$. $\odot$ is the element-wise product between two matrices.

Given the Convs layer implemented above, a temporal GCN (Convt) layer can then be implemented via a 2D convolutional operation with its kernel size set to $\Gamma \times 1$. As Fig. 4 shows, a basic GCN block is stacked with a Convs layer, a Convt layer, and an

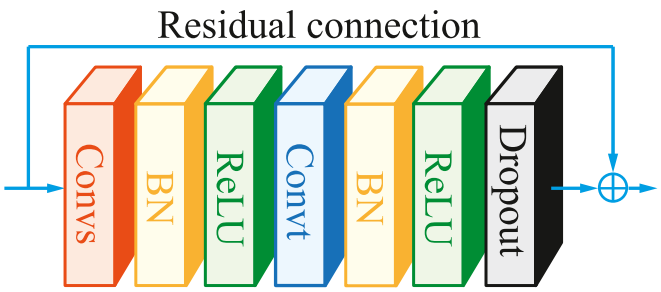

Fig. 4. Illustration of a GCN block. Convs represents the spatial GCN, and Convt represents the temporal GCN, both of which are followed by a BN layer and a ReLU layer. Moreover, a residual connection is added to the block.

additional dropout layer with the drop rate set to 0.5 to prevent overfitting. Convs and Convt are each followed by a batch normalization (BN) layer and a ReLU layer. To stabilize training, similar to Yan et al. [32], a residual connection is added for the GCN block.

4.3. Network architecture

The proposed 2T-GCN model is constructed from a stack of basic blocks. There are 9 GCN blocks in total as shown in the middle part of Fig. 5. The first three blocks, the middle three blocks, and the last three blocks have 64, 128, and 256 output channels, respectively. The temporal kernel size of these GCN blocks is set to 9. The strides of the 4th and 7th GCN blocks are both set to 2. To normalize the input data, a BN layer is added at the beginning. As a common practice in DL, we add a global average pooling layer at the end of the GCN blocks to transform the data into a vector with 256 dimensions. As the data are binarily labeled as either normal or abnormal, we use a 2D convolutional layer to transform the 256-dimensional feature vector into a 2-dimensional vector. Finally, we feed the 2-dimensional vector to a SoftMax classifier to infer the abnormality of a given exercise.

We adopt the probability results of the SoftMax layer to infer the evaluation score of an exercise repetition. More specifically, we retrieve the first dimension of the probability distribution from the SoftMax layer of the model, which can be calculated as

$$f_{score}(a,b) = \frac{e^a}{(e^a + e^b)} \tag{3}$$

where a and b respectively represent the first and second neuron output values of the fully connected layer that are used to calculate the probability distribution of the SoftMax layer.

This numerical score can indicate the exercise quality without the supervision of subjective human evaluation scores [6] or arbitrary scores calculated by a function as in [5]. In our experiments, we investigated whether this score could be consistent with the clinical evaluation.

4.4. Optimization

To learn the weights Θ of the 2T-GCN model G, we define the objective as a classification problem supervised by binary labels y. The binary labels for our EHE dataset are transformed from clinical evaluations with no AD set into 1, indicating no abnormality. Otherwise, the label is set to 0 to indicate abnormal action. By using the cross-entropy loss for abnormality detection, the objective can be formulated as

$$\arg\max_{\Theta} \frac{1}{N} \sum_{i=1}^{N} - \sum y^{(i)} log(\sigma(G(\theta, S^{(i)}))) \tag{4}$$

where $G(\theta, S^{(i)})$ is the defined graph convolutional operation as shown in the middle of Fig. 5. σ represents the SoftMax function that transfers the probability distribution results to the abnormality result.

5. Experiments

In this section, we introduce the experiments in which we applied our 2T-GCN method to two datasets UI-PRMD [7] and our EHE in terms of abnormality detection and exercise quality evaluation. We validated two perspectives on UI-PRMD. First, because we employed Kinect v2 to collect our EHE dataset, we referred to the UI-PRMD dataset [7] to validate the suitability of using Kinect v2 by comparing the device's output with that of the Vicon optical tracking system. Second, we validated our proposed method by comparing its representation ability with that proposed in [5] in terms of evaluation metrics. Regarding the EHE dataset, we validated the potential of our method for monitoring AD progression.

5.1. Datasets

5.1.1. UI-PRMD

The UI-PRMD dataset [7] consists of skeleton data gathered from 10 healthy subjects with every subject performing 10 repetitions of 10 rehabilitation exercises such as "deep squat", "hurdle step", and "sit to stand" in a correct and incorrect manner. Two sensors (i.e., Kinect v2 and the Vicon optical tracking system) were used to collect the dataset, with both sensors providing position and orientation features of skeleton data. We referred to the consistent version of the dataset processed by [5], in which inconsistent samples due to measurement error or subjects using the wrong limbs to perform an exercise were removed.

5.1.2. Our EHE dataset

As described in Table 2, our EHE dataset contained 869 action repetitions performed by 25 older people. One objective of this dataset involved investigating which skeleton features primarily contributed to abnormality detection. We used the eight features described in Section 4 in varying combinations to explore which features would be best for evaluating certain exercises. Another goal was to examine whether the action quality evaluation of our 2T-GCN method reflected the severity level of AD as evaluated by clinical staff. To do so, we adopted a cross-validation criterion to evaluate the HAE ability of our proposed method.

5.2. Implementation details

To train the proposed 2T-GCN model, we adopted the same experimental setting for the UI-PRMD and EHE datasets. Specifically, we used stochastic gradient descent to optimize our 2T-GCN model by setting the initial learning rate to 0.1. At epochs of 10, 50, and 100, we decayed the learning rate by multiplying it by 0.1. The training process was terminated once the model either achieved 100% accuracy or reached epoch 200. The batch was set to 16. All experiments were performed on a workstation with 2 GTX 1080 Ti GPUs.

5.3. Evaluation metrics

To test the representation ability of the GCN model, we referred to the concept of separation degree (SD) defined in [5]. For a pair of positive numbers x and y, the SD can be defined as $SD(x,y) = \frac{x-y}{x+y} \in [-1,1]$. Then the SD between two positive sequences $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ can be calculated as

$$SD(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} S_D(x_i, y_j) \tag{5}$$

Fig. 5. Illustration of the 2T-GCN model structured with 9 GCN blocks (B1'B9). The three numbers for each block represent the number of input channels, the number of output channels, and the stride, respectively. GAP represents the global average pooling layer. The model output can then be classified by a SoftMax classifier to infer abnormality; the results will also deliver a numerical evaluation score by utilizing the probability output of the SoftMax layer.

Table 3
Separation degree of exercises with different features and sensors in UI-PRMD.

Method	Exercise ID	Separation Degree (Std. Deviation)			
		Kinect v2		Vicon	
		3D Position	Angular	3D Position	Angular
Our Method	E1	0.908 (0.127)	0.977 (0.039)	0.468 (0.292)	0.986 (0.035)
	E2	0.942 (0.088)	0.992 (0.029)	0.875 (0.214)	0.713 (0.282)
	E3	0.821 (0.165)	0.930 (0.167)	0.591 (0.300)	0.945 (0.077)
	E4	0.712 (0.191)	0.909 (0.086)	0.009 (0.010)	0.699 (0.190)
	E5	0.691 (0.192)	0.953 (0.117)	0.547 (0.257)	0.955 (0.074)
	E6	0.961 (0.081)	0.902 (0.111)	0.785 (0.147)	0.832 (0.174)
	E7	0.873 (0.179)	0.867 (0.190)	0.867 (0.163)	0.975 (0.072)
	E8	0.890 (0.148)	0.899 (0.097)	0.657 (0.299)	0.750 (0.238)
	E9	0.785 (0.264)	0.937 (0.112)	0.301 (0.329)	0.913 (0.095)
	E10	0.267 (0.030)	0.953 (0.078)	0.422 (0.250)	0.968 (0.094)
Our Method	Average	0.793 (0.148)	**0.933** (0.100)	0.545 (0.244)	**0.878** (0.128)
MV [5]	Average	-	-	-	0.344 (0.049)
PCA [5]	Average	-	-	-	0.360 (0.060)
AE [5]	Average	-	-	-	0.515 (0.106)

To validate the HAE ability of our method in monitoring AD progression, we investigated whether the quality evaluation scores delivered by our model were consistent with clinical evaluation results. To do so, we considered the Euclidean distance (ED) and correlation (CORREL) between HAE results with clinical labels. For an n-dimensional space, the ED of two vectors $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ is calculated as $ED(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$. While the CORREL between $\boldsymbol{x}$ and $\boldsymbol{y}$ can be defined as $CORREL(\boldsymbol{x}, \boldsymbol{y}) = \frac{\sum_{i=1}^{n} (x-\bar{x})(y-\bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i-\bar{x})^2 \sum_{i=1}^{n} (y_i-\bar{y})^2}}$, where $\bar{x}$ and $\bar{y}$ are the average values of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. Smaller $ED(\boldsymbol{x}, \boldsymbol{y})$ and larger $CORREL(\boldsymbol{x}, \boldsymbol{y})$ indicate that the evaluation result is more consistent with the AD severity observation of a human expert and vice versa.

5.4. Results and analysis

5.4.1. UI-PRMD

Table 3 presents the SD results of our method and existing methods. Various feature extractors such as maximum variance (MV), principal component analysis (PCA), and autoencoder (AE) were attempted with a combination of a log-likelihood Gaussian mixture model(GMM) in [5]. Our method achieved an average SD of 0.878 by using the same experimental setting as [5], resulting in a substantial improvement over the best model (with an average SD of 0.515) based on a combination of log-likelihood GMM and AE in [5]. Using the angular feature of the Kinect v2 sensor, our method achieved an even higher average SD of 0.933; therefore, Kinect v2 appears more capable of HAE tasks, as it outperformed the Vicon optical tracking system in terms of 3D positions and angular features.

To further validate the proposed method, as displayed in Fig. 6, we visualized the exercise quality evaluation scores of "deep squat" and "hurdle step" in UI-PRMD by using the 3D position feature of Kinect v2. Correct and incorrect repetitions were clearly classified based on the exercise quality evaluation score calculated with the probability output of the SoftMax layer of the proposed 2T-GCN model with Eq. 4. According to the results in [5], the correct and incorrect pairs could not be clearly separated by their DL method as most of the incorrect repetitions get evaluation scores around 0.9 (given that 1 is the fully correct score), which needs to select a proper threshold to determine the correctness of an exercise repetition. With our method, the correctness of an exercise repetition is directly given by the result. Meanwhile, we could see that the scores of incorrect repetitions are below 0.5, whereas the correct repetitions have scores that are over 0.5. This indicates that our method achieves a clearer judgment on the correctness of an exercise repetition.

5.4.2. Elderly home exercise

Because Kinect v2 provides position and orientation attributes for skeleton joints, we assessed its capabilities when dealing with different attributes in various combinations. As Table 4 indicates, we conducted experiments on five feature combinations by comparing the training accuracy for different exercises. Our results show that some features were well suited to evaluating specific exercises. For example, "*xyz*" was appropriate for the first 4 exercises but inappropriate for Exercise 5 (E5). The average accuracy of all attribute combinations was greater than 90%, verifying the representation power of our 2T-GCN method. Because our model in-

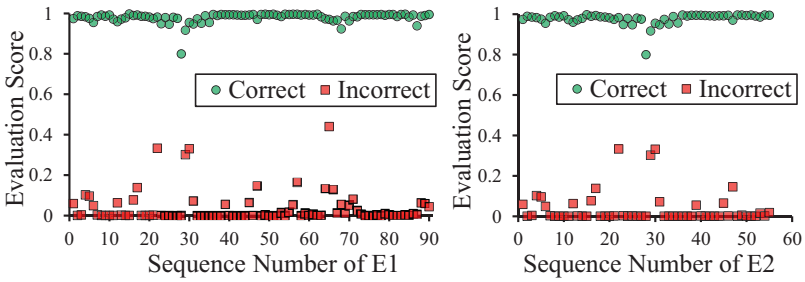

Fig. 6. Quality evaluation scores of E1 and E2 (i.e., "deep squat" and "hurdle step", respectively.

Table 4
Training accuracies of different attribute combinations on our EHE dataset.

Exercise ID	Training Accuracy (%)				
	angular	angw	xyz	xyzh	xyzhangw
E1	100.00	100.00	100.00	100.00	100.00
E2	99.31	100.00	100.00	100.00	100.00
E3	100.00	100.00	100.00	100.00	100.00
E4	100.00	100.00	100.00	95.70	100.00
E5	86.49	89.19	56.76	66.22	100.00
E6	100.00	100.00	98.59	85.92	100.00
Average	97.63	98.2	99.72	91.31	**100.00**

Note: "*xyz*" refers to the 3D position attributes (x, y, z), "*xyzh*" refers to (x, y, z, h), "*angular*" refers to (X, Y, Z), "*angw*" refers to (X, Y, Z, W), and "*xyzhangw*" represents using all attributes (x, y, z, h, X, Y, Z, W).

Table 5
Subjects in each cross-validation folds.

CV Fold	Subject ID
Fold 1	2, 3, 1, 4, 7
Fold 2	5, 6, 8, 9, 16
Fold 3	10, 11, 17, 18, 23
Fold 4	12, 13, 20, 21, 24
Fold 5	14, 15, 19, 22, 25

volved a learning-based method, we fed all attributes to the model to determine whether it could automatically learn discriminative features from the data. Ultimately, the model returned promising results by using all attributes of skeleton joints with 100% training accuracy on all exercises.

We also performed prediction experiments to determine whether our method could be used to evaluate exercise quality and reflect the severity of AD. To make our experiments less biased compared to simply splitting the data into a training set and test set, we adopted the *k*-fold cross-validation (CV) evaluation method, which is commonly used to estimate the skill of a machine learning model when the sample size is relatively small [40]. We set *k* to 5 for CV by splitting the data of each exercise to 5 folds based on a cross-subjects evaluation criterion as detailed in Table 5.

According to the comparison of different attribute combinations in Table 4, we used all skeleton joint attributes for abnormality prediction. Table 6 shows the experimental results of our CV evaluation. Findings consistently indicated that exercises such as "wave hands" and "bend waist to right" (i.e., E1 and E4, respectively) effectively reflected AD severity, slightly outperforming exercises such as "hands up and down" (E2) and "bend waist to left" (E3). However, walking-related exercises did not demonstrate good performance in terms of AD abnormality prediction.

To further clarify the effectiveness of the proposed method, we visualize the average evaluation scores of all exercises for 25 subjects in Figs. 7 and 8 (differently shaped markers). The lines in

Table 6
Abnormality prediction accuracy for cross-validation evaluation.

CV Fold	Exercise Abnormality Prediction (%)					
	E1	E2	E3	E4	E5	E6
Fold 1	100.00	88.89	81.48	90.00	57.14	57.14
Fold 2	97.06	74.07	82.86	92.31	75.00	75.00
Fold 3	97.56	86.67	76.19	91.43	75.00	62.50
Fold 4	86.67	80.00	73.08	96.15	62.50	80.00
Fold 5	95.56	83.33	85.00	83.72	62.50	71.43
Avg	95.37	82.59	79.72	90.72	66.43	69.21

Fig. 7 indicate that the prediction results of non-walking-related exercises coincided with the clinical severity evaluations of AD. In Fig. 8, no discriminative features were apparent for walking-related exercises such as "walking forward" and "walking backward". This result could inform the design of exercises for rehabilitation therapies; imitating basic actions such as "walking forward" and "walking backward" is less dependent on praxis conditions. Given that our data were collected in a natural environment compared to UI-PRMD [7], whose data were gathered from young subjects mimicking abnormal exercises, our work reflects actual circumstances for patients with AD. On one hand, these findings imply that patients with AD usually exhibit unnoticeable abnormal symptoms during regular walking exercises. On the other hand, the proposed method can potentially capture AD severity from exercises that require good praxis conditions (e.g., imitation, production, or recognition of gestures).

To quantify the performance of our 2T-GCN method, we considered two association parameters (ED and CORREL) as introduced in Section 5.3. We focused on investigating which exercises containing which skeleton data attributes could achieve optimal exercise evaluation performance by comparing their associations with clinical severity labels of AD. When considering normal subjects, Fig. 9 shows that E1 (i.e., waving hands) may be the best exercise for inferring exercise quality: its evaluation scores had the lowest ED and highest correlation compared with clinical evaluation. In terms of evaluating AD severity, based on the correlation and normalized ED in Fig. 10, it remains challenging to model an exercise evaluation score that is highly associated with clinical observation. Notably, the results of E1 and E3 are at least consistent and positively associated with the clinical evaluation of AD, thereby validating the contribution of this study.

5.4.3. Discussion

As listed in Tables 3, 4, and 6, different exercises are capable of identifying AD-related abnormalities to varying degrees. In particular, walking-related exercises in our EHE dataset displayed the weakest capability, namely because these exercises are daily actions that rely less on good praxis conditions for imitation. Actions that are not common in everyday life generally require good praxis conditions (e.g., imitation, production, or recognition of gestures) for good performance.

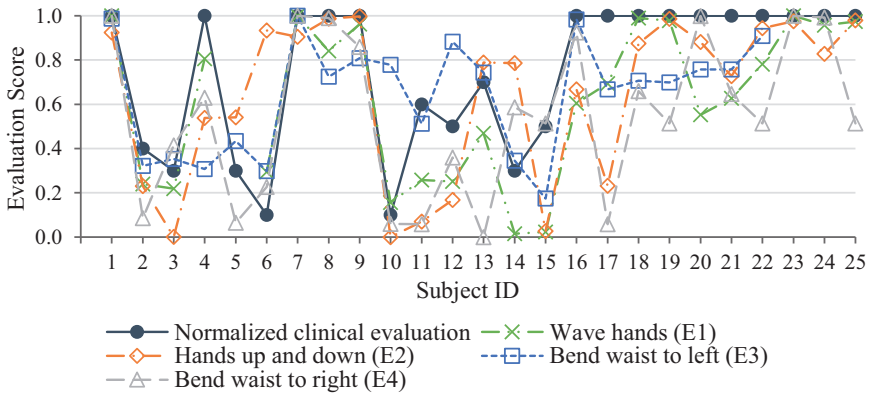

Fig. 7. Average action quality evaluation scores of Exercises 1 to 4.

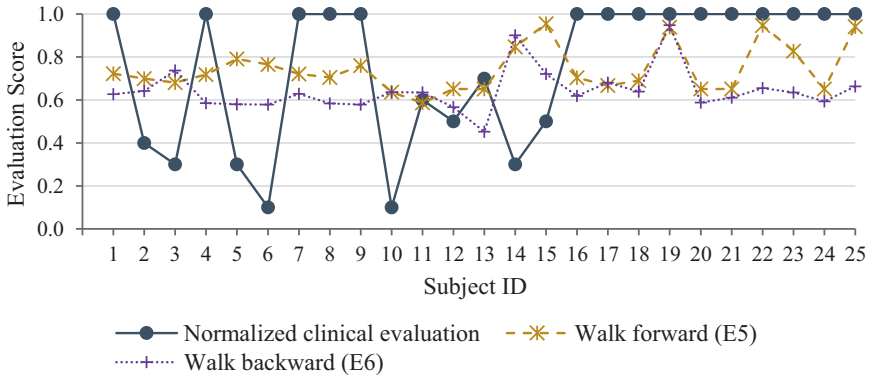

Fig. 8. Average action quality evaluation scores of Exercises 5 and 6.

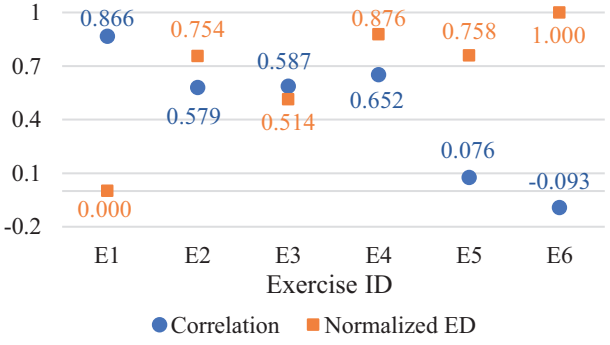

Fig. 9. Comparison of HAE ability with different exercises considering all subjects (numbers are colored).

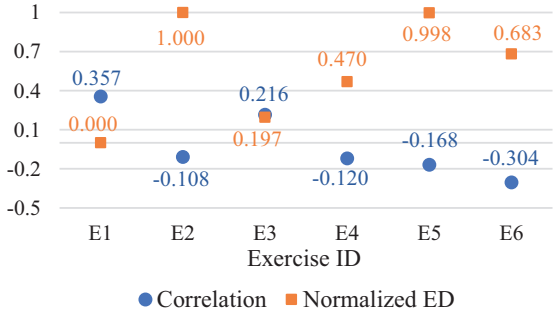

Fig. 10. Comparison of HAE ability with different exercises considering subjects with AD (numbers are colored).

Although this work underscores the potential of using a skeleton-based method to infer AD symptoms, several limitations could be addressed in future research. First, our dataset was collected at one time point. The exercises were performed daily, and patients symptoms might progress over time; therefore, additional rounds of data collection could further validate the proposed algorithm as well as the exercises effectiveness in alleviating AD progression. Another limitation involves the lack of inter-rater validation: we did not compare other mainstream approaches to AD diagnosis because data were unavailable for various biomarkers.

6. Conclusions

In sum, this paper presents a skeleton-based HAE method called 2T-GCN to support AD diagnosis and related behavioral therapies. With respect to diagnosis, the proposed method can perform abnormality detection based on skeleton exercise data. Regarding behavioral therapies, the numerical evaluation score generated from 2T-GCN can reflect elderly individuals praxis health conditions. The proposed method was first validated on the UI-PRMD dataset and displayed superior performance compared with existing methods in terms of the SD. Experimental results from a real-world EHE dataset that we collected in an elderly home also indicated that our 2T-GCN method could successfully predict abnormality from non-walking-related exercises, showing promise to support AD diagnosis. Additionally, numerical evaluation scores generated using the proposed model were positively associated with AD severity. Our method thus appears useful for monitoring the progress of exercise-based interventions. In the future, we will expand these experiments over a longer monitoring period and integrate the proposed method in real-world applications.

Declaration of Competing Interest

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

Acknowledgements

This work is supported by The Hong Kong Polytechnic University for the Project "Music Therapy for Dementia Prevention Using Algorithmic Composition" under Grant No. P0008740. The authors would also like to thank Miss Zhu, a clinical staff member at the Hezhuang Elderly Home, and all elderly patients involved in the EHE dataset.

References

[1] B. Liu, H. Cai, Z. Ju, H. Liu, RGB-D sensing based human action and interaction analysis: a survey, Pattern Recognit. 94 (2019) 1–12.
[2] Q. Lei, J.-X. Du, H.-B. Zhang, S. Ye, D.-S. Chen, A survey of vision-based human action evaluation methods, Sensors 19 (19) (2019) 4129.
[3] F. Patrona, A. Chatzitofis, D. Zarpalas, P. Daras, Motion analysis: action detection, recognition and evaluation based on motion capture data, Pattern Recognit. 76 (2018) 612–622.
[4] L. Li, A. Vakanski, Generative adversarial networks for generation and classification of physical rehabilitation movement episodes, Int. J. Mach. Learn.Comput. 8 (5) (2018) 428.
[5] Y. Liao, A. Vakanski, M. Xian, A deep learning framework for assessing physical rehabilitation exercises, IEEE Trans. Neural Syst.Rehabil. Eng. 28 (2) (2020) 468–477.
[6] A. Elkholy, M. Hussein, W. Gomaa, D. Damen, E. Saba, Efficient and robust skeleton-based quality assessment and abnormality detection in human action performance, IEEE J. Biomed. Health Inf. (2019).
[7] A. Vakanski, H.-p. Jun, D. Paul, R. Baker, A data set of human body movements for physical rehabilitation exercises, Data 3 (1) (2018) 2.
[8] J. Xu, K.D. Kochanek, S.L. Murphy, B. Tejada-Vera, Deaths: final data for 2014 (2016).
[9] J. Gill, M.J. Moore, The state of aging & health in America 2013(2013).
[10] P. Cao, X. Shan, D. Zhao, M. Huang, O. Zaiane, Sparse shared structure based multi-task learning for MRI based cognitive performance prediction of Alzheimer's disease, Pattern Recognit. 72 (2017) 219–235.
[11] M.A. El-Yacoubi, S. Garcia-Salicetti, C. Kahindo, A.-S. Rigaud, V. Cristancho-Lacroix, From aging to early-stage Alzheimer's: uncovering handwriting multimodal behaviors by semi-supervised learning and sequential representation learning, Pattern Recognit. 86 (2019) 112–133.
[12] B. Dubois, H.H. Feldman, C. Jacova, S.T. DeKosky, P. Barberger-Gateau, J. Cummings, A. Delacourte, D. Galasko, S. Gauthier, G. Jicha, Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria, Lancet Neurol. 6 (8) (2007) 734–746.
[13] B. Reisberg, S.H. Ferris, M.J. de Leon, T. Crook, The global deterioration scale for assessment of primary degenerative dementia, Am. J. Psychiatry (1982).
[14] Seven stages of dementia | symptoms & progression., 2016, (https://www.dementiacarecentral.com/aboutdementia/facts/stages/).
[15] P. Heyn, B.C. Abreu, K.J. Ottenbacher, The effects of exercise training on elderly persons with cognitive impairment and dementia: a meta-analysis, Arch. Phys. Med. Rehabil. 85 (10) (2004) 1694–1704.
[16] S.M. Arkin, N. Morrow-Howell, Elder rehab: a student-supervised exercise program for Alzheimer's patients, Gerontologist 39 (6) (1999) 729–735.
[17] L. Teri, L.E. Gibbons, S.M. McCurry, R.G. Logsdon, D.M. Buchner, W.E. Barlow, W.A. Kukull, A.Z. LaCroix, W. McCormick, E.B. Larson, Exercise plus behavioral management in patients with alzheimer disease: a randomized controlled trial, Jama 290 (15) (2003) 2015–2022.
[18] E. Dove, A. Astell, The kinect project: group motion-based gaming for people living with dementia, Dementia 18 (6) (2019) 2189–2205.
[19] A. Diamond, Executive functions, Annu. Rev. Psychol. 64 (2013) 135–168.
[20] Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh, Realtime multi-person 2D pose estimation using part affinity fields, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7291–7299.
[21] F. Moreno-Noguer, 3D human pose estimation from a single image via distance matrix regression, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 1561–1570.
[22] G. Pavlakos, X. Zhou, K.G. Derpanis, K. Daniilidis, Coarse-to-fine volumetric prediction for single-image 3D human pose, in: Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, IEEE, 2017, pp. 1263–1272.
[23] F. Han, B. Reily, W. Hoff, H. Zhang, Space-time representation of people based on 3D skeletal data: a review, Comput. Vis. Image Underst. 158 (2017) 85–105.
[24] B. Galna, G. Barry, D. Jackson, D. Mhiripiri, P. Olivier, L. Rochester, Accuracy of the microsoft kinect sensor for measuring movement in people with Parkinson's disease, Gait Posture 39 (4) (2014) 1062–1068.
[25] Y. Chen, M. Duff, N. Lehrer, H. Sundaram, J. He, S.L. Wolf, T. Rikakis, A computational framework for quantitative evaluation of movement during rehabilitation, in: AIP Conference Proceedings, vol. 1371, American Institute of Physics, 2011, pp. 317–326.
[26] M.J. Fard, S. Ameri, R. Darin Ellis, R.B. Chinnam, A.K. Pandya, M.D. Klein, Automated robot-assisted surgical skill evaluation: predictive analytics approach, Int. J. Med. Rob.Comput. Assisted Surg. 14 (1) (2018) e1850.
[27] L. Tao, A. Paiement, D. Damen, M. Mirmehdi, S. Hannuna, M. Camplani, T. Burghardt, I. Craddock, A comparative study of pose representation and dynamics modelling for online motion quality assessment, Comput. Vis. Image Underst. 148 (2016) 136–152.
[28] L.L. Presti, M. La Cascia, 3d skeleton-based human action classification: a survey, Pattern Recognit. 53 (2016) 130–147.
[29] N. Sadawi, A. Miron, W. Ismail, H. Hussain, C. Grosan, Gesture correctness estimation with deep neural networks and rough path descriptors, in: 2019 International Conference on Data Mining Workshops (ICDMW), IEEE, 2019, pp. 595–602.
[30] J. Liu, G. Wang, P. Hu, L.-Y. Duan, A.C. Kot, Global context-aware attention LSTM networks for 3D action recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1647–1656.
[31] J. Liu, A. Shahroudy, D. Xu, A.C. Kot, G. Wang, Skeleton-based action recognition using spatio-temporal LSTM network with trust gates, IEEE Trans. Pattern Anal. Mach.Intell. 40 (12) (2017) 3007–3021.
[32] S. Yan, Y. Xiong, D. Lin, Spatial temporal graph convolutional networks for skeleton-based action recognition, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
[33] L. Shi, Y. Zhang, J. Cheng, H. Lu, Two-stream adaptive graph convolutional networks for skeleton-based action recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 12026–12035.
[34] Z. Liu, H. Zhang, Z. Chen, Z. Wang, W. Ouyang, Disentangling and unifying graph convolutions for skeleton-based action recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 143–152.
[35] G. Hinton, O. Vinyals, J. Dean, Distilling the knowledge in a neural network, arXiv preprint arXiv:1503.02531 (2015).
[36] M.A.R. Ahad, A.D. Antar, O. 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, pp. 1–11.
[37] D. González-Ortega, F. Díaz-Pernas, M. Martínez-Zarzuela, M. Antón-Rodríguez, A kinect-based system for cognitive rehabilitation exercises monitoring, Comput. Methods Programs Biomed. 113 (2) (2014) 620–631.
[38] F. Sardari, A. Paiement, M. Mirmehdi, View-invariant pose analysis for human movement assessment from RGB data, in: International Conference on Image Analysis and Processing, Springer, 2019, pp. 237–248.
[39] J. Antunes, A. Bernardino, A. Smailagic, D.P. Siewiorek, AHA-3D: a labelled dataset for senior fitness exercise recognition and segmentation from 3D skeletal data, in: BMVC, 2018, p. 332.

[40] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical learning, vol. 112, Springer, 2013.

Bruce X. B. Yu obtained Ph.D. in the Department of Computing from The Hong Kong Polytechnic University, Hong Kong. He is now with the Hong Kong Polytechnic University as a Research Associate. His research interests include big data analytics, machine learning, deep learning, human motion analysis and healthcare.

Yan Liu obtained Ph.D. degree in Computer Science from Columbia University in the US. In 2005, she joined The Hong Kong Polytechnic University, Hong Kong, where she is currently an Associate Professor with the Department of Computing and the director of Cognitive Computing Lab. Her research interests span a wide range of topics, ranging from brain modeling and cognitive computing, image/video retrieval, computer music to machine learning and pattern recognition.

Keith C. C. Chan obtained Ph.D. degrees in systems design engineering from the University of Waterloo, Waterloo, ON, Canada. He then worked as a Software Analyst for the development of multimedia and software engineering tools with the IBM Canada Laboratory, Toronto, ON, Canada. In 1994, he joined The Hong Kong Polytechnic University, Hong Kong, where he is currently a Professor with the Department of Computing. He has authored or coauthored more than 250 publications.

Qintai Yang, M.D., is currently the vice president of The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. He has published more than 70 papers in international journals such as The New England Journal of Medicine (NEJM). Prof. Yang also serves as an editor of the journal Therapeutics and Clinical Risk Management (TCRM).

Xiaoying Wang received Ph.D. degree from Sun Yat-sen University. His research interests include rehabilitation, intelligent healthcare, and music therapy. He is currently the project manager of Medical AI center at The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, China.