



HAL
open science

PhysioFormer: A Spatio-Temporal Transformer for Physical Rehabilitation Assessment

Aleksa Marusic, Sao Mai Nguyen, Adriana Tapus

► **To cite this version:**

Aleksa Marusic, Sao Mai Nguyen, Adriana Tapus. PhysioFormer: A Spatio-Temporal Transformer for Physical Rehabilitation Assessment. ICSR 2024, Oct 2024, Odense, Denmark. 10.1145/3568294.3580153 . hal-04857956

HAL Id: hal-04857956

<https://hal.science/hal-04857956v1>

Submitted on 29 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

PhysioFormer: A Spatio-Temporal Transformer for Physical Rehabilitation Assessment

Aleksa Marusic, Sao Mai Nguyen, and Adriana Tapus

*Autonomous Systems and Robotics Lab/U2IS
ENSTA Paris, Institut Polytechnique de Paris
Paris, France*

{aleksa.marusic, sao-mai.nguyen, adriana.tapus}@ensta-paris.fr

Abstract. Studies indicate that physical rehabilitation exercises recommended by healthcare professionals can enhance physical function, improve quality of life, and promote independence for physically disabled individuals. In response to the lack of immediate expert feedback on performed actions, developing an automated system for monitoring such actions and providing feedback is very much needed. In this work, we focus on skeleton-based exercise assessment, which uses skeleton data to evaluate human motion and provide a score on how well a patient performed a movement. There are several approaches to this issue, with Spatio Temporal Graph Convolutional Networks (GCN) being among the most recent. GCNs model skeleton data as graphs and utilize temporal and spatial convolutions to capture relationships between joints more effectively than previous methods. In this research, we propose a new Transformer based model, PhysioFormer. It is inspired by SkateFormer method for human action recognition, with enhanced structure to fit the task of physical rehabilitation assessment. The model leverages skeletal-temporal self-attention across different groups based on relations between joints. The evaluation is done on the KIMORE, UI-PRMD, and KERAAL datasets, benchmark datasets that provide skeleton data captured by Kinect motion system. Our model is surpassing state-of-the-art methods significantly.

Keywords: Transformer · Automated assessment · Physical Rehabilitation

1 INTRODUCTION

In healthcare, physical rehabilitation exercises are essential for post-surgery recovery and the management of various musculoskeletal conditions. Typically, these exercises are overseen by clinicians within hospital or clinic settings. The behavior patients exhibit during these sessions is a key indicator of their health status and treatment progress. However, physiotherapists cannot monitor every patient daily, and patients often need to continue part of their rehabilitation at home without direct supervision. Patients must carefully perform prescribed exercises at home and periodically return to the hospital for progress evaluations.

However, the lack of supervision and timely feedback from healthcare providers can reduce patient engagement during rehabilitation. This diminished motivation and inadequate oversight increase the risk of incorrect exercise execution, potentially hindering recovery and raising the likelihood of re-injury [17].

One potential solution to address this challenge involves the utilization of a robotic coach designed for physical rehabilitation exercises. Such a robot should possess the capability to both learn and perform rehabilitation exercises, as well as evaluate patients' movements [2]. Effective provision of feedback relies heavily on the analysis of human motion, thus emphasizing the significance of human motion analysis within this context [22].

Human activity analysis represents a dynamic and pivotal research area, tackling the complex challenge of interpreting human body movements based on the dynamics of diverse body joints, skeletons, and muscles. This field finds application across diverse domains including video retrieval, entertainment, human-computer interaction, behavior analysis, and many others. [14]

The primary objective of automated monitoring systems for physical rehabilitation is to identify the activity being undertaken, assess its quality, and provide as much information as possible about the error and possible improvement [24,17]. Over recent years, numerous studies have explored the machine learning techniques to assess movement correctness. Among the initial methodologies proposed for this purpose are algorithms based on distance functions, such as Dynamic Time Warping [26] and Mahalanobis distance, as well as probabilistic models like hidden Markov models (HMMs) [3] and Gaussian mixture models (GMMs) [7].

Recently, skeleton-based human action recognition has attracted significant attention for its computational efficiency and resilience to environmental variations and different camera angles. Given that skeletons can be naturally represented as graphs and the increasing capabilities of Graph Neural Networks (GNNs), it is unsurprising that Graph Convolution Networks (GCNs) have emerged as the leading method for skeleton-based action recognition, capitalizing on their ability to process topological information [31].

Deb et al. [6] were among the first to apply Spatio-Temporal GCN (STGCN) to physical rehabilitation assessment, significantly surpassing previous best results in this field. The authors in [21] improved this algorithm, by combining STGCN with Transformer architecture to obtain the current state-of-the-art. On the other hand, Transformer based architectures are getting increasing attention in Human Action Recognition [31], reaching and beating state-of-the-art results on the most popular datasets in Human Action Recognition. Inspired by recent advancements in this field, we propose an extended architecture based on a Transformer for physical rehabilitation assessment, incorporating temporal and spatial partitions, thus reducing the number of parameters needed to be learned.

The rest of the paper is structured as follows: Section 2 is about related works. Section 3 describes the methodology; Section 4 introduces the experimental setup

and presents the results, compared with the state-of-the-art. The conclusions are part of Section 5.

2 RELATED WORKS

An overview of research on skeleton-based action recognition and algorithms for assessing physical rehabilitation exercises is provided in this section.

2.1 Skeleton-based Action Recognition

Recently, significant progress has been made in skeleton-based action recognition. Although initial research was primarily focused on handcrafted features exploiting relative 3D rotations and translations between joints [29,12], Deep Learning based algorithms revolutionized this field by being more robust and achieving better performance [30,14]. They can be split into three categories depending on how they extract features from skeleton data for classification :

- Recurrent Neural Networks (RNNs), such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), consider skeleton data primarily as a temporal sequence of continuous features [18,15,10].
- Convolutional Neural Networks (CNNs) extract spatial information as some studies transformed skeleton data into pseudo-images to apply CNNs [13,19].
- Graph Neural Networks (GNNs) and Graph Convolutional Networks (GCNs): Skeleton data consists of joints and bones, which naturally correspond to vertices and edges in a graph. As a result, research in this field has largely shifted to GNN which can extract both spatial and temporal data by exploiting information contained in the natural topological graph structure of the human skeleton [32,34,5].

2.2 Physical Rehabilitation Assessment

Evaluating movement quality in physical rehabilitation is essential for developing tools and devices that support home-based rehabilitation. Automatic assessment of rehabilitation exercises aims to deliver a quality score based on a sequence of body movements, necessitating a more precise model than typical gesture classification models.

Initially, studies on exercise evaluation employed traditional machine learning methods for classification, such as Adaboost classifiers, K-Nearest Neighbors (KNNs), or Bayesian classifiers, while others tried using distance function-based models [1,11]. Further research used probabilistic approaches like Hidden Markov Models or Gaussian Mixture Models, which provide assessments based on the likelihood that the given exercises are drawn from a trained model [7,27]. These models can incorporate the stochastic nature of human motion, providing a quality score on the correctness of the movement, but such models do not exploit fully information such as joint or spatial connections among body parts.

Liao et al. [16] created a deep neural network model to generate quality scores for input movements. They proposed a deep learning architecture for hierarchical spatio-temporal modeling, combining GMMs, CNNs, and LSTMs to provide a quality score. With the recent development of Graph Neural Networks, it is now possible to exploit the spacial information using the skeletal graph [6] and [9] used Graph Convolutional Networks (GCNs) to assess physical rehabilitation, achieving state-of-the-art scores on commonly used KIMORE, UI-PRMD, and KERAAL datasets. Additionally, Yu et al [35] employed an ensemble of two GCNs, one for position and one for orientation features of the skeleton joints.

3 METHODOLOGY

3.1 Representation of skeleton data

In our work, we consider movement data captured by non-invasive imaging systems such as RGB or depth cameras. These systems capture movement data at a fixed frame rate, recording the spatial information of several skeletal joints of the human pose in 2 or 3 dimensions : joint positions, and in some cases orientations. We use joint orientation data in our experiments because it is more generalizable and maintains accuracy, as noted in [20].

Let us consider a system that tracks V skeletal landmarks on a human subject. The performance of a single repetition of an exercise by a human subject will generate data for T frames. At each frame, joint position information for all V landmarks will be captured in a vector $x(i), i \in [1 \dots T]$, of dimension $D = V \times C$, where C is 2 or 3, depending on whether we capture 2D or 3D data. Stacking these vectors for each frame results in a tensor $X \in \mathbb{R}^{T \times V \times C}$, representing a single repetition.

3.2 Transformer and self-attention

The original concept behind Transformer self-attention is to enable the encoding of correlations between words in a sentence, both short and long-range. The self-attention mechanism allows a network to focus on relevant parts of the input. Encoder-decoder models are designed to learn long-term contextual information. Hidden states in the encoder and decoder networks summarize the input sequence. Attention allows each element in the sequence to have a corresponding hidden state for decoding, taking into account all elements for precise decoding.

This approach can also be applied to skeleton-based action recognition, where correlations between nodes are important in both spatial and temporal dimensions. We treat the joint positions as words and frames as sentences, and we use self-attention to capture more complex relationships between neighbor and distant joints, as well as between neighbor and distant frames.

3.3 Model architecture

As mentioned above, the main processing idea is inspired by SkateFormer from Do. et al. [8]. We adapt their model to our regression problem, and remove

data augmentation from the pipeline, while the overall architecture, which is depicted in Figure 1, remains the same. A skeleton sequence is represented as $\mathbf{X} \in \mathbb{R}^{T \times V \times C}$, as explained above. The 3 linear layers map raw skeleton data into a higher-dimensional feature space. Further, a skeletal-temporal positional embedding is performed before passing the data into a certain number of blocks, each comprising a self-attention layer that propagates features via skeletal-temporal relations and a feed-forward layer that refines the features. The final features undergo skeletal-temporal pooling to produce the outcome $\hat{\mathbf{y}} \in \mathbb{R}$, which represents a score of how well the exercise has been performed. The model is trained via backpropagation through a regression loss function by comparing $\hat{\mathbf{y}}$ with the true label \mathbf{y} - a score given by the physiotherapists.

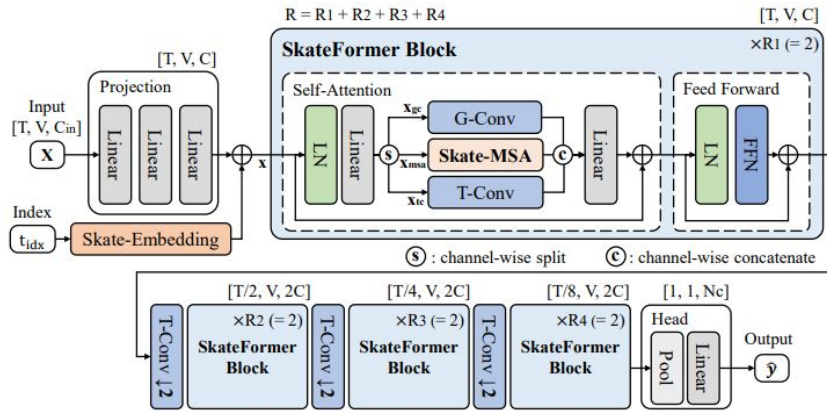


Fig. 1. Overview of the model architecture. Image sourced from [8].

4 RESULTS

In this section, we present the results of comprehensive comparative experiments conducted to assess the performance of our model. Initially, we outline the rehabilitation exercise datasets utilized and the metrics employed for evaluation. Finally, we provide a quantitative comparison between our proposed approach and several state-of-the-art methods.

4.1 Datasets

KIMORE [4] The Kinematic Assessment of Movement for Remote Monitoring of Physical Rehabilitation dataset comprises RGB-D videos along with ground-truth score annotations for 5 distinct exercise types, as shown on Figure 2 It is

organized into 2 primary groups: a control (consisting of expert and non-expert participants) and a group characterized by pain and postural disorders (including Parkinson’s disease, back pain, and stroke). Within the control group, there are 44 healthy subjects, among whom 12 are specialized physiotherapists, while the remaining 32 are non-expert healthy subjects. The pain and postural disorder group consists of 34 subjects grappling with chronic motor disabilities.

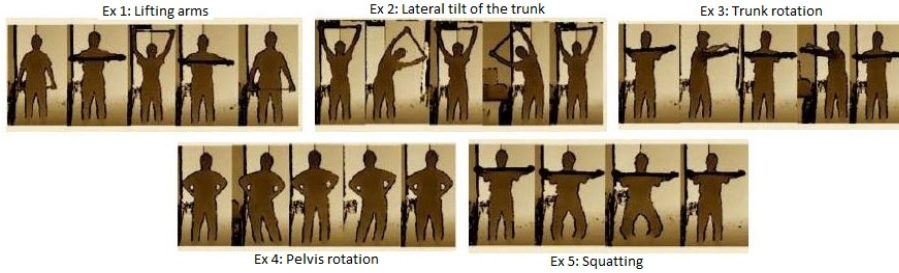


Fig. 2. 5 exercises in the Kimore dataset, namely: (1) Lifting of the arms, (2) Lateral tilt of the trunk with the arms in extension, (3) Trunk rotation, (4) Pelvis rotations on the transverse plane, and (5) Squatting. Image sourced from [4].

UI-PRMD [28] This dataset comprises movements corresponding to common exercises typically undertaken by patients in physical rehabilitation programs. For this dataset, 10 healthy participants executed 10 repetitions of diverse physical therapy movements, with data captured using both a Vicon optical tracker and a Microsoft Kinect sensor. This dataset contains detailed data capturing the positions and angles of full-body joints throughout the exercise routines.

KERAAL [23] The Keraal dataset has been recorded within a long-term rehabilitation double-blind clinical study which includes 31 low-back patients, aged 18 to 70 years. The data includes recordings from both healthy subjects and 12 rehabilitation patients, performing each of 3 predefined exercises. The full dataset encompasses videos and skeleton data collected from a Microsoft Kinect 2 and Vicon. The recordings collected from patients were annotated by a physiotherapist, where each exercise was labeled as either correct or incorrect, but also with a mention of how big the error was, which body part and in which moment the error was made.

4.2 Evaluation metrics

Similar to prior studies [6,16,21], we assess the performance of our model using three key metrics: mean absolute error (MAE), mean absolute percentage error

(MAPE), and root mean square error (RMSE). Lower scores indicate greater predictive accuracy. In the equations below, n is the sample size, and y and \hat{y} are the ground truth and predicted value, respectively.

MAE Mean absolute error represents the average absolute deviation between actual and predicted values. However, MAE is sensitive to scale differences across datasets, limiting its comparability between methods applied to data with varying scales.

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

MAPE In contrast, mean absolute percentage error provides a scale-independent evaluation metric, but it tends toward infinity or becomes undefined when the ground truth value equals zero for any sample.

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}$$

RMSE Root mean squared error, calculated as the square root of the squared errors, penalizes larger deviations due to its squared term. Like MAE, RMSE is also influenced by scale differences across datasets.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

4.3 Implementation details

PhysioFormer is implemented using Python 3.10 and the PyTorch 2.3 + CUDA 12.1 framework. The system used for implementation was a PC equipped with an Intel Core i9-9900KF Silver 4215R CPU, 32GB of RAM, and an Nvidia RTX 2080 Ti graphics card with 11GB of RAM. The model was trained with the Adam optimizer for 1500 epochs, with a learning rate set to 1×10^{-4} , and using batch sizes of 10, 3 and 3 for the KIMORE, UI-PRMD, and KERAAL datasets, respectively. 20% of the data has been devoted to the test set, which was used to select the best model during training. Performance measures were recorded for each run and averaged to ensure accuracy.

4.4 Comparison with the State-of-the-art

Tables 1 and 2 display our results for MAE, RMSE, and MAPE performances, along with those of the state-of-the-art, on the KIMORE and UI-PRMD datasets, respectively, while 3 display our results for F1 score as KERAAL dataset is more suitable for binary scores. Our algorithms show the best on most tests, and in the other tests our score is the second best or very close to the best.

Metric	Ex	PhysioFormer	Mourchid et al. [21]	Deb et al. [6]	Song et al.[25]	Liao et al.[16]	Yan et al.[33]	Du et al.[10]
MAE	Ex1	0.450	<u>0.641</u>	0.799	0.977	1.141	0.889	1.271
	Ex2	0.659	<u>0.753</u>	0.774	1.282	1.528	2.096	2.199
	Ex3	0.198	<u>0.210</u>	0.369	1.105	0.845	0.604	1.123
	Ex4	<u>0.220</u>	0.206	0.347	0.715	0.468	0.842	0.880
	Ex5	0.350	<u>0.399</u>	0.621	1.536	0.847	1.218	1.864
RMSE	Ex1	0.714	2.020	2.024	2.165	2.534	<u>2.017</u>	2.440
	Ex2	0.882	<u>1.468</u>	2.120	3.345	3.738	3.262	4.297
	Ex3	0.460	<u>0.487</u>	0.556	1.929	1.561	0.799	1.925
	Ex4	0.519	<u>0.527</u>	0.644	2.018	0.792	1.331	1.676
	Ex5	0.682	<u>0.735</u>	1.181	3.198	1.914	1.951	3.158
MAPE	Ex1	1.419	<u>1.623</u>	1.926	2.605	2.589	2.339	3.228
	Ex2	0.905	<u>0.974</u>	1.272	3.296	3.976	6.136	6.001
	Ex3	0.603	<u>0.613</u>	0.728	2.968	2.023	1.727	3.421
	Ex4	<u>0.599</u>	0.541	0.824	2.152	2.333	2.325	2.584
	Ex5	1.189	<u>1.217</u>	1.591	4.959	2.312	3.802	5.620

Table 1. Results of our method in comparison with other state-of-the-art approaches on the KIMORE dataset. The best score is in bold, the second are underlined.

Metrics	MAE		RMSE		MAPE	
	PhysioFormer	Mourchid et al. [21]	PhysioFormer	Mourchid et al. [21]	PhysioFormer	Mourchid et al. [21]
Ex1	0.010	0.011	0.017	0.019	1.213	1.289
Ex2	0.009	0.009	0.013	0.014	1.155	1.105
Ex3	0.011	0.013	0.017	0.020	1.483	1.592
Ex4	0.009	0.009	0.012	0.011	0.942	0.984
Ex5	0.008	0.009	0.012	0.013	0.987	1.032
Ex6	0.012	0.013	0.019	0.020	1.401	1.476
Ex7	0.019	0.022	0.029	0.034	2.426	2.697
Ex8	0.018	0.020	0.030	0.032	2.296	2.362
Ex9	0.013	0.013	0.020	0.019	1.501	1.455
Ex10	0.013	0.014	0.021	0.023	1.592	1.619

Table 2. Results of our method in comparison with other state-of-the-art approaches on the UI-PRMD dataset.

Exercise	Torso rotation		Flank stretch		Hiding face	
	PhysioFormer	GMM [23]	PhysioFormer	GMM [23]	PhysioFormer	GMM [23]
F1 score	0.9	0.68	0.82	0.68	0.87	0.72

Table 3. Results of our method in comparison with other state-of-the-art approaches on the Keraal dataset.

5 CONCLUSIONS

In this paper, we propose a Transformer-based model for assessing physical rehabilitation exercises. Our model inputs skeleton data of human movement to predict a quality score for the performed exercise relative to the prescribed version. We adapt a novel SkateFormer architecture to our specific problem, incorporating a partition-specific attention strategy to capture key features while minimizing computational complexity, which is crucial given our small datasets. Our model achieves state-of-the-art performance on the KIMORE, UI-PRMD and KERAAL physical rehabilitation datasets.

References

1. Ar, I., Akgul, Y.S.: A computerized recognition system for the home-based physiotherapy exercises using an rgbd camera. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **22**(6), 1160–1171 (2014)
2. Blanchard, A., Nguyen, S.M., Devanne, M., Simonnet, M., Goff-Pronost, M.L., Rémy-Néris, O.: Technical feasibility of supervision of stretching exercises by a humanoid robot coach for chronic low back pain: The r-cool randomized trial. *BioMed Research International* **2022**, 1–10 (mar 2022)
3. Capecci, M., Ceravolo, M.G., Ferracuti, F., Iarlori, S., Kyrki, V., Monteriù, A., Romeo, L., Verdini, F.: A hidden semi-markov model based approach for rehabilitation exercise assessment. *Journal of Biomedical Informatics* **78**, 1–11 (2018)
4. Capecci, M., Ceravolo, M.G., Ferracuti, F., Iarlori, S., Monteriu, A., Romeo, L., Verdini, F.: 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)
5. Chi, H.G., Ha, M.H., Chi, S., Lee, S.W., Huang, Q., Ramani, K.: Infogcn: Representation learning for human skeleton-based action recognition. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
6. Deb, S., Islam, M.F., Rahman, S., Rahman, S.: Graph convolutional networks for assessment of physical rehabilitation exercises. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **30**, 410–419 (2022)
7. Devanne, M., Nguyen, S.M.: Multi-level motion analysis for physical exercises assessment in kinaesthetic rehabilitation. In: 2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids) (2017)
8. Do, J., Kim, M.: Skateformer: Skeletal-temporal transformer for human action recognition (2024)
9. Du, C., Graham, S.A., Depp, C.A., Nguyen, T.Q.: Assessing physical rehabilitation exercises using graph convolutional network with self-supervised regularization. *International Conference of the IEEE Engineering in Medicine & Biology Society*
10. Du, Y., Wang, W., Wang, L.: Hierarchical recurrent neural network for skeleton based action recognition. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
11. Houmanfar, R., Karg, M., Kulić, D.: Movement analysis of rehabilitation exercises: Distance metrics for measuring patient progress. *IEEE Systems Journal* **10** (2016)
12. Hussein, M.E., Torki, M., Gowayyed, M.A., El-Saban, M.: Human action recognition using a temporal hierarchy of covariance descriptors on 3d joint locations. p. 2466–2472. *IJCAI '13*, AAAI Press (2013)
13. Ke, Q., Bennamoun, An, S., Sohel, F., Boussaïd, F.: A new representation of skeleton sequences for 3d action recognition. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
14. Le, V.T., Tran-Trung, K., Hoang, V.T.: A comprehensive review of recent deep learning techniques for human activity recognition. *Comput. Intell. Neurosci.*
15. Lev, G., Sadeh, G., Klein, B., Wolf, L.: Rnn fisher vectors for action recognition and image annotation. *ArXiv abs/1512.03958* (2015)
16. Liao, Y., Vakanski, A., Xian, M.: A deep learning framework for assessing physical rehabilitation exercises. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **PP**, 1–1 (01 2020)
17. Liao, Y., Vakanski, A., Xian, M., Paul, D., Baker, R.: A review of computational approaches for evaluation of rehabilitation exercises. *Computers in biology and medicine* **119**, 103687 (2020)

18. Liu, J., Wang, G., Yu Duan, L., Abdiyeva, K., Kot, A.C.: Skeleton-based human action recognition with global context-aware attention lstm networks. *IEEE Transactions on Image Processing* **27**, 1586–1599 (2017)
19. Liu, M., Liu, H., Chen, C.: Enhanced skeleton visualization for view invariant human action recognition. *Pattern Recognition* **68** (03 2017)
20. Marusic, A., Nguyen, S.M., Tapus, A.: Evaluating kinect, openpose and blazepose for human body movement analysis on a low back pain physical rehabilitation dataset. In: *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. pp. 587–591 (2023)
21. Mouchid, Y., Slama, R.: D-stgcnt: A dense spatio-temporal graph conv-gru network based on transformer for assessment of patient physical rehabilitation. *Computers in Biology and Medicine* **165**, 107420 (2023)
22. Nguyen, S.M., Tanguy, P.: Cognitive architecture of a humanoid robot for coaching physical exercises in kinaesthetic rehabilitation. In: *International Workshop on Cognitive Robotics* (2016)
23. Nguyen, S.M., Devanne, M., Remy-Neris, O., Lempereur, M., Thepaut, A.: A medical low-back pain physical rehabilitation dataset for human body movement analysis. In: *International Joint Conference on Neural Networks* (2024)
24. Sardari, S., Sharifzadeh, S., Daneshkhah, A., Nakisa, B., Loke, S.W., Palade, V., Duncan, M.J.: Artificial intelligence for skeleton-based physical rehabilitation action evaluation: A systematic review. *Computers in Biology and Medicine* (2023)
25. Song, Y.F., Zhang, Z., Wang, L.: Richly activated graph convolutional network for action recognition with incomplete skeletons. In: *2019 IEEE International Conference on Image Processing (ICIP)* (2019)
26. Su, C.J., Chiang, C.Y., Huang, J.Y.: Kinect-enabled home-based rehabilitation system using dynamic time warping and fuzzy logic. *Appl. Soft Comput.*
27. Vakanski, A., Ferguson, J., Lee, S.: Mathematical modeling and evaluation of human motions in physical therapy using mixture density neural networks. *Journal of Physiotherapy & Physical Rehabilitation* **1** (10 2016)
28. Vakanski, A., Jun, H.p., Paul, D., Baker, R.: A data set of human body movements for physical rehabilitation exercises. *Data* **3**(1) (2018)
29. Vemulapalli, R., Arrate, F., Chellappa, R.: Human action recognition by representing 3d skeletons as points in a lie group. *2014 IEEE Conference on Computer Vision and Pattern Recognition* pp. 588–595 (2014)
30. Wang, L., Huynh, D.Q., Koniusz, P.: A comparative review of recent kinect-based action recognition algorithms. *IEEE Transactions on Image Processing*
31. Xin, W., Liu, R., Liu, Y., Chen, Y., Yu, W., Miao, Q.: Transformer for skeleton-based action recognition: A review of recent advances
32. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. *AAAI'18/IAAI'18/EAAI'18*, AAAI Press (2018)
33. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 32 (2018)
34. Ye, F., Pu, S., Zhong, Q., Li, C., Xie, D., Tang, H.: Dynamic gcn: Context-enriched topology learning for skeleton-based action recognition. *MM '20*, Association for Computing Machinery, New York, NY, USA (2020)
35. Yu, B.X., Liu, Y., Zhang, X., Chen, G., Chan, K.C.: Egcn: An ensemble-based learning framework for exploring effective skeleton-based rehabilitation exercise assessment. In: *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*