2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Assessing Physical Rehabilitation Exercises using Graph Convolutional Network with Self-supervised regularization

Chen Du[1], Sarah Graham[2], Colin Depp[2], and Truong Nguyen[1]

***Abstract*— Computer-vision techniques provide a way to conduct low-cost, portable, and real-time evaluations of exercises performed as a part of physical rehabilitation. Recent data-driven methods have explored using deep learning on 3D body-landmark sequences for automatic assessment of physical rehabilitation exercises. However, existing deep learning methods using convolutional neural networks (CNN) fail to utilize the spatial connection information of the human body, which limits the accuracy of these assessments. To overcome these limitations and provide a more accurate method to assess physical rehabilitation exercises, we propose a deep learning framework using a graph convolutional network (GCN) with self-supervised regularization. The experimental results on an existing benchmark dataset validate that the proposed method achieves state-of-the-art performance with lower error than other CNN methods, and the self-supervised learning improves the prediction accuracy.**

***Clinical relevance*—This work established a supervised learning method to automatically assess physical rehabilitation exercises in the home environment using computer vision. This low-cost, portable, and real-time evaluation may provide clinicians with a way to provide feedback to patients about their exercise performance without having to provide in-person supervision.**

I. INTRODUCTION

Physical rehabilitation exercises are important in postoperative recovery and treatment of various musculoskeletal conditions [1]–[4]. It is critical for patients to correctly perform prescribed exercises to gain the expected outcome for recovery [5]–[7]. The execution of rehabilitation exercises is typically monitored in a hospital or clinic environment by a clinician; however, patients are only offered a limited number of supervised sessions due to high cost [8]. Continued correct performance of rehabilitation exercises in the home environment is necessary to promote full recovery. Therefore, an automated computer-vision solution provides a low-cost, portable, and real-time approach to evaluate the performance of physical rehabilitation exercises performed in a home environment. The benefits of a home-based solution are further increased during the COVID-19 pandemic.

In recent years, with the fast development of affordable and portable 3D motion sensors, several methods have been proposed for automatically assessing physical rehabilitation exercises from 3D body landmark sequences. [9]–[11] proposed to use an interactive video game with a video sensor Kinect to capture body motion during exercise and used it to provide supportive feedback to players. [12]–[15] applied a dynamic time-warping algorithm [16] to calculate the degree of matching distance between the motion of trainers and trainees and provide quantitative feedback on the quality of the movement performance. [17], [18] applied probabilistic approaches, such as Markov models and mixtures of Gaussian distributions, to estimate a performance score based on the likelihood of individual observed sequences with a reference model trained on trainers' movements. The latest deep learning approach [19] proposed to use a combination of convolutional neural networks (CNN) and recurrent neural network (RNN) to model exercises and evaluate performance, which achieved state-of-the-art performance on an existing 3D physical rehabilitation dataset [20]. However, there is limited exploration into utilizing the natural spatial information of the human body for these purposes. Although [19] proposed a multi-stream neural network model for different body part inputs to exploit the spatial information of rehabilitation movements, the spatial relationship among different body parts, and joint connections within each part, is not efficiently represented in CNN, thus limiting the accuracy of a performance score prediction.

To overcome the limitations of existing methods, we propose a deep learning framework using a graph convolutional network (GCN) [21] to learn features for quantitative scoring of physical rehabilitation exercises from graph-represented 3D body landmark sequences. Moreover, we propose a self-supervised learning approach to regularize the training

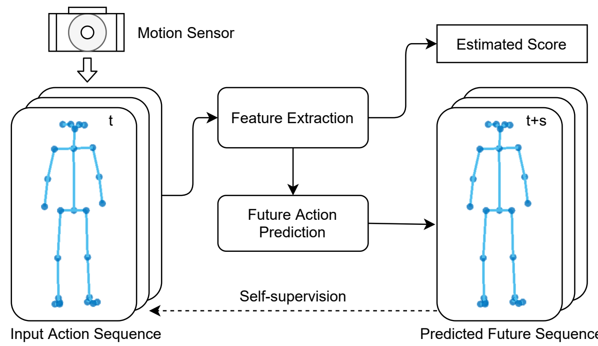


Fig. 1. Proposed framework for assessing physical rehabilitation exercises. The input 3D body landmark sequences are captured by motion sensors during the rehabilitation exercises. The proposed network model extracts graph features from the input sequences and then performs regression of the performance score based on learned features. During the training phase, we optimized a self-supervised regularization branch to predict the future action frames from the extracted features.

[1]C. Du and T. Q. Nguyen are with the Department of Electrical and Computer Engineering, University of California, San Diego, CA 92093 USA. c9du@eng.ucsd.edu, tqn001@eng.ucsd.edu

[2]Sarah Graham and Colin Depp are with the Department of Psychiatry and Sam and Rose Stein Institute for Research on Aging, University of California, San Diego, CA 92093 USA. sagraham@health.ucsd.edu cdepp@health.ucsd.edu

*This study was supported in part by the IBM Research AI

process and force the network to learn the representations of action patterns, increasing the robustness of the proposed model. The proposed framework is shown in Fig. 1.

We evaluate the proposed method on the UI-PRMD [20] benchmark dataset and compare the performance with other state-of-the-art methods. The evaluation result validates that the proposed GCN-based method outperforms other CNN methods by achieving the lowest average prediction error, and that the self-supervised regularization could improve the performance of the regression model.

II. RELATED WORK

A. Physical Rehabilitation Exercise Assessment

Recent studies have proposed several methods for automatic assessment of rehabilitation exercises using body landmark inputs, including game-based guiding systems [9]–[11], quantitative modeling using distance functions [12]–[15] and probabilistic approaches [17], [18]. In this paper, we focus on deep learning methods for assessing exercise performance using 3D body landmark sequences. The most relevant work to our paper is [19], which proposed a temporal-pyramid deep neural network, combining CNN and RNN, to predict performance scores for a variety of different exercises using multi-stream inputs of different body parts. The network model in [19] exploited the spatial information of the human body by using five sub-networks to take five streams of body landmark sequences from two arms, two legs, and the trunk. However, the spatial connections among the five body parts, and joint connections within each part, were not efficiently represented with this method. Since GCN [21] has been validated to learn features from body landmarks for various applications, such as action recognition [21]–[24], we propose to use a GCN [21] as a feature extractor to efficiently exploit the spatial information of body movements.

B. Action Quality Assessment

Action quality assessment (AQA), which quantifies action patterns into scores based on RGB video or body landmarks, has various applications including scoring sport activities, such as diving, skating, and vault [25]–[28]; skill assessment, such as piano playing [29]; and biomechanical metrics estimation [30]. Recent studies have adapted GCNs to learn features for AQA tasks: [31] proposed a diving score estimation method on RGB videos using 3D convolutional network and GCN, and [30] proposed a center of pressure metrics estimation method on 3D body landmarks using a multi-task model combining GCN and long short-term memory (LSTM). In this paper, we adapt GCN for the assessment of a variety of physical rehabilitation exercises, and propose a framework using a GCN as a feature extractor with self-supervised regularization to improve robustness of the quantitative exercise score regression.

III. PROPOSED METHOD

The quantitative physical rehabilitation exercise assessment is a regression task that predicts a performance score from an input body landmark sequence. First, we formulate the proposed assessment framework by defining the data representation of the input body landmark sequences and an objective function for optimization. Next, we propose a regression model using a GCN to extract the graph-structured features from the input data with a self-supervision branch that predicts future body landmarks for regularization.

A. Data representation and objective function

To exploit the spatial information of body joints, we define the graph of body landmark sequences following the spatial-temporal graph representation proposed in [21]. The spatial connections of joints follow the structure of the human skeleton, and each joint is connected with itself among neighbored frames in the temporal dimension, thus forming a spatial-temporal graph. To provide a spatial order of the neighbored joints for graph convolution operation, we apply a partitioning strategy following [21] by selecting the sternum as the center, in which the connections towards the center are labeled as inward and the others labeled outward. Each joint then has three sub-neighbor sets including itself, inward neighbors, and outward neighbors.

To compare different exercises in the same coordinate system, we normalize the score values into a range of 0-1 following [19]. Then we use the mean binary cross-entropy as the objective function for a total of N input sequences

$$l_{score} = -\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i \log y_i + (1-\hat{y}_i)\log(1-y_i)) \qquad (1)$$

where $\hat{y}_i$ is the predicted performance score, and y_i is the corresponding ground truth. We minimize the objective function during the network training process.

B. Score regression model

Based on the graph representation and objective function, we propose a regression model using the spatial-temporal graph convolution (st-gcn) network proposed in [21]. For input body landmark sequences with J joints, the spatial graph convolution is implemented by

$$\mathbf{f}_{out} = \sum_{p=1}^{3} W_p(\mathbf{f}_{in}{\Lambda_p}^{0.5} A_p {\Lambda_p}^{0.5}) \odot M_p \qquad (2)$$

where p is the spatial index according to the spatial partitioning. A_p represents a $J \times J$ adjacency matrix of which the element A_p^{ij} indicates if vertex v_j belongs to the p sub neighbor set v_j. Λ_p is a diagonal normalization matrix whose elements Λ_p^{ii} are the column-sum of A_p^{ij}. M_p is a learnable $J \times J$ attention map [21]. A 2D convolution along the temporal dimension is implemented afterwards.

We sequentially apply four st-gcn blocks with 64 output channels to learn features from the input body landmark sequences. We set the temporal kernel size to 9 and the temporal stride to 1 for a consistent temporal dimension of the output features. For an input body landmark sequence of size $T * D * J$ where T, D, J represents the temporal frame number, the dimension of each joint, and the number of joints respectively, we apply a global average pooling layer

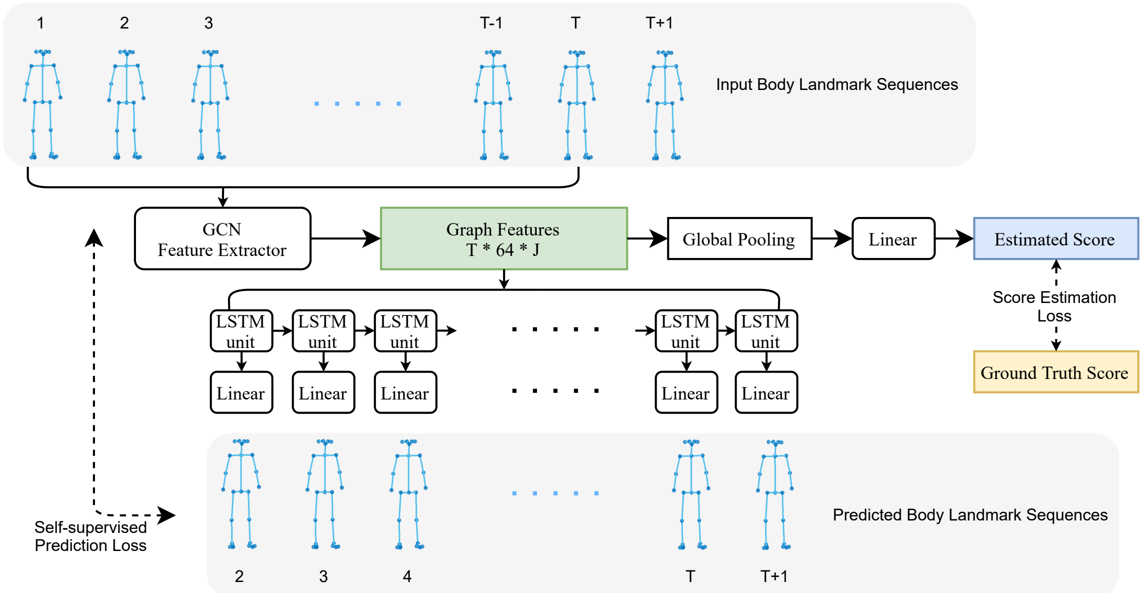


Fig. 2. Proposed regression model with self-supervised regularization. The learned features from the GCN feature extractor are fed into a linear layer to predict the estimated performance scores. During the training phase, we trained a self-supervision branch in parallel to predict the future body landmarks from the learned graph features.

to aggregate the learned features of size $T * 64 * J$ into a $(64J) * 1$ vector, followed by a linear layer with a softmax activation to output the performance score. The architecture of the proposed model is shown in Fig. 2.

C. *Self-supervised regularization*

We propose a self-supervision branch predicting future body landmarks, as regularization during the training phase, to increase the robustness of the performance score regression, which is shown in Fig. 2. Since the prediction of future body landmarks is a sequence-to-sequence regression, we apply a two-layer LSTM with a hidden state size of 512 following the GCN feature extractor as the self-supervision branch. For an input sequence of T frames, the input of each LSTM unit is the learned features at temporal position of frame t, ranging from 1 to T, and the output is the body landmarks at a future temporal position with a shift s at frame $t+s$. We set s to 1 for UI-PRMD dataset [20].

For optimization of the future body landmark prediction, we use a mean square error (MSE) function between the predicted value and ground truth as the objective function

$$l_{self} = \frac{1}{T * J * D} \sum_{t=2}^{T+1} \sum_{j=1}^{J} \sum_{d=1}^{D} (\hat{s}_{tjd} - s_{tjd})^2 \qquad (3)$$

where the $\hat{s}_{tjd}$ is the predicted body landmark value of joint j at frame t along the d-th dimension, and s_{tjd} is the corresponding ground truth. During the training phase, we optimize the objective function of self-supervision branch in parallel with the score regression model via back propagation. During the testing phase, we use the score regression model without the self-supervision branch to predict the performance score; therefore, we avoid extra computational cost during the application of a trained model. Since the performance score is highly correlated with the action patterns, the self-supervision branch, by predicting future body landmarks, forces the GCN feature extractor to learn features that represent rehabilitation action patterns. This step increases the robustness of the score predictions, especially for the existence of some outlier label values.

IV. EXPERIMENTS

A. *Implementation Details*

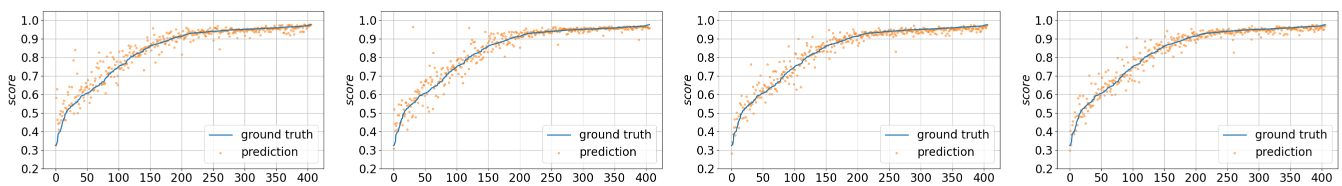
We evaluated the proposed method on the UI-PRMD dataset [20], which is a 3D movement benchmark dataset for physical rehabilitation exercises. The dataset consists of 10 different exercises targeting different regions of the body performed by 10 healthy individuals, with each exercise repeated 10 times in both a correct manner and incorrect manner, respectively. The movements were recorded by a Vicon optical tracker and a Kinect camera. The performance scores were proposed in [19], based on the log-likelihood of a Gaussian mixture model, which encoded a low-dimensional data representation obtained by a deep auto-encoder network. To focus on the comparison of different network models, we followed [19] to use the same inputs of 117-dimensional (39 joints $*$ 3 dimensions) sequences of angular joint displacements recorded by the Vicon tracker, with the same pre-processing including normalization and interpolation.

For each exercise, we trained the proposed regression model on the training set using the Adam optimizer with an initial learning rate of 0.0001 and a batch size of 16. We set the maximum epoch number to 300 and stopped training after observing no improvement for 100 epochs. For the self-supervision branch, we applied a dropout for LSTM with a dropout rate of 0.5. As an ablation study, we also report the results using the same parameters without the self-supervised regularization during the training phase. We performed all experiments using the PyTorch platform on a machine with an Intel i7 4.20 GHz processor and Nvidia RTX 2080-Ti graphic card.

TABLE I

QUANTITATIVE RESULTS (MAE) OF REHABILITATION EXERCISE ASSESSMENT ON UI-PRMD DATASET

Exercise	Proposed GCN + self supervision	Proposed GCN	Liao et al. [19]	Deep CNN [19]	Co-occurrence [32]	Deep LSTM [19]	PA-LSTM [33]	Hierarchical LSTM [20]	Two-stream CNN [34]
E1	**0.00895**	**0.01005**	0.01077	0.01357	0.01052	0.01670	0.01839	0.03010	0.28798
E2	**0.02039**	**0.01932**	0.02824	0.02953	0.02905	0.04934	0.04413	0.07742	0.22349
E3	**0.03643**	**0.03877**	0.03980	0.04141	0.05577	0.09382	0.08094	0.13766	0.20493
E4	0.01448	0.01594	**0.01185**	0.01640	**0.01347**	0.01609	0.02347	0.03580	0.36033
E5	0.01478	**0.01425**	0.01870	**0.01300**	0.01687	0.02536	0.03156	0.06367	0.12332
E6	0.02031	0.02012	**0.01779**	0.02349	**0.01886**	0.02166	0.03426	0.04676	0.21119
E7	**0.02175**	**0.02204**	0.03819	0.03346	0.02733	0.04090	0.04954	0.19280	0.05016
E8	**0.02255**	0.02585	**0.02305**	0.02905	0.02464	0.04590	0.05070	0.07260	0.04337
E9	0.02568	0.02694	**0.02271**	**0.02495**	0.02720	0.04419	0.04313	0.06508	0.14411
E10	**0.02615**	**0.02857**	0.04162	0.03667	0.04657	0.05198	0.07727	0.16009	0.11044
Average	**0.02115**	**0.02219**	0.02527	0.02615	0.02703	0.04059	0.04534	0.08819	0.17593



(a) Deep CNN [19] (b) Liao et al. [19] (c) Proposed GCN (d) Proposed GCN + self supervision

Fig. 3. Aggregated regression results of the physical rehabilitation exercise performance score on UI-PRMD dataset [20], sorted by ground truth values. The horizontal axis indicates the indices of sorted data points (exercise movement sequences), while the vertical axis indicates the performance score. The blue lines represent the ground truth values that are sorted in ascending order. The dots in four subplots represent the corresponding predicted values of (a) Deep CNN [19], (b) Liao et al. [19], (c) Proposed GCN, (d) Proposed GCN with self-supervised regularization.

For quantitative evaluation, we report the mean absolute error (MAE) between the prediction values and ground truth performance scores on the validation set for the 10 exercises. For each exercise consisting of N sequences, the MAE was computed as

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |\hat{y}_i - y_i| \tag{4}$$

where $\hat{y}_i$ was the prediction value, and y_i was the corresponding ground truth score.

For comparison, we report the performance using other state-of-the-art methods including the spatio-temporal neural network model, deep CNN model, and deep LSTM model proposed in [19], as well as the co-occurrence model [32], PA-LSTM [33], two-stream CNN [34], and hierarchical LSTM [20] with the results provided in [19].

B. Evaluation Results

The quantitative evaluation results on the UI-PRMD dataset [20] are reported in Table I. For better visualization, we highlighted the top two methods with the lowest MAE for each exercise and underlined the top method. Comparing the proposed framework with the other methods shows that the proposed GCN methods achieved better performance with lower MAE for most exercises, with a significant improvement for the average MAE of all exercises. Considering the largest MAE among all exercises, we observe that the proposed GCN method is more robust, in that it achieves the lowest MAE for those exercises that are relatively challenging for other methods to assess, such as E2, E3, E7, and E10, by efficiently utilizing the spatial information. Comparing the results of the proposed GCN model with and without the self-supervised regularization, we observe that the self-supervised regularization achieved better performance with a lower average MAE value.

We also present the visualization of aggregated regression results in Fig. 3 using Deep CNN [19], Liao et al. [19], and the proposed model with and without the self-supervised regularization. We aggregated the predicted results of all 10 exercises and sorted them in an ascending order of the corresponding ground truth values. From the plots we observe that the proposed framework is more robust especially for those movements with lower scores under 0.9, while the prediction of the other two CNN-based methods suffers relatively larger error values. Comparing Fig. 3c and 3d, we observe that the proposed model achieved fewer significant outlier points with the self-supervised regularization.

The quantitative evaluation and visualization both validate that the proposed method achieves state-of-the-art performance on the UI-PRMD [20] benchmark dataset, and that the self-supervised regularization could improve the proposed model by increasing the robustness to outliers.

V. CONCLUSIONS

In this paper we present a deep learning framework using GCN and self-supervised regularization to assess the performance of physical rehabilitation exercises. The experiments on an existing benchmark dataset validate that our proposed method achieved state-of-the-art performance with lower prediction MAE and improved robustness. This work will promote a low-cost, portable, and real-time technique to assess the performance of physical rehabilitation exercises.

References

[1] E. Weening-Dijksterhuis, M. H. de Greef, E. J. Scherder, J. P. Slaets, and C. P. van der Schans, "Frail institutionalized older persons: A comprehensive review on physical exercise, physical fitness, activities of daily living, and quality-of-life," *American journal of physical medicine & rehabilitation*, vol. 90, no. 2, pp. 156–168, 2011.

[2] J. S. Brach, S. FitzGerald, A. B. Newman, S. Kelsey, L. Kuller, J. M. VanSwearingen, and A. M. Kriska, "Physical activity and functional status in community-dwelling older women: a 14-year prospective study," *Archives of internal medicine*, vol. 163, no. 21, pp. 2565–2571, 2003.

[3] M. A. Buhagiar, J. M. Naylor, I. A. Harris, W. Xuan, S. Adie, and A. Lewin, "Assessment of outcomes of inpatient or clinic-based vs home-based rehabilitation after total knee arthroplasty: a systematic review and meta-analysis," *JAMA network open*, vol. 2, no. 4, pp. e192 810–e192 810, 2019.

[4] G. Hernández-Molina, S. Reichenbach, B. Zhang, M. Lavalley, and D. T. Felson, "Effect of therapeutic exercise for hip osteoarthritis pain: Results of a meta-analysis," *Arthritis Care & Research: Official Journal of the American College of Rheumatology*, vol. 59, no. 9, pp. 1221–1228, 2008.

[5] S. F. Bassett and H. Prapavessis, "Home-based physical therapy intervention with adherence-enhancing strategies versus clinic-based management for patients with ankle sprains," *Physical therapy*, vol. 87, no. 9, pp. 1132–1143, 2007.

[6] K. Jack, S. M. McLean, J. K. Moffett, and E. Gardiner, "Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review," *Manual therapy*, vol. 15, no. 3, pp. 220–228, 2010.

[7] S. M. McLean, M. Burton, L. Bradley, and C. Littlewood, "Interventions for enhancing adherence with physiotherapy: a systematic review," *Manual therapy*, vol. 15, no. 6, pp. 514–521, 2010.

[8] S. R. Machlin, J. Chevan, W. W. Yu, and M. W. Zodet, "Determinants of utilization and expenditures for episodes of ambulatory physical therapy among adults," *Physical therapy*, vol. 91, no. 7, pp. 1018–1029, 2011.

[9] B. Lange, C.-Y. Chang, E. Suma, B. Newman, A. S. Rizzo, and M. Bolas, "Development and evaluation of low cost game-based balance rehabilitation tool using the microsoft kinect sensor," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 1831–1834.

[10] B. Lange, S. Koenig, E. McConnell, C.-Y. Chang, R. Juang, E. Suma, M. Bolas, and A. Rizzo, "Interactive game-based rehabilitation using the microsoft kinect," in *Virtual Reality Conference, IEEE*. IEEE Computer Society, 2012, pp. 171–172.

[11] C.-Y. Chang, B. Lange, M. Zhang, S. Koenig, P. Requejo, N. Somboon, A. A. Sawchuk, and A. A. Rizzo, "Towards pervasive physical rehabilitation using microsoft kinect," in *2012 6th international conference on pervasive computing technologies for healthcare (PervasiveHealth) and workshops*. IEEE, 2012, pp. 159–162.

[12] C.-J. Su, C.-Y. Chiang, and J.-Y. Huang, "Kinect-enabled home-based rehabilitation system using dynamic time warping and fuzzy logic," *Applied Soft Computing*, vol. 22, pp. 652–666, 2014.

[13] Z. Zhang, Q. Fang, and X. Gu, "Objective assessment of upper-limb mobility for poststroke rehabilitation," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 859–868, 2015.

[14] D. Antón, A. Goni, and A. Illarramendi, "Exercise recognition for kinect-based telerehabilitation," *Methods of information in medicine*, vol. 54, no. 02, pp. 145–155, 2015.

[15] X. Yu and S. Xiong, "A dynamic time warping based algorithm to evaluate kinect-enabled home-based physical rehabilitation exercises for older people," *Sensors*, vol. 19, no. 13, p. 2882, 2019.

[16] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series." in *KDD workshop*, vol. 10, no. 16. Seattle, WA, USA:, 1994, pp. 359–370.

[17] A. Vakanski, J. Ferguson, and S. Lee, "Mathematical modeling and evaluation of human motions in physical therapy using mixture density neural networks," *Journal of physiotherapy & physical rehabilitation*, vol. 1, no. 4, 2016.

[18] M. Capecci, M. G. Ceravolo, F. Ferracuti, S. Iarlori, V. Kyrki, A. Monteriù, L. Romeo, and F. Verdini, "A hidden semi-markov model based approach for rehabilitation exercise assessment," *Journal of biomedical informatics*, vol. 78, pp. 1–11, 2018.

[19] Y. Liao, A. Vakanski, and M. Xian, "A deep learning framework for assessing physical rehabilitation exercises," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 2, pp. 468–477, 2020.

[20] A. Vakanski, H.-p. Jun, D. Paul, and R. Baker, "A data set of human body movements for physical rehabilitation exercises," *Data*, vol. 3, no. 1, p. 2, 2018.

[21] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[22] Y. Du, W. Wang, and L. Wang, "Hierarchical recurrent neural network for skeleton based action recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1110–1118.

[23] Y. Tang, Y. Tian, J. Lu, P. Li, and J. Zhou, "Deep progressive reinforcement learning for skeleton-based action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5323–5332.

[24] K. Liu, L. Gao, N. M. Khan, L. Qi, and L. Guan, "A multi-stream graph convolutional networks-hidden conditional random field model for skeleton-based action recognition," *IEEE Transactions on Multimedia*, 2020.

[25] H. Pirsiavash, C. Vondrick, and A. Torralba, "Assessing the quality of actions," in *European Conference on Computer Vision*. Springer, 2014, pp. 556–571.

[26] P. Parmar and B. Tran Morris, "Learning to score olympic events," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 20–28.

[27] P. Parmar and B. Morris, "Action quality assessment across multiple actions," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 1468–1476.

[28] J. Gao, W.-S. Zheng, J.-H. Pan, C. Gao, Y. Wang, W. Zeng, and J. Lai, "An asymmetric modeling for action assessment," in *European Conference on Computer Vision*. Springer, 2020, pp. 222–238.

[29] P. Parmar, J. Reddy, and B. Morris, "Piano skills assessment," *arXiv preprint arXiv:2101.04884*, 2021.

[30] C. Du, S. Graham, S. Jin, C. Depp, and T. Nguyen, "Multi-task center-of-pressure metrics estimation from skeleton using graph convolutional network," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 2313–2317.

[31] J.-H. Pan, J. Gao, and W.-S. Zheng, "Action assessment by joint relation graphs," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 6331–6340.

[32] C. Li, Q. Zhong, D. Xie, and S. Pu, "Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation," *arXiv preprint arXiv:1804.06055*, 2018.

[33] A. Shahroudy, J. Liu, T.-T. Ng, and G. Wang, "Ntu rgb+ d: A large scale dataset for 3d human activity analysis," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1010–1019.

[34] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," *arXiv preprint arXiv:1406.2199*, 2014.