# FineRehab: A Multi-modality and Multi-task Dataset for Rehabilitation Analysis

Jianwei Li*†[1], Jun Xue*[1], Rui Cao[1], Xiaoxia Du[2],
Siyu Mo[1], Kehao Ran[1], Zeyan Zhang[2]

[1]School of Sports Engineering, Beijing Sport University, China
[2]Department of Neurorehabilitation, Rehabilitation Research Center, China
[1]{jianwei, jxue, caorui, mosiyu, rkh117}@bsu.edu.cn
[2]364906784@qq.com [2]hyzhangzeyan@163.com

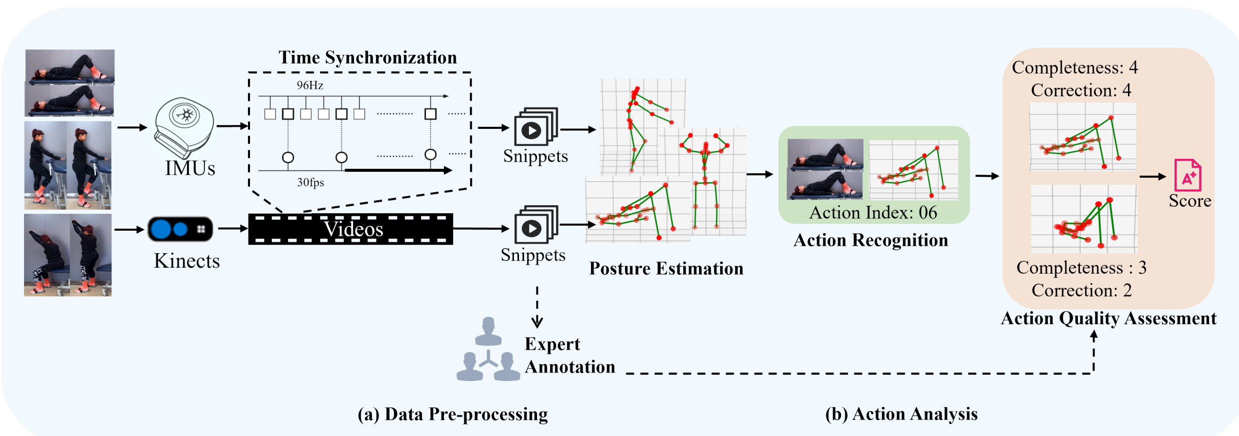


Figure 1. The pipeline of our proposed rehabilitation exercise analysis method.

## Abstract

*The assessment of rehabilitation exercises for neurological and musculoskeletal disorders are crucial for recovery. Traditionally, assessment methods have been subjective, with inherent uncertainty and limitations. This paper introduces a novel multi-modality dataset named FineRehab§ to prompt the study of rehabilitation movement analysis, leveraging advancements in sensor technology and artificial intelligence. FineRehab collects 16 actions from 50 participants, including both patients with musculoskeletal disorders and healthy individuals, and consists of 4,215 action samples captured by two Kinect cameras and 17 IMUs. To benchmark FineRehab, we present a reliable approach to analyze rehabilitation exercises, and make experiments to evaluate the comprehensive movement quality from across multi-dimensions. Comparative experimental analyses have verified the validity of our dataset in distinguishing between the movement of the normal population and patients, which can offer a quantifiable basis for personalized rehabilitation feedback. The introduction of FineRehab will encourage researchers to apply, develop and adapt various methods for rehabilitation exercise analysis.*

## 1. Introduction

Home-based rehabilitation therapy is crucial for the recovery of patients suffering from neurological and musculoskeletal disorders, such as strokes and fractures. Traditionally, the assessment of rehabilitation progress has relied on subjective evaluation scales, including the Fugl-Meyer Scale and Brunnstrom approach, among others [11, 17]. These scales, assessing a broad spectrum of functional capabilities—ranging from joint mobility and muscle strength to motor patterns and overall body coordination—are comprehensive yet flawed. Their efficacy is compromised by a reliance on the evaluators' experience, leading to variability

*Equal contribution
†Corresponding author
§Our dataset can be found at: https://bsu3dvlab.github.io/FineRehab

in assessments and potentially undermining the consistency of patient care. Moreover, the subjective and labor-intensive nature of these assessments does not align well with the needs of an aging population, posing challenges to global healthcare systems.

The emergence of advanced sensor technologies, coupled with the rapid advancements in artificial intelligence (AI) for action analysis, has been instrumental in the development of data-driven assessment tools [24]. This innovative integration facilitates the precise, real-time analysis of patient movements, marking a significant leap forward from traditional assessment methods. By providing quantifiable metrics of rehabilitation progress, this technology heralds a new era in rehabilitation medicine. Specifically, action recognition technology automates the identification of patient movements, while action quality assessment algorithms evaluate the correctness and efficiency of these movements [15]. This synergistic approach ensures the delivery of personalized feedback and recommendations, significantly advancing the rehabilitation assessment towards a fully automated, scalable solution.

The success of AI in analyzing movements, especially in rehabilitation, hinges on creating detailed, high-quality datasets. These datasets are essential for training AI to recognize and evaluate various patient movements accurately. Our research is dedicated to assembling this pivotal dataset, aiming to elevate the precision and reliability of motion recognition technologies within the domain of telemedicine rehabilitation, providing patients with real-time, personalized rehabilitation support, ultimately fostering better patient engagement and improving clinical outcomes. Our main **contributions** are as follows:

1. We propose a multi-modality and multi-task dataset for rehabilitation movement analysis.
2. Independently spatial and temporal categories are proposed to further explore fine-grained action recognition and quality assessment.
3. We provide comparative experimental comparisons between the normal population and patients on FineRehab dataset.

## 2. Related work

### 2.1. Action recognition

Human action recognition (HAR) technology is crucial for identifying human motions, providing an advanced framework for automated monitoring and analysis of patient movements during rehabilitative exercises. The development of three-dimensional skeletal data acquisition through economical depth sensors and pose estimation algorithms, has revolutionized action recognition, particularly in sports domain [6].

Skeletal data represents a breakthrough in HAR by reducing data volume compared to traditional methods like RGB and optical flow, leading to improved computational efficiency essential for real-time rehabilitation assessments. Its robustness to lighting and background variations, along with its insensitivity to camera viewpoint changes, makes it ideal for the varied and unpredictable environments typical of rehab settings [7]. Notably, skeletal data captures detailed motion dynamics over time, providing a nuanced understanding of rehabilitative exercises. The skeletal data provides intricate structural details within each frame and ensures strong temporal continuity across frames, providing a comprehensive spatio-temporal perspective on patient movements. This level of granularity and precision enables more refined analysis and evaluation in rehabilitation practices, supporting the development of treatment modalities that are more tailored and effective.

The primary challenge of traditional methods lies in modeling spatio-temporal features, as they often extract motion patterns from specific skeletal sequences and use manually designed features for representation [6]. While effective in certain scenarios, these handcrafted features struggle to generalize across diverse datasets, limiting the widespread applicability of behavior recognition algorithms. With the rise of deep learning, methods like RNNs [5], CNNs [1], and GCNs [27] have been applied to action recognition using skeletal data. RNN-based methods can process temporal data effectively. However, these methods lack effective modeling capabilities for the spatial relationships of skeletal joints. In contrast, CNN-based methods [5] have the natural ability to learn structural information from two-dimensional arrays, filling some of the gaps in RNN methods. To explore spatial information more clearly, many researchers encode skeletal joints into multiple two-dimensional pseudo-images [7]. In contrast, CNN-based methods naturally learn structural information from two-dimensional arrays, complementing RNNs. To enhance spatial information further, researchers encode skeletal joints into multiple two-dimensional pseudo-images [7].

Skeletal data inherently forms a natural graph topology, with joints and bones analogous to graph nodes and edges, respectively. Recent investigations emphasize the efficacy of Graph Convolutional Networks (GCNs) [27] in skeletal data recognition. Yan *et al*. extended graph convolution to Spatio-Temporal Graph Convolutional Networks (ST-GCN) [29], modeling dynamic skeletons based on the temporal sequence of human joint positions. However, conventional GCNs require manual adaptation of adjacency matrices tailored to specific human topologies, hindering model generalization across datasets and necessitating dataset-specific model training from scratch. UNIK [30] introduces a pioneering initialization strategy, leveraging multiple dependency matrices derived from various attention spectra to facilitate multi-head aggregation. This enables the model to assimilate spatio-temporal skeletal features by exploiting information from diverse representation subspaces, significantly enhancing cross-dataset generalization capabilities.

In this paper, we undertake a comprehensive comparison between two distinct skeleton-based action recognition methods. Through rigorous testing and evaluation, we seek to provide insights into the efficacy of these methods in various application scenarios, thereby contributing to the enhancement of action recognition technology within the field.

### 2.2. Action quality assessment

Action quality assessment (AQA) transcends the simple recognition of physical actions, probing into the subtleties of movement efficacy and accuracy. This exploration is crucial for providing nuanced feedback and guidance throughout rehabilitation processes. Initially, AQA in exercise science was primarily focused on action recognition, treated as a classification problem, where spatial and temporal features from video data were extracted and classified to evaluate action quality using methods such as support vector machine [24], random forests [4], decision trees and neural networks [8], to distinguish between "good" and "bad" execution of actions.

With the advent of deep learning, end-to-end networks have further refined AQA, enabling more sophisticated assessments including regression predictions of movement scores [15]. Vakanski *et al*. [14] proposed three deep learning models based on CNNs, RNNs, and HNNs for this purpose, normalizing the evaluation scores between 0 and 1 to assess movement quality accurately. Yet, predicting performance directly via regression models faces challenges due to the complexity of motor skills and the variability in assessment criteria across dynamic movement stages. This has led to an increased demand for precise and personalized feedback, pushing AQA towards more detailed parameters and multidimensional assessment system, including movement duration, rhythm, joint angles, movement distances and changes in the body's center of gravity [8, 12, 32].

Multi-task learning frameworks have proven effective in enhancing the performance of fine-grained assessments. Xu *et al*. [28] developed a model combining Self-attentive LSTM and Multi-scale Convolutional Skip LSTM for comprehensive quality assessments of sports movements in long videos, significantly improving assessment precision by analyzing both local and global video information and segmenting scores into overall and component scores.

Establishing a reasonable standard for AQA remains challenging. Some studies standardized the movements of healthy individuals or coaches as templates. For example, Baptista *et al*. [31] utilized Dynamic Time Warping (DTW) to compare participant movements with template movements for feedback, while You *et al*. [29] compared patients' movements with therapists' in terms of angle and trajectory similarity. Jain *et al*. [10] introduced a novel metric learning-based Siamese neural network for comparing action videos, combining relevance scores with sub-scores to create a comprehensive final score.

Different from the above work, our research aims to develop a quantitative and comprehensive performance evaluation system that combines traditional and deep learning methods for fine-grained movement analysis, thereby offering patients intuitive and specific feedback.

### 2.3. Rehabilitation exercise dataset

The development and validation of comprehensive datasets for rehabilitation exercises have been a pivotal area of research, underpinning the advancement of both action recognition and quality assessment technologies. HPTE movement dataset [2] provided video and depth image data from Kinect for eight exercises performed by five subjects, while the dataset constructed by Nishiwaki *et al*. [19] is limited to EMG recordings of three lower limb exercises. While these datasets offer valuable insights, their limited movement diversity and data modality range significantly affect the accuracy of machine learning models in recognizing and assessing rehabilitation movements.

To surmount the limitations imposed by single-source data, recent advancements have seen a shift towards multi-view, multimodal data acquisition approaches [23]. The PHYtMO [9] dataset, for example, leverages an optical motion capture system combined with four IMUs on the limbs to achieve a more accurate and holistic understanding of movements. Similarly, the UI-PRMD [26] dataset integrates Vicon optical tracking with Kinect cameras, effectively addressing joint occlusion issues and enriching the dataset for a deeper analysis of rehabilitation exercises.

The challenge of collecting real patient data has led to a reliance on data from healthy individuals performing simulated exercises, which introduces a significant imbalance and fails to accurately represent patient behaviors and movement nuances. Previous efforts to enrich datasets often involve subjects mimicking erroneous movement patterns or patient-like behaviors, yet these do not fully capture the authentic conditions of rehabilitation exercises performed by patients [18, 22, 26]. For personalized treatment and feedback, it is imperative for datasets to include annotations that reflect the quality of movement execution for effective model training. Many studies resort to binary or simple multi-level labels such as "incorrect or correct" or using a scale of 0/1/2 to evaluate movements [9, 18]. However, such coarse or inadequately detailed annotations fall short of enabling the development of models that can meet the precision required for individualized patient care.

This paper proposes a fine-grained and multi-task rehabilitation exercise dataset designed to encompass a broad spectrum of movement types through multi-modality data acquisition strategies. Our dataset characterizes by multi-dimensional and fine-grained expert evaluation labels, offering a robust foundation for to achieve precise assessments and improve algorithms in rehabilitation contexts. Our work aims to bridge the gap in current datasets, facilitating advancements in personalized rehabilitation treatments and enhancing clinical outcomes.

## 3. Methods

We carried out rehabilitation movements recognition and evaluation based on multimodal data in FineRehab. Fig. 1 illustrates the entire process of data pre-processing and rehabilitation exercise analysis. Firstly, we gathered data using IMUs and Kinects, then aligned the timestamps between the two datasets to obtain synchronized multimodal data. Secondly, the multimodal data was fed into recognition and evaluation networks for multi-task analysis. Besides, we conducted pose estimation on the images captured by Kinect.

### 3.1. Data acquisition

The dataset was collected by two Microsoft Kinect Azure cameras and 17 inertial measurement units (IMUs System Perception Neuron Studio from Noitom)[20] to capture the kinematic data of rehabilitation movements. The Kinect cameras was configured to a sampling rate of 30 Hz, while the IMUs were set to 96 Hz. The IMUs were strapped to the subject's *head*, *shoulder blades*, *upper back*, *lower back*, *arms*, *forearms*, *hands*, *thighs*, *calves*, and *feet*. Videos were recorded by two Kinect Azure cameras from front and side view simultaneously. This arrangement was designed to acquire image data from multiple perspectives, addressing the challenge of occlusion of joint points during movements and improving the accuracy of evaluator annotation. The setup of our experiment is detailed in supplement material.

Data collection was undertaken at the China Rehabilitation Research Center and Beijing Sport University, facilitated by a collaborative effort among patients, students, and physiotherapists. The rehabilitation exercises, were vetted by rehabilitation experts for clinical applicability and relevance. Fig. 2 shows 16 movements of three positions (upright, seated, and supine) to meet the needs at different stages of recovery. Detailed descriptions of the selected exercises are provided in supplement material, with a schematic representation of rehabilitation movements.

This study was conducted in strict compliance with ethical standards, adhering to local clinical trial protocols, and was formally approved by the respective ethics committee. To ensure consistency in the execution of these exercises, all subjects received instructions from the same physiotherapist. Each movement was performed in three repetitions to

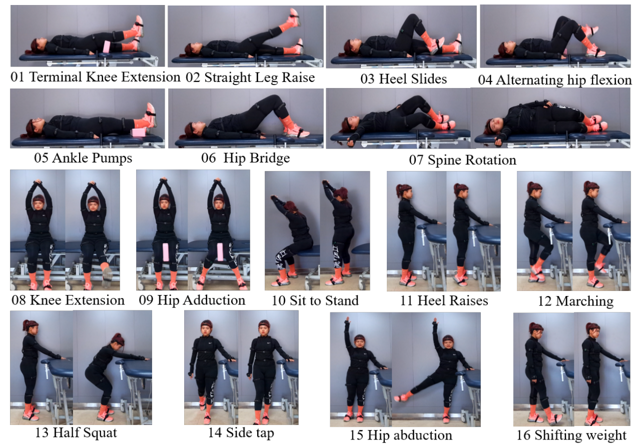


Figure 2. Schematic of 16 movements in FineRehab.

increase the diversity of our dataset. In total, we recorded more than 60 hours of videos over the course of several days.

### 3.2. Subjects information

Fifty volunteers (22 females and 28 males) participated in this study. Among these participants, 30 are musculoskeletal disorders patients, exhibiting varying degrees of stroke-induced motor dysfunctions due to musculoskeletal disorders. The remaining 20 subjects are healthy individuals who reported no movement disorders or other health problems that could affect their mobility. Detailed inclusion and exclusion criteria are available in the supplementary materials. In addition, further details on demographic and anthropometric measurements are provided in FineRehab Dataset.

### 3.3. Data pre-processing

To get a complete and accurate 3D skeleton pose for rehabilitation exercise analysis, we processed the action sequences mainly through the following steps:

**Camera calibration**: The calibration for Kinect cameras were performed before yoga pose capturing. 15-25 checkerboard images were selected in each calibration. The intrinsic matrices of each camera were obtained from Kinect azure SDK, while the geometric relationship of two cameras in front and side views were computed by the tool of stereo camera calibration with MATLAB. The grids in the checkerboard were 13×8, and the actual side length of each grid was 30 cm. The average reprojection error of stereo camera calibration was 2.55 pixels.

**Video cropping**: Since subjects may perform actions other than those required movement during data acquisition, to avoid the influence of this part on the subsequent experiments, we cropped the long video into short repetition-united snippets. At the same time, the incomplete and redundant clips of the original videos were removed.

**Time synchronization**: Each Azure Kinect sensor device included 3.5-mm synchronization portals, which were used for image synchronization among the different devices. Due to the disparity in timestamp recording between Kinect and IMUs,, we developed a synchronous acquisition program based on the synchronization trigger provided by Point Grey industrial camera SDK[21] to ensure the simultaneous start-up of both devices.To effectuate temporal synchronization between two devices, the following procedure was executed for each frame of both devices:

$$f_n = f_{cur} - f_{\text{int}} \tag{1}$$

where $n \in (0, N)$ is the length of an action sequence. The processed timestamp is $f_n$, $f_{cur}$ is the timestamp of each frame, and $f_{int}$ is the first frame timestamp. Subsequently, timestamps were uniformly rounded to three decimal places, and frames sharing identical timestamps were retained as synchronized frames.

**Posture estimation**: Due to significant loss of 3D keypoints estimated through tracking technology of Kinect SDK [25], especially when processed the supine posture movement, thus, our study employed MediaPipe [16] for pose estimation on video data recorded by Kinects. The skeletal data acquired by 17 IMUs and estimated by MediaPipe are shown in Fig. 3. In the Sec. 5, data from these two skeletons were used for comparative experiment.

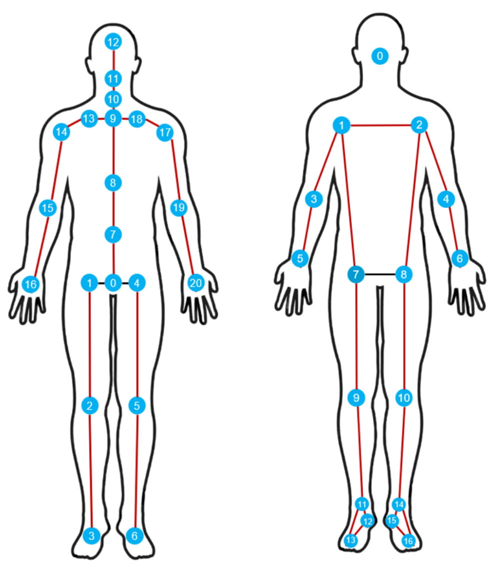


Figure 3. The skeleton data from IMUs and MediaPipe.

# 4. FineRehab dataset

## 4.1. Data annotation

In order to assess movement quality in this study, we developed a detailed multi-dimensional evaluation framework for rehabilitation exercises with suggestions from rehabilitation experts and clinical doctors. This tool evaluates the quality of exercise execution across three dimensions—completeness, correction, and smoothness using a 0 to 4 scale to assess 16 specific rehabilitative actions based on joint-specific performance. The evaluation framework, including dimension parameters, evaluation purposes, objects, criteria, and scores, is summarized in Tab. 1.

All the snippets were independently annotated by three annotators majored in kinesiology and rehabilitation medicine through watching the simultaneous videos captured from multi-view Kinects. In order to ensure the quality of annotation, all annotators participated in rigorous, uniform training sessions instructed by two experts in musculoskeletal neurology before annotation. The inter-rater reliability of annotators resulted in an ICC of 0.784 (95% CI: 0.878, 0.958) and was considered good[13]. The final annotation of single snippets was determined by a majority vote. The statistical results of data annotation are presented in the supplementary materials. Eventually, our FineRehab dataset accrued 51,547 annotated labels as the ground truth for following action quality assessment experiments.

Besides, we further required experts providing the contributory significance of each body parts and evaluation dimensions within each movements in FineRehab. Based on that, the score of overall action quality was accumulated by each dimension score, and defined by following formula:

$$Score = \sum_{i=1}^{k} W_{di} \sum_{j=1}^{m} W_{bj} A_{ij} \tag{2}$$

where $W_d$ and $W_b$ is the weight of dimension and body part of each movement determined by experts, $A_{ij}$ denotes the final annotation result of the $j$-th body part within the $i$-th evaluation dimension.

## 4.2. Data organization

The dataset introduced in this study comprised 4,215 movement snippets in total, each corresponding to a single repetition of rehabilitative exercise. The structure of FineRehab in a hierarchical format is illustrated in Fig. 4. This root directory included two sub-folders: one dedicated to image data and the other to skeletal data. Besides, there were two *CSV* files containing expert annotations and subject information in the root folder.

**Skeletal data**: The skeletal data was categorized by Kinects and IMU sensors, and stored as *JSON* files. Each file contained time stamps, subject IDs, joints number,

Table 1. Evaluation Dimensions and criteria of movement quality assessment.

| Dimension | Evaluation Purpose | Evaluation Object | Evaluation Criteria | Score |
|---|---|---|---|---|
| Completeness | Joint mobility& muscle strength | Primary active joints | The primary active joint has achieved the required range of motion needed for recovery. | 0-4 |
| Correction | Compensation situation | All involved joints | Each joint moves according to the correct kinetic chain without misdirection or compensatory behavior. | 0-4 |
| Smoothness | Movement control | Initiation phase& Reverse phase | Movements are performed smoothly without lag or abnormal jitter andobvious acceleration. | 0-4 |

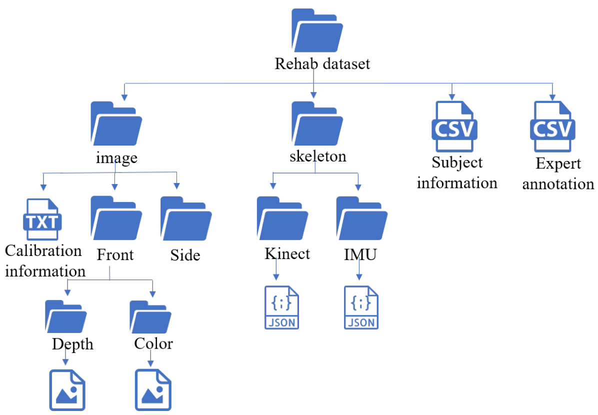


Figure 4. The organization of FineRehab.

3D spatial coordinates and quaternions (IMU only) of joints. As we distinguish between the left side only, right side only and two sides together, the *SideID* was used to record the main limb part during the movement. The nomenclature of the files was as follows: *SubjectID_MovementID_SideID_RepetitionNo.json*. For example, *24_12SL_Cut01.json* referred to the exercise performed by the subject with ID 24, on the first repetition of the movement labeled 12, and performed primarily by left body side.

**Image data**: The image data was systematically divided further into two sub-folders representing the perspectives from front and side views. In addition, calibration information was also provided in Image folder. Then, the sub-folders of image folder consisted of depth images and color images. The nomenclature of the files was similar to the rules above, shown as: *SubjectID_CameraID_ExerciseID_SideID_RepetitionNo_frame.jpg*.

## 5. Rehabilitation exercise analysis

In this section, we made experiments with rehabilitation exercise analysis methods on FineRehab dataset. All experiments were carried out on a 16GB NVIDIA Tesla P100 GPU.

To better understand the performance of prominent action recognition models on this proposed dataset, we benchmarked two skeleton-based models on FineRehab. The 3D skeletal data was trained on ST-GCN [29] and UNIK [30], which offers advanced ways to understand and recognize human actions from skeletal data. ST-GCN captures spatio-temporal dynamics by modeling spatial relations and temporal variations, while UNIK automates the learning of spatial dependencies, enhancing generalization and efficiency in action recognition from skeletal data.

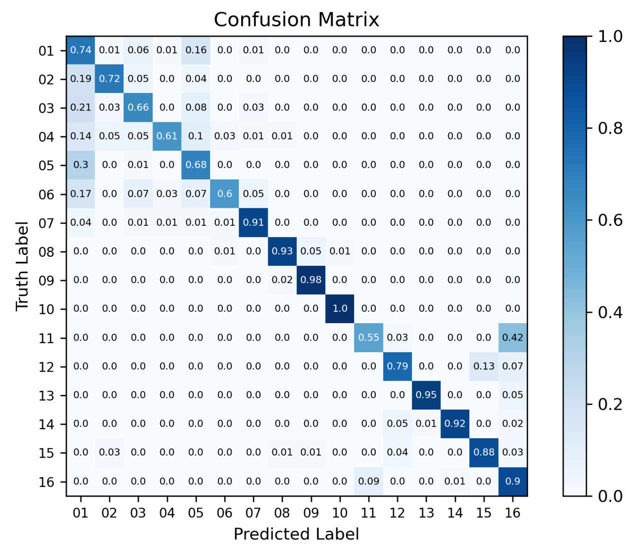


Figure 5. Confusion matrix of 16 movements classification.

### 5.1. Action recognition

The dataset was divided into a training set and a test set at a ratio of 3:1 according to the subject partition strategy, thus, the test set only consisted of samples from unseen subjects not included in the training set. Three augmentation methods, including rotation, transformation, and Gaussian noise, were applied to the data. Our benchmark experiment focused on multi-modality action recognition, cross-subject datasets comparison. The parameters of all models are provided in the supplemental material.

Tab. 2 presents the action recognition results across two models, two groups of subjects and data derived from two modality. Experimental results demonstrates that the dense network structure of UNIK is capable of learning motion information more effectively, achieving an accuracy of 92.63% on our dataset. Moreover, it also converges significantly faster than ST-GCN. The result also indicates that both modals, when trained with data from healthy subjects, generally outperformed that of the patient and mixed groups. Overall, the experimental accuracies of the skeletal data obtained by pose estimation are lower than that of the skeletal data collected directly. Fig. 5 displays the confusion matrix for action classification using mixed group's skeletal data estimated by MediaPipe. It suggests that the classification accuracies for supine movement (01-07) are relatively low, which may associate with the inferior pose estimation performance for supine posture.

Due to UNIK achieving exceptionally high accuracy without data augmentation, the comparison of results with data augmentation was conducted solely on ST-GCN, as shown in Tab. 3. After data augmentation, the accuracy of healthy individual group increases almost 6%. Conversely, data augmentations in the patient group have no effect on accuracy improvement, or even causes it to decline as compared to the baseline, which may be attributed to the greater variability in the movements of patients.

### 5.2. Action quality assessment

The action quality assessment within this study was to comprehensively evaluate the rehabilitative movements performance. We employed traditional methods for comparative analyses between healthy individuals and patients, as well as self-comparisons between a patient's affected and unaffected sides. Features were manually derived, encompassing metrics such as the maximum, minimum, and average rotation angles, alongside movement distances of specific joints. Furthermore, in consideration of the prevalent asymmetries observed in individuals afflicted with unilateral impairments, the discrepancy between left and right body sides was computed for both rotation angles and movement distances. Dynamic Time Warping (DTW)[3] was applied to align and compare movement sequences of patient and healthy subject. Fig. 6 shows the ankle moving distances during movement 11 (*Heel Raises*) between patient and "healthy template" which derived from healthy individual averages. The DTW distances are 13.04 and 13.36 for the patient's left and right side, respectively, versus the template, with an inter-limb distance of 4.78 for the patient and 2.72 for the template. This underscores significant mobility range disparities and a lack of functional symmetry in patients.

Traditional machine learning were applied for binary classification of exercises as abnormal or normal. The data

Table 2. The Top-1 accuracy of action recognition cross-models, cross-skeletons and cross-subjects.

| Models | Data | Estimated | Best Epoch | Accuracy |
|---|---|---|---|---|
| ST-GCN | H[1] | | 174 | 84.35% |
| ST-GCN | H[1] | ✓[4] | 74 | 77.07% |
| ST-GCN | P[2] | | 124 | 77.61% |
| ST-GCN | P[2] | ✓[4] | 59 | 52.32% |
| ST-GCN | H&P[3] | | 139 | **87.94%** |
| ST-GCN | H&P[3] | ✓[4] | 79 | 80.02% |
| UNIK | H[1] | | 35 | **92.63%** |
| UNIK | H[1] | ✓[4] | 37 | 84.58% |
| UNIK | P[2] | | 63 | 76.07% |
| UNIK | P[2] | ✓[4] | 68 | 59.23% |
| UNIK | H&P[3] | | 54 | 86.49% |
| UNIK | H&P[3] | ✓[4] | 34 | 80.33% |

[1] H represents healthy participants.
[2] P represents patients.
[3] H&P refers to all participants.
[4] "✓" indicates that the skeletal data was estimated using MediaPipe, while the absence of "✓" indicates that the skeletal data was directly collected via IMUs.

Table 3. The Top-1 accuracy of action recognition cross-subjects and data augmentation.

| Model | Data | Aug | Best Epoch | Accuracy |
|---|---|---|---|---|
| ST-GCN | H[1] | | 174 | 84.35% |
| ST-GCN | H[1] | ✓[4] | 119 | **90.34%** |
| ST-GCN | P[2] | | 124 | 77.61% |
| ST-GCN | P[2] | ✓[4] | 94 | 75.51% |
| ST-GCN | H&P[3] | | 139 | 87.94% |
| ST-GCN | H&P[3] | ✓[4] | 94 | 87.44% |

[1] H represents healthy participants.
[2] P represents patients.
[3] H&P refers to all participants.
[4] "✓" indicates that the skeletal data was augmented.

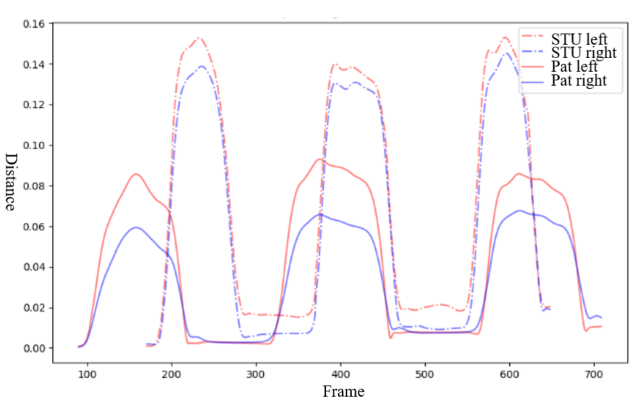


Figure 6. Difference of kinematic metrics between patients and healthy subjects.

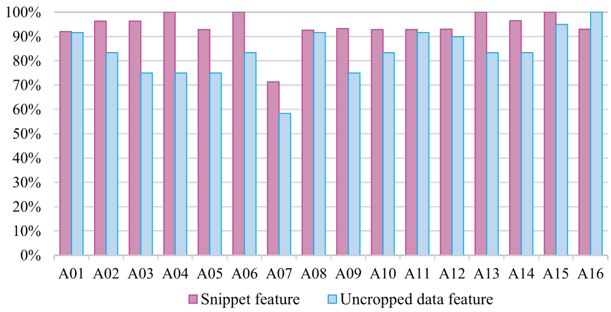


Figure 7. Classification accuracies of assessment results for each dimension by traditional machine learning.

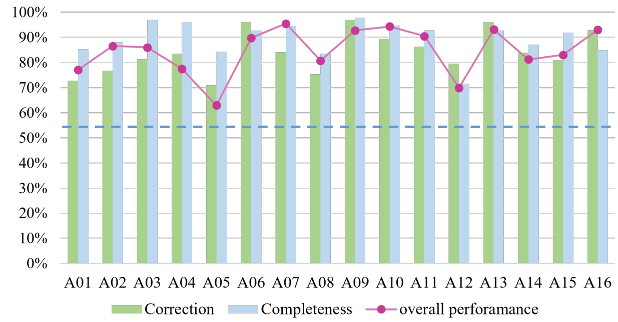


Figure 8. Classification accuracies of assessment results for each dimension by ST-GCN.

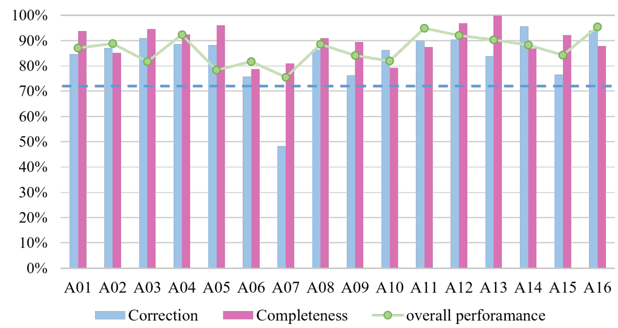


Figure 9. Classification accuracies of assessment results for each dimension by UNIK.

was split, allocating 70% for training and 30% for testing, enhanced by five-fold cross-validation for performance evaluation. Techniques tested in our experiments included Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), K-Nearest Neighbors (KNN), Bayesian methods, and Neural Networks (NN). Fig. 7 illustrates the accuracies of the best-performing algorithm across 16 movements. The experiment also compared outcomes from features extracted from snippets (cropped data, pink) versus repeated movement (uncropped data, blue). The results show segmented features demonstrated greater efficacy in classifying rehabilitation movements, achieving an average accuracy of 93.84%, which illustrates the importance of the pre-processing process of action segmentation. Moreover, instances of classification accuracy falling below the mean are predominantly from multi-joint movements, as manual feature extraction may result in crucial information loss.

In this section, we also employed deep learning methods, such as UNIK and ST-GCN, to classify the movements quality. The rating of action quality was based on the score calculated by Eq. (2). The final score for each action was obtained by aggregating weighted scores across three dimensions: completion, accuracy, and smoothness. Each dimension's score was determined by the multiplication of each joint's assessment labels with its relative weight. Subsequently, the final scores of all movements were used to establish a distribution, from which five levels were defined to represent the overall performance of movement execution: *poor*, *below average*, *average*, *above average*, and *good*. Fig. 8 and Fig. 9 respectively show the performance from 16 movements. The experiment also incorporated other two dimensions assessment experiments :*completeness* and *correction*, which referred in Tab. 1. The blue dashed line in Fig. 8 and Fig. 9 presents the accuracy of level prediction using whole 16 movement data in training progress, achieving 71.9% with UNIK and 54.1% with ST-GCN, respectively. When training with data from specific movements under identical parameters, the accuracy noticeably surpasses that of combined training, indicating significant potential for improvement in learning the unified action evaluation criteria of different rehabilitation movements.

# 6. Conclusion

In this study, we introduce the FineRehab dataset, a multi-modality and multi-task resource designed for fine-grained rehabilitation movement analysis. We have conducted exhaustive action recognition and multi-dimensional action quality assessment on our dataset using both traditional and deep learning methods, and offer insights into differences between healthy individuals and patients, validating our dataset's utility for personalized rehabilitation. Future work may include designing a network to efficiently assess the quality of diverse rehabilitation movements.

# Acknowledgements

This work is partially supported by the National Key R&D Program of China (No.2022YFC3600300 and 2022YFC3600305), Beijing Higher Education Undergraduate Teaching Reform and Innovation Project (No.202310043003), and the Fundamental Research Funds for Central Universities (No.2021TD006 and 2022YB007).

## References


[1] Ayman Ali, Ekkasit Pinyoanuntapong, Pu Wang, and Mohsen Dorodchi. Skeleton-based Human Action Recognition via Convolutional Neural Networks (CNN), 2023. 2

[2] Ilktan Ar and Yusuf Sinan Akgul. 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. 3

[3] Karl Bringmann, Nick Fischer, Ivor van der Hoog, Evangelos Kipouridis, Tomasz Kociumaka, and Eva Rotenberg. Dynamic Dynamic Time Warping, 2023. 7

[4] Chuang Cai, Yang Chunxi, Sheng Lu, Guanbin Gao, and Jing Na. Human Motion Pattern Recognition Based on the Fused Random Forest Algorithm. Preprint, SSRN, 2023. 3

[5] Guanzhou Chen, Sheng Liu, and Jingting Xu. Memory-boosting RNN with dynamic graph for event-based action recognition. *Optoelectronics Letters*, 19(10):629–634, 2023. 2

[6] Huseyin Coskun, David Joseph Tan, Sailesh Conjeti, Nassir Navab, and Federico Tombari. Human Motion Analysis with Deep Metric Learning, 2018. 2

[7] Haodong Duan, Yue Zhao, Kai Chen, Dahua Lin, and Bo Dai. Revisiting Skeleton-based Action Recognition. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2959–2968, New Orleans, LA, USA, 2022. IEEE. 2

[8] Fotos Frangoudes, Maria Matsangidou, Eirini C. Schiza, Kleanthis Neokleous, and Constantinos S. Pattichis. Assessing Human Motion During Exercise Using Machine Learning: A Literature Review. *IEEE Access*, 10:86874–86903, 2022. 3

[9] Sara García-de-Villa, Ana Jiménez-Martín, and Juan Jesús García-Domínguez. A database of physical therapy exercises with variability of execution collected by wearable sensors. *Scientific Data*, 9(1):266, 2022. 3

[10] Hiteshi Jain, Gaurav Harit, and Avinash Sharma. Action Quality Assessment Using Siamese Network-Based Deep Metric Learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(6):2260–2273, 2021. 3

[11] Michael Jones. Therapeutic Exercise: Foundations and techniques. *Physiotherapy*, 77(2):96, 1991. 1

[12] Moeko Kojima, Kyo Kutsuzawa, Dai Owaki, and Mitsuhiro Hayashibe. Game-based Evaluation of Whole-body Movement Functions with CoM Stability and Motion Smoothness. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 4354–4357, Glasgow, Scotland, United Kingdom, 2022. IEEE. 3

[13] Terry K. Koo and Mae Y. Li. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2):155–163, 2016. 5

[14] Yalin Liao, Aleksandar Vakanski, and Min Xian. A Deep Learning Framework for Assessing Physical Rehabilitation Exercises. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(2):468–477, 2020. 3

[15] Yalin Liao, Aleksandar Vakanski, Min Xian, David Paul, and Russell Baker. A review of computational approaches for evaluation of rehabilitation exercises. *Computers in Biology and Medicine*, 119:103687, 2020. 2, 3

[16] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. MediaPipe: A Framework for Building Perception Pipelines, 2019. 5

[17] Hamza Y. Madhoun, Botao Tan, Yali Feng, Yi Zhou, Cuijuan Zhou, and Lehua Yu. Task-based mirror therapy enhances the upper limb motor function in subacute stroke patients: A randomized control trial. *European Journal of Physical and Rehabilitation Medicine*, 56(3), 2020. 1

[18] Alina Miron, Noureddin Sadawi, Waidah Ismail, Hafez Hussain, and Crina Grosan. IntelliRehabDS (IRDS)—A Dataset of Physical Rehabilitation Movements. *Data*, 6(5):46, 2021. 3

[19] Gaston Ariel Nishiwaki, Yukio Urabe, and Kosuke Tanaka. EMG Analysis of Lower Extremity Muscles in Three Different Squat Exercises. *Journal of the Japanese Physical Therapy Association*, 9(1):21–26, 2006. 3

[20] Noitom. Perception neuron studio. `https://www.noitom.com.cn/perception-neuron-studio.html`, 2023. 4

[21] Edmund optics. Point grey grasshopper. `https://www.edmundoptics.cn/f/flir-grasshopper3-high-performance-usb-30-cameras/14785/`, 2023. 5

[22] Adeline Paiement, Lili Tao, Massimo Camplani, Sion Hannuna, Dima Damen, and Majid Mirmehdi. Online quality assessment of human motion from skeleton data. In *Proceedings of the British Machine Vision Conference 2014*, pages 79.1–79.12, Nottingham, 2014. British Machine Vision Association. 3

[23] Manuel Palermo, João M. Lopes, João André, Ana C. Matias, João Cerqueira, and Cristina P. Santos. A multi-camera and multimodal dataset for posture and gait analysis. *Scientific Data*, 9(1):603, 2022. 3

[24] Paritosh Parmar and Brendan Tran Morris. Measuring the quality of exercises. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2241–2244, Orlando, FL, USA, 2016. IEEE. 2, 3

[25] Microsoft Development Team. Synchronize multiple azure kinect dk devices. `https://docs.microsoft.com/en-us/azure/kinect-dk/multi-camera-sync`, 2019. 5

[26] Aleksandar Vakanski, Hyung-pil Jun, David Paul, and Russell Baker. A Data Set of Human Body Movements for Physical Rehabilitation Exercises. *Data*, 3(1):2, 2018. 3

[27] Jiangning Wei, Bo Yu, Hao Zhang, and Jun Liu. Skeleton Based Graph Convolutional Network Method for Action Recognition in Sports: A Review. In *2023 8th IEEE International Conference on Network Intelligence and Digital Content (IC-NIDC)*, pages 66–70, Beijing, China, 2023. IEEE. 2

[28] Chengming Xu, Yanwei Fu, Bing Zhang, Zitian Chen, Yu-Gang Jiang, and Xiangyang Xue. Learning to Score Figure Skating Sport Videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(12):4578–4590, 2020. 3

[29] Sijie Yan, Yuanjun Xiong, and Dahua Lin. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), 2018. 2, 3, 6

[30] Di Yang, Yaohui Wang, Antitza Dantcheva, Lorenzo Garattoni, Gianpiero Francesca, and Francois Bremond. UNIK: A Unified Framework for Real-world Skeleton-based Action Recognition, 2021. 2, 6

[31] Yu You, Tai-Qi Wang, Keisuke Osawa, Megumi Shimodozono, and Eiichiro Tanaka. Kinect-based 3D Human Motion Acquisition and Evaluation System for Remote Rehabilitation and Exercise. In *2022 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 1213–1218, Sapporo, Japan, 2022. IEEE. 3

[32] Yan Zhu, Dong Wang, Run Zhao, Qian Zhang, and Anna Huang. FitAssist: Virtual fitness assistant based on wifi. In *Proceedings of the 16th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, pages 328–337, Houston Texas USA, 2019. ACM. 3