



HAL
open science

A Medical Low-Back Pain Physical Rehabilitation Dataset for Human Body Movement Analysis

Sao Mai Nguyen, Maxime Devanne, Olivier Remy-Neris, Mathieu Lempereur,
Andre Thepaut

► **To cite this version:**

Sao Mai Nguyen, Maxime Devanne, Olivier Remy-Neris, Mathieu Lempereur, Andre Thepaut. A Medical Low-Back Pain Physical Rehabilitation Dataset for Human Body Movement Analysis. IJCNN 2024 - International Joint Conference on Neural Networks, IEEE, Jun 2024, Yokohama, Japan. hal-04629541

HAL Id: hal-04629541

<https://inria.hal.science/hal-04629541>

Submitted on 29 Jun 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

A Medical Low-Back Pain Physical Rehabilitation Dataset for Human Body Movement Analysis

Sao Mai Nguyen
IMT Atlantique
Lab-STICC, UMR 6285
and FLOWERS U2IS,
ENSTA, IP Paris & Inria,
France
nguyensmai@gmail.com

Maxime Devanne
Université de
Haute-Alsace
IRIMAS EA 7499,
France

Olivier Remy Neris
Université Brest
CHU Brest
INSERM, UMR 1101
F 29200 Brest,
France

Mathieu Lempereur
Université Brest
CHU Brest
INSERM, UMR 1101
F 29200 Brest,
France

André Thépaut
IMT Atlantique
Lab-STICC, UMR 6285,
F-29238 Brest,
France

Abstract—While automatic monitoring and coaching of exercises are showing encouraging results in non-medical applications, they still have limitations such as errors and limited use contexts. To allow the development and assessment of physical rehabilitation by an intelligent tutoring system, we identify in this article four challenges to address and propose a medical dataset of clinical patients carrying out low back-pain rehabilitation exercises. The dataset includes 3D Kinect skeleton positions and orientations, RGB videos, 2D skeleton data, and medical annotations to assess the correctness, and error classification and localisation of body part and timespan. Along this dataset, we perform a complete research path, from data collection to processing, and finally a small benchmark. We evaluated on the dataset two baseline movement recognition algorithms, pertaining to two different approaches: the probabilistic approach with a Gaussian Mixture Model (GMM), and the deep learning approach with a Long-Short Term Memory (LSTM). This dataset is valuable because it includes rehabilitation relevant motions in a clinical setting with patients in their rehabilitation program, using a cost-effective, portable, and convenient sensor, and because it shows the potential for improvement on these challenges.

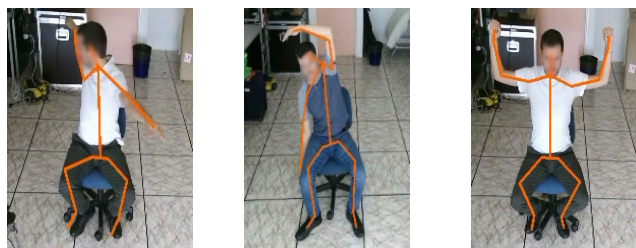
Index Terms—rehabilitation, machine learning, human pose estimation, movement analysis

I. INTRODUCTION

a) Physical rehabilitation

Back pain is the 3rd common musculoskeletal disabling condition in the 45-65 years old population, and therapists consider that regular physical exercises is essential to alleviate back pain. 50 to 80% of the world population suffers at a given

This work is partially supported by the EU FP7 grant ECHORD++ KERAAL and by the European Regional Fund via the VITAAL Contrat Plan Etat Region.



(a) torso rotation (b) flank stretch (c) hiding face
Fig. 1: The three rehabilitation exercises in our dataset

moment from back pain [1], [2], caused by accident, surgery, old age or unfit working habits. While active rehabilitation is considered more effective for physical rehabilitation than usual care [3], [4], therapists are concerned about the decreasing engagement of the patients throughout the months-long repetition of physical exercises [5], primarily attributed to the lack of supervision and monitoring of the patient performance.

To promote physical activity, games [6], mobile phone applications [7], virtual agents [8], [9] and robots [10]–[13] can enhance the engagement of the patient. These approaches rely on body motion analysis of patients. Leveraging machine learning and computer vision, 3D physical body movements analysis has improved the performance of its algorithms. Unfortunately, their performance relies on the quality of the training dataset, and most human motion datasets have not been carried out in the medical context [14]: they lack medical annotation and variability representative of patients, and most contain data from only healthy subjects.

b) Challenges

Our main motivation for creating the presented dataset is to create an intelligent tutoring system capable of supervising rehabilitation sessions autonomously such as presented in [15], by providing instructions for each exercise of the program and real-time feedback how to improve the efficiency of the patient’s performance. With the aim of rehabilitation using human body movement analysis, intelligent tutoring systems (ITS) need to analyse complex full-body exercises that can involve several parts of the body but not necessarily all parts of the body. The assessment algorithm should be able to understand which parts are important, and what are the ranges of freedom that are acceptable. The ITS should encapsulate the tolerated variance for each joint and time frame. Thus, we have identified 4 challenges in rehabilitation movements analysis:

- 1) Rehabilitation motion assessment. The goal is to assess an observed motion sequence by detecting if the rehabilitation exercise is correctly performed or not.
- 2) Rehabilitation error recognition. The goal is to classify the observed error among a set of known errors, so as to explain and give feedback.

- 3) Spatial localization of the error. In addition to recognizing the error, the goal is also to identify which body part is responsible of the error.
- 4) Temporal localization of the error. The goal is to detect the temporal segment where the detected error occurred along the sequence.

While most rehabilitation datasets only provide annotations for the rehabilitation motion assessment, we propose a medical dataset of clinical patients carrying out low back-pain rehabilitation exercises. The dataset includes 3D Kinect skeleton positions and orientations, RGB videos, 2D skeleton data, and medical annotations to assess the correctness, label and timing errors of each movement part. The article also provides initial benchmarks with 2 movement analysis algorithms.

II. RELATED WORK

a) Capture system

A number of human activity datasets have been recently reported. They were mostly captured in two modalities: marker-based motion capture (MoCap) tracking and vision-based (marker-less) tracking [14]. On the other hand, vision-based technologies do not hinder movements. RGB-D cameras, such as the Microsoft Kinect, are affordable, require little calibration or setup. Moreover, the Kinect has been validated against standard motion capture systems despite a lower precision, and against other vision-based tracking algorithms, and wearable sensors [16]–[21]. More recently, human pose estimation from RGB images [22]–[28] have also made impressive improvements, leveraging deep learning models and large datasets. Cameras can easily blend into patient’s homes or a clinical environment. They are affordable, require little calibration or setup. This makes them suitable for use in our clinical test where patients are recorded in daily sessions of 30 minutes. Thus, we provide data from both RGB cameras and the Kinect.

b) Medical Human Movement Dataset

Although several datasets contribute to research in human motion analysis, such as the Kinect datasets [37]–[40] or the multi-sensor datasets [41]–[43], they can hardly be applied in the medical context, including physical rehabilitation.

The K3Da dataset [29] is the first Kinect based dataset in a healthcare setting based on common clinical assessments of gait and balance. Nevertheless, they did not specifically recruit patients and the data were not labelled with medical criteria. The HPTE dataset [31] records with a Kinect therapy movements for computerized monitoring at home of 8 shoulder and knee exercise movements. Again, the participants are healthy. The EMG Squat dataset [30] is restricted to EMG electrodes recordings of 3 exercises of lower limbs performed by 9 healthy participants. The UI-PRMD [32] proposes a dataset of 10 healthy subjects performing physical therapy recorded by a Vicon marker-based tracker and a Kinect. Moreover, the labeling only indicates if the execution is correct, but not the type of error or their timing.

c) Human Movement Dataset with Patients

Targeting low back pain too, the EmoPain dataset [35], to study the effects of pain during physical rehabilitation, provides

high-resolution face videos, audios, full body joint motions, and electromyographs (EMG) from back muscles. The labels include pain facial expressions and pain-related movements. Like K3Da, EmoPain is labeled for recognition of exercises or mental states, they do not address the challenges stated in Sec. I. The TRSP dataset [33] proposes a dataset of clinically relevant motions during robotic rehabilitation exercises focusing on strength exercises with a haptic robot tabletop, captured with a Kinect. Likewise, IRDS [34] provides Kinect data of patients doing general rehabilitation. In TRSP and IRDS, The data from both healthy and patients are assessed by clinicians with scores and error labels, so they can address the challenges 2 and 3. But the labels do detail which body part and timespan of the execution was wrong. The Kimore dataset [36] proposes 5 whole body exercises from patients and healthy subjects. It is the closest to ours : it also targets low-back pain and is labelled by medical annotators. The labels are both at the global and the body part level to address challenges 1, 2 and 3. However, Kimore does not yet provide with the temporal information of errors to address challenge 4.

In comparison with these datasets, our Keraal dataset has been recorded within a long-term rehabilitation program, targeting low back pain. Contrarily to EMG Squat, HPTE and K3Da, UI-PRMD, we have recruited rehabilitation patients and have data labelled by a doctor. Like the HPTE dataset [31], our work is in a long-term effort of enabling physiotherapy exercises at home and our data is extracted from a 4-weeks evolution of each patient. A full comparison can be read in Table I. Thus our Keraal dataset is the only benchmarking set with clinical patients and with labels provided by a physician for the four challenges : motion assessment, error recognition, spatial and temporal localization.

III. DATASET AND FRAMEWORK

This section describes the protocol and rationale for how the Keraal dataset was created, the participants included, the hardware setup and the experimental protocol. The dataset and the code are available on <http://nguyensmai.free.fr/KeraalDataset.html> and <https://github.com/nguyensmai/KeraalDataset>.

a) Rehabilitation program

31 patients, aged 18 to 70 years, were recruited in the double blind study. This prospective, centrally randomized, controlled, single-blind and bi-centric study was conducted from October 2017 to May 2019. 12 patients suffering from low-back pain were included and were asked to perform each of the three predefined exercises the best they can from its demonstration. The details on this clinical trial, including the patient care, the rehabilitation sessions, the inclusion and exclusion criteria, the characteristics of the patients, the efficiency of the care have been reported in [15]. This study has received a Legal authorisation for clinical tests from the medical ethics board of the hospital at Brest (CHRU Brest). All subjects have given their informed consent to participate in the study.

A. Exercises and errors

A list of three exercises have been chosen in conjunction with therapists as common rehabilitation exercises that are also

TABLE I: Comparison with other datasets (Participants column reports the total number of participants including the patients. NA = Not Available, annotators column report the labelling method : if the cases were recorded under instructions or if a medical or non medical human annotator labelled the data)

Dataset	Activities	Exercises	Nb exer.	Pat-ients	Par-ticip-ant ¹	EMG	RGB D	MoCap /wear-able	Nb record.	Annotations	Annotators
K3Da [29]	clinical assess. of gait, balance	posture tests	13	0	54	NA	Kinect2 D, JP	NA	525	None	None
EMG Squat [30]	Therapy to protect the anterior cruciate ligament (ACL)	Squat	3	0	9	leg	NA	NA	81	Exercise type	instructions
HPTE [31]	Physiotherapy exercises at home	Shoulder and knee exercises	8	0	5	NA	Kinect1 RGB D	NA	240	Exercise type	instructions
UI-PRMD [32]	Physical therapy and rehabilitation	Whole-body exercises	10	0	10	NA	Kinect1JP JO	Vicon	1000	Correct/ Incorrect	instructions
TRSp [33]	Post-stroke physical rehab	Arm movements using a haptic robot tabletop	2	9	19	NA	Kinect1 JP JO	NA	190	Error label	2 Med
IRDS [34]	general physical rehabilitation	arms and legs	9	15	29	NA	Kinect2 JP	NA	2589	Exercise type position (sitting or standing), correctness label (correct, incorrect, unrecognisable)	2 Non Med
EmoPain [35]	Chronic low back pain	Physical exercises	7	22	50	back	multi- RGB A	IMU	unk.	pain expression & pain related mvt	8 Non Med + 6 Med
KIMORE [36]	low back pain	whole body	5	34	78	NA	Kinect2 RDB D JP JO	NA	1000	3 scores (quantify the accuracy of the subjects) for total error and body part	3 Med
Keraal	Posture and back pain rehabilitation	Upper-body exercises	3	12	21	NA	Kinect2 RGB, JP, JP	Vicon (Gr3)	2622	Error label + body part, timespan)	2 Med

used for low-back pain treatment, under the condition that they can be coached by an intelligent tutoring system using visual assessment. Illustrations of these exercises can be seen in Fig. 1 . The 3 exercises are centered on spine stretching: a left rotation of the trunk followed by a the same right rotation, a left and right lateral bending of the trunk and a breathing exercise with the upper limbs flexed 90°at shoulder and elbow.

A list of common errors was defined in conjunction with the experience of the therapists of CHRU Brest. Errors are illustrated in Fig. 2.

B. Participants

The dataset contains data from three groups of participants:

- Group1 : Rehabilitation Patients :
The data of the daily sessions of the 12 patients recruited is split into two subsets:
 - Group1a: 14 recordings per exercise among 6 patients were annotated as detailed in Sec.III-D.
 - Group1b: the remaining recordings of the 12 patients without annotation.
- Group2 : Healthy participants with Kinect V2 recordings : Six healthy adults performed the exercises after the

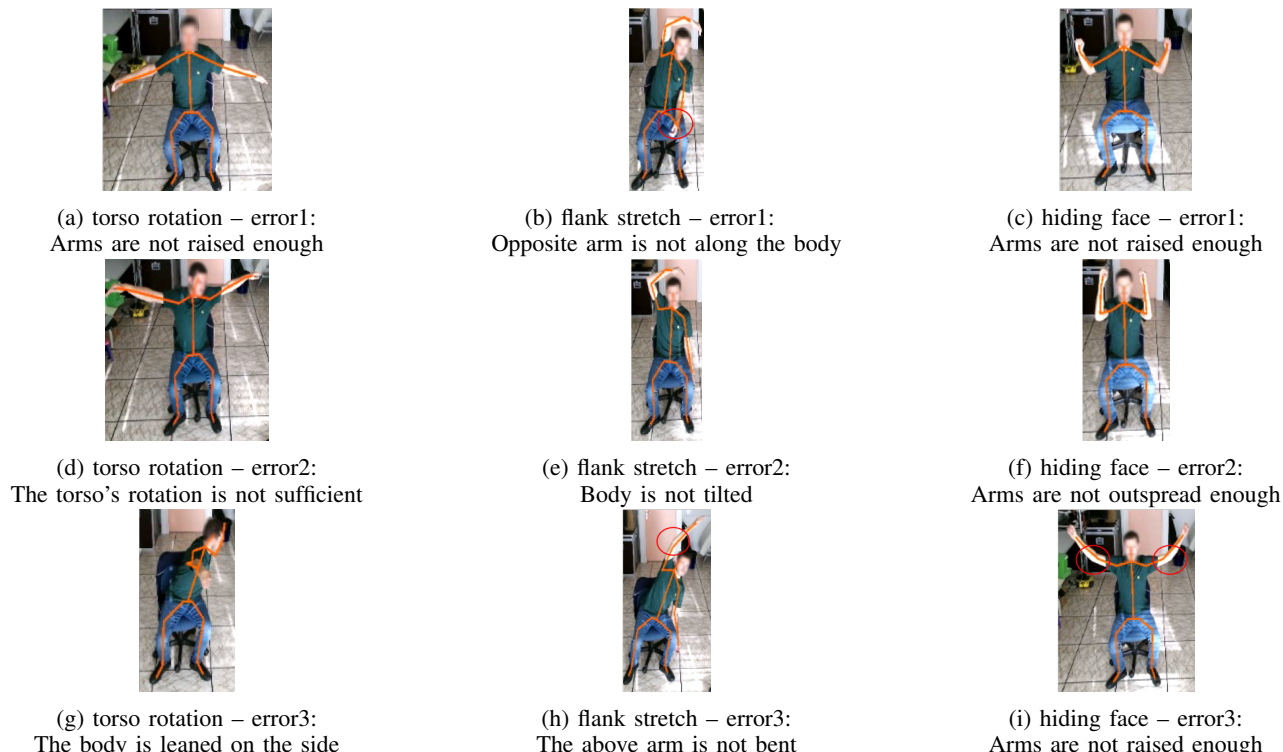


Fig. 2: Each column illustrates 3 errors for each exercise.

same instructions. They are free to execute the exercise correctly or with errors. As for group1, the recordings of the 6 healthy adults are split into two subsets:

- Group2a: 51 recordings per exercise were annotated as detailed in Sec. III-D.
- Group2b: the remaining recordings of the 6 healthy adults without annotation.
- Group3 : Healthy participants : Three healthy adults performed correct execution of exercises and simulated the identified common errors described in Sec. III-D.

C. Sensor system

Using the Microsoft Kinect V2 sensor, we obtained the RGB video with the skeleton drawn, and the skeleton joint positions and orientations information. From the RGB videos, we can also obtain additional estimation of joint positions and orientations using the human body keypoint detection libraries OpenPose [44] and BlazePose [25]. Moreover, as the Vicon system is considered the best system for precision, for Group3, we also recorded with MoCap using the Vicon system. For synchronisation purpose, the two systems were activated simultaneously.

Comparing the human pose estimation methods, [45] showed that OpenPose and BlazePose data lead to comparable performances of GMM to classify correct/incorrect exercises (challenge 1). Thus in the following, we report results with Kinect data.

D. Dataset annotation

The videos collected from patients with the Kinect V2 skeleton drawn were annotated by a medical doctor in physiotherapy and a physiotherapist, using the Anvil video annotation research tool³. The videos without blurred faces) are annotated at three levels related to the 4 challenges described in Sec. I. On a global evaluation level, an assessment is given as either correct or incorrect (Challenge 1). In the case of an incorrect error, they can indicate if the execution has no errors but finished before the end (label code 4: incomplete) or the participant did not start the execution of the exercise (label code 5 : motionless). On the error classification level, in the case of an incorrect movement, annotations first indicate whether the error is significant or small as well as the label of the error (Fig. 2) (Challenge 2). Moreover the body part causing the error is also indicated (Challenge 3). On a temporal level, the annotators can also indicate the time window where the error occurs (Challenge 4), and the same information as previously: whether the error is significant or small, the error label, and the body part causing the error. The annotation is carried out a frame level.

For challenge 1 to annotate a movement as correct or incorrect, we carried out an interannotator agreement analysis. The results show substantial agreement between the two medical annotators (Cohen's $\kappa = 0.63$ and Krippendorff's $\alpha = 0.62$ [46], [47]).

³<http://www.anvil-software.org>

E. Dataset description

Our dataset is composed of :

- The therapist’s annotations indicate whether the execution is correct, the label of the error, the body-part causing the error and the temporal description of the beginning and ending timestamps of the error.
- Anonymised RGB videos.
- The positions and orientations of each joint of the Microsoft Kinect V2 skeleton.
- The 2D positions of each joint of the OpenPose, Alphapose and Blazepose skeletons in the COCO pose or the 33 3D landmarks output format⁴.
- the positions and orientations of each joint of the Vicon skeleton.

We have summarized the data and the number of recordings available per participant group in Table II.

The Keraal dataset has a total of 2622 recordings, with 1881 recordings of patients and 741 recordings of healthy participants. Therefore, our number of patients recording already outnumbers the total number of recordings of all the other datasets, but the IRDS dataset.

IV. BENCHMARKS OF THE KERAAL DATASET

A. Two algorithms

In order to evaluate the dataset, we propose to assess the performance of human motion analysis algorithms in the context of rehabilitation. These baseline performances can serve as reference for future comparison with more sophisticated approaches. The first approach is a probabilistic approach uses a Gaussian Mixture Model (GMM) on a Riemannian Manifold [48], [49]. The second method is a deep learning-based approach employing a LSTM [50] as applied for motion analysis in [51], [52] as the main part of its architecture. We use a 3-layers LSTM with *tanh* activation followed by dropout layer as the core of two different architectures employed for rehabilitation motion assessment and rehabilitation motion recognition. For both configurations, we use the Adam optimizer with a learning rate of 0.01 and a batch size of 32 on a simple laptop. These two approaches are evaluated and compared for the challenges of rehabilitation motion assessment (Challenge 1) and rehabilitation motion recognition (Challenge 2). Position and orientation features of the upper body of the Kinect skeleton data are used in the following. For rehabilitation motion assessment, we employ our LSTM model within an autoencoder architecture in order to assess movements as correct or incorrect (Fig. 3a). For rehabilitation motion error classification, we add a fully-connected layer with *softmax* activation to train over 1000 epochs (Fig. 3b).

B. Rehabilitation motion assessment

The training data correspond to healthy subjects’ correct demonstrations (Group2 and Group3 cf. III-B), while testing data are patients’ performances acquired during rehabilitation

sessions. and labelled (Group 1a cf. III-B). The unlabelled data from patients (Group 1b) can be used for unsupervised or semi-supervised learning.

We compare the ability of the two baselines to detect incorrect motion sequences while only correct demonstrations are available during training. After determining by grid search the best threshold values for classification, Fig. 4 shows the detection results for the best F1-score of the GMM baseline and LSTM baseline, respectively. While colors represent normalized values, we left absolute values within the matrices so as to emphasize that classes are imbalanced.

Most of the sequences are classified as correct, as shown Fig. 4. The majority of incorrect motion sequences are misclassified as correct, as highlighted in red color. This is a critical point showing that the proposed baselines may not be appropriate for rehabilitation motion analysis as it does not allow patients to improve their performance. This suggests that selecting a lower threshold may allow the approaches to correctly detect errors with the cost of also misclassifying correct sequences as incorrect. Depending on the requirement, a trade-off can be chosen. To further facilitate the choice, we show in Fig. 6 the evolution of the number of true positives (correct motion) and true negatives (incorrect motion) with respect to the thresholds for both baselines.

Further, [53] studied the impact of the amount of training data on this performance for GMM and Spatio-Temporal Graph Convolutional Networks with this dataset.

C. Rehabilitation motion error classification

For the challenge of error classification, we aim to compare the two baselines for supervised classification. The goal is to classify an observed exercise among 4 classes corresponding to *correct*, *error1*, *error2* and *error3*, for each exercise separately. We evaluate the two baselines: 1) GMM-based features combined with a SVM classifier and 2) LSTM classifier. These two approaches are compared in term of classification accuracy. Classification results of GMM-based features combined with SVM classifier is reported in TableIII. For the LSTM classifier, as it may depend on initialization, we run the experiment 10 times and report mean accuracy with standard deviation. The best accuracy among the 10 runs is also reported.

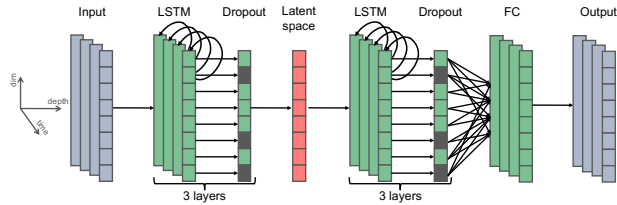
The LSTM classifier obtains much better accuracy for the three exercises. However, the latter approach obtains a maximum accuracy of 64.44% for the torso rotation exercise, showing that recognizing errors is still quite challenging.

From the confusion matrices in Fig. 8, we indeed notice that the majority of correct sequences are not recognized using the GMM-based features with SVM classifier in comparison to the LSTM classifier which is more accurate to recognize correct sequences. Nevertheless, the LSTM classifier is not able to efficiently recognize performed errors for all exercises. This especially explains the lower accuracy obtained for the flank stretch exercise in contrast to the two other exercises, as reported in TableIII. Indeed, for this exercise, fewer test sequences are annotated as correct by the medical expert. Moreover, for the two other exercises (torso rotation and hiding

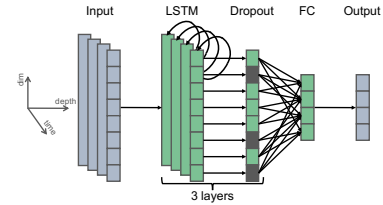
⁴https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/02_output.md

TABLE II: Available recording modalities per group and the format of the data

Gr.	Annotation	RGB videos	Kinect	Openpose/Blazepose	Vicon	Nb rec
1a	xml anvil : err label, bodypart, timespan	mp4, 480x360	tabular	dictionary	NA	249
1b	NA	mp4, 480x360	tabular	dictionary	NA	1631
2a	xml anvil : err label, bodypart, timespan	mp4, 480x360	tabular	dictionary	NA	51
2b	NA	mp4, 480x360	tabular	dictionary	NA	151
3	error label	avi, 960x544	tabular	dictionary	tabular	540

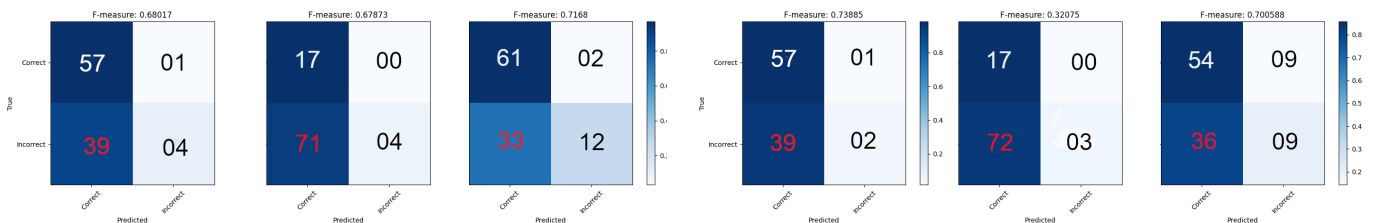


(a) Architecture of the LSTM-autoencoder



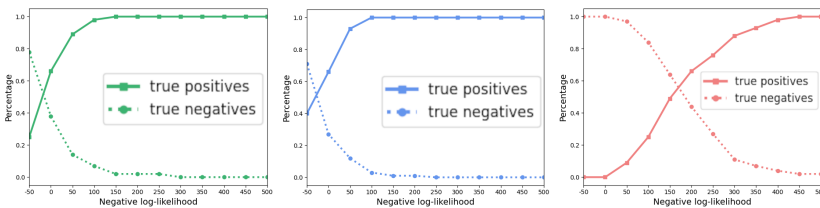
(b) Architecture of the LSTM classifier

Fig. 3: LSTM models for assessment and classification



(a) GMM torso rotation (b) GMM flank stretch (c) GMM hiding face (d) LSTM torso rotation (e) LSTM flank stretch (f) LSTM hiding face

Fig. 4: Confusion matrix of error detection using SVM and LSTM autoencoder.



(a) GMM torso rotation (b) GMM flank stretch (c) GMM hiding face

Fig. 5: GMM baseline : true positives and true negatives evolution with respect to different thresholds employed to evaluate a movement as correct or incorrect.

TABLE III: Accuracies of GMM-based SVM classifier and the LSTM classifier.

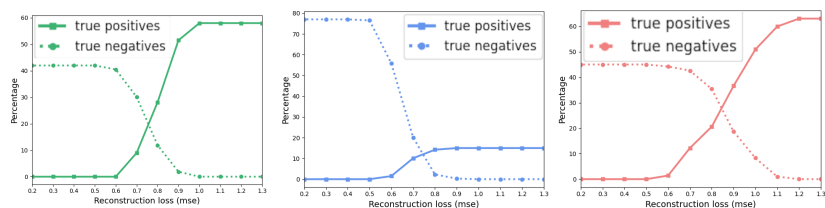
Exercise	GMM & SVM	LSTM	
	Accuracy	Accuracy (mean±std)	Best Accur.
torso rotation	27.78 %	53.89±4.82 %	64.44 %
flank stretch	25.32 %	31.64±6.48 %	43.03%
hiding face	33.33 %	49.1±3.28 %	56.19%

face), we observe a lot of confusion for the error2. As described in Fig.2, this error is related to the insufficient medial rotation (yaw) of arms. While it shows that the proposed baseline is not able to handle this medial rotation, it can also be explained by the precision of the skeleton data provided by Kinect.

V. CONCLUSION

In this paper, we formalized the challenges of automatic coaching of physical rehabilitation exercises from human pose movements as four folds, including motion assessment, error

classification, spatial and temporal localization. To benchmark the performances of machine learning algorithms with respect to these four challenges, we introduced the Keraal dataset for human body movement analysis in the context of low-back pain physical rehabilitation. This dataset has been acquired during a clinical study where patients were performing rehabilitation exercises. The limitations of this article are : the dataset includes a small number of exercises and patients, and the baseline algorithms have only standard performance. However,



(a) LSTM torso rotation (b) LSTM flank stretch (c) LSTM hiding face

Fig. 6: LSTM baseline : true positives and true negatives evolution with respect to different thresholds employed to evaluate a movement as correct or incorrect.

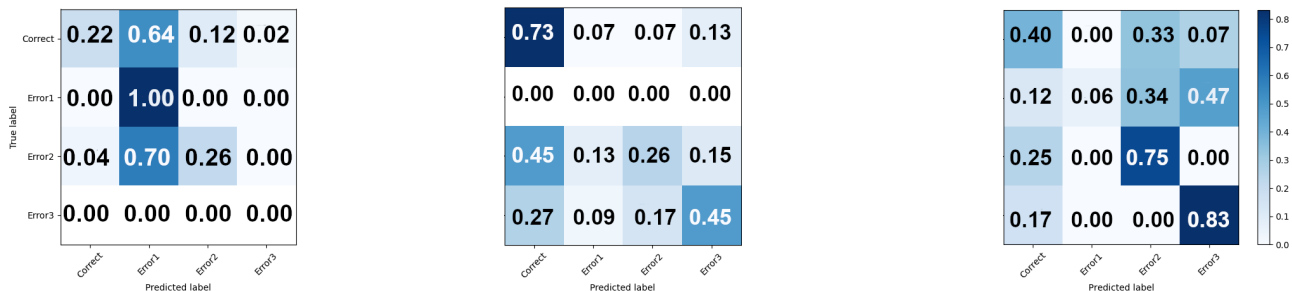
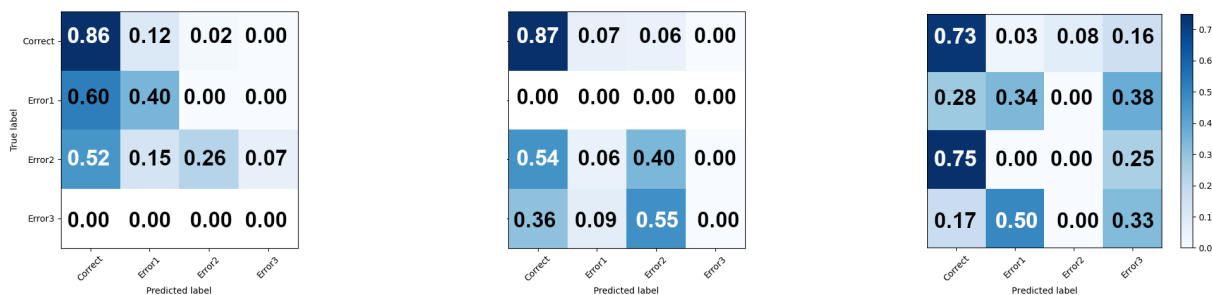


Fig. 7: Confusion matrices of rehabilitation movements classification using SVM-based features with SVM. Rows full of 0.00 mean that the corresponding error is not present in test set.



(a) LSTM torso rotation (b) LSTM flank stretch (c) LSTM hiding face

Fig. 8: Confusion matrices for rehabilitation movements classification using LSTM. Rows full of 0.00 mean that the corresponding error is not present in test set.

the dataset has detailed annotations. It includes 3D skeleton sequences captured by a Kinect, color video sequences and 2D skeleton data estimated from videos. Moreover, medical expert annotations are associated to each patient’s performance for assessment of correctness, recognition of errors, spatio-temporal localization of errors. The performance of the two proposed baselines show that the dataset is challenging enough to be a benchmark. We believe it can serve the research community of various fields from computer vision, machine learning, robotics, virtual agents, physical medicine and bio-mechanics, in the long run, allowing the patients to have limited access to perform rehabilitation movements on a regular basis.

In addition, to further facilitate its use, we evaluate and compare two baseline motion analysis algorithms, pertaining to two different approaches, for the tasks of rehabilitation motion assessment and error recognition. While the experiment introduces how to use the dataset, it also demonstrates that the targeted tasks are still challenging. This suggests that specific and more accurate methods should be designed so as to deeply assess rehabilitation movements and differentiate slight errors

from correct sequences. This latter investigation is part of our future work. Moreover, we also aim to investigate the challenges of spatial and temporal localization of errors along a motion sequence. Finally, we also want to extend the number of rehabilitation exercises considered by our dataset, as well as annotating more samples.

This dataset allows the development of better intelligent tutoring systems for physical rehabilitation and for physical exercises in general. Its social impact is to enable telemedicine and allow access to better exercising for those with difficult access to rehabilitation centers.

ACKNOWLEDGEMENT

This research was made possible by the support of the Hi! PARIS Engineering Team.

REFERENCES

[1] K. Mounce, “Back pain,” *Rheumatology*, vol. 41, pp. 1–5, 2002.
 [2] WHO, “The burden of musculoskeletal conditions at the start of the new millennium,” WHO, Geneva, Tech. Rep., 2003.

- [3] P. Kent and P. Kjaer, "The efficacy of targeted interventions for modifiable psychosocial risk factors of persistent nonspecific low back pain - a systematic review," *Man Ther.*, vol. 17, no. 5, pp. 385–401, Oct 2012.
- [4] G. Everard, A. Luc, I. Doumas, K. Ajana, G. Stoquart, M. G. Edwards, and T. Lejeune, "Self-rehabilitation for post-stroke motor function and activity—a systematic review and meta-analysis," *Neurorehabilitation and Neural Repair*, vol. 35, no. 12, pp. 1043–1058, 2021.
- [5] R. Gordon and S. Bloxham, "A systematic review of the effects of exercise and physical activity on non-specific chronic low back pain," *Healthcare*, vol. 4, no. 2, p. 22, Apr 2016.
- [6] D. Pasco, C. Bossard, C. Buche, and G. Kermarrec, "Using exergames to promote physical activity: A literature review," *Sport Science Review*, vol. 1-2, pp. 77–93, 2011.
- [7] G. Kermarrec, Y. Guillodo, D. Mutambayi, and L. Ballarin, "Mobile phones app to promote daily physical activity: Theoretical background and design process," *European Project Space on Computational Intelligence, Knowledge Discovery, and Systems Engineering for Health and Sports*. Scitepress: Rome, 2015.
- [8] T. Waltemate, F. Hülsmann, T. Pfeiffer, S. Kopp, and M. Botsch, "Realizing a low-latency virtual reality environment for motor learning," in *Proceedings of ACM Symposium on Virtual Reality Software and Technology (VRST)*, 2015.
- [9] K. Anderson, E. André, T. Baur, S. Bernardini, M. Chollet, E. Chrissyafidou, I. Damian, C. Ennis *et al.*, "The tardis framework: intelligent virtual agents for social coaching in job interviews," in *Advances in computer entertainment*. Springer, 2013, pp. 476–491.
- [10] J. Fasola and M. Mataric, "A socially assistive robot exercise coach for the elderly," *Journ. of HRI*, vol. 2, no. 2, 2013.
- [11] M. Devanne, S. M. Nguyen, A. Thépaut, O. Remy-Neris, B. L. G. Garnett, and G. Kermarrec, "A co-design approach for a rehabilitation robot coach for physical rehabilitation based on the error classification of motion errors," in *Proceedings of IEEE International Conference on Robotic Computing*, 2018. [Online]. Available: <https://ieeexplore.ieee.org/document/8329938>
- [12] M. Hun Lee, D. P. Siewiorek, A. Smailagic, A. Bernardino, and S. Bermúdez i Badia, "Design, development, and evaluation of an interactive personalized social robot to monitor and coach post-stroke rehabilitation exercises," *User Modeling and User-Adapted Interaction*, vol. 33, no. 2, pp. 545–569, Apr 2023.
- [13] R. Feingold-Polak, O. Barzel, and S. Levy-Tzedek, "A robot goes to rehab: a novel gamified system for long-term stroke rehabilitation using a socially assistive robot—methodology and usability testing," *Journal of NeuroEngineering and Rehabilitation*, vol. 18, no. 1, p. 122, 2021.
- [14] Y. Liao, A. Vakanski, M. Xian, D. Paul, and R. Baker, "A review of computational approaches for evaluation of rehabilitation exercises," *Computers in Biology and Medicine*, vol. 119, p. 103687, 2020.
- [15] A. Blanchard, S. M. Nguyen, M. Devanne, M. Simonnet, M. L. Goff-Pronost, and O. Rémy-Néris, "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*, vol. 2022, pp. 1–10, mar 2022.
- [16] R. A. Clark, Y.-H. Pua, K. Fortin, C. Ritchie, K. E. Webster, L. Denehy, and A. L. Bryant, "Validity of the microsoft kinect for assessment of postural control," *Gait & posture*, vol. 36, no. 3, pp. 372–377, 2012.
- [17] Y. Yang, F. Pu, Y. Li, S. Li, Y. Fan, and D. Li, "Reliability and validity of kinect rgb-d sensor for assessing standing balance," *IEEE Sensors Journal*, vol. 14, no. 5, pp. 1633–1638, 2014.
- [18] E. Stone and M. Skubic, "Evaluation of an inexpensive depth camera for in-home gait assessment," *Journal of Ambient Intelligence and Smart Environments*, vol. 3, no. 4, pp. 349–361, 2011.
- [19] M. Jebeli, A. Bilesan, and A. Arshi, "A study on validating KinectV2 in comparison of vicon system as a motion capture system for using in health engineering in industry," *Nonlinear Engineering*, vol. 6, no. 2, jan 2017. [Online]. Available: <https://doi.org/10.1515%2Fnleng-2016-0017>
- [20] G. Faity, D. Mottet, and J. Froger, "Validity and reliability of kinect v2 for quantifying upper body kinematics during seated reaching," *Sensors*, vol. 22, no. 7, 2022.
- [21] X. Xu, M. Robertson, K. B. Chen, J. hua Lin, and R. W. McGorry, "Using the microsoft kinect™ to assess 3-d shoulder kinematics during computer use," *Applied Ergonomics*, vol. 65, pp. 418–423, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0003687017300868>
- [22] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
- [23] Y. Xiu, J. Li, H. Wang, Y. Fang, and C. Lu, "Pose flow: Efficient online pose tracking," 2018. [Online]. Available: <https://arxiv.org/abs/1802.00977>
- [24] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5693–5703.
- [25] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "Blazepose: On-device real-time body pose tracking," *CoRR*, vol. abs/2006.10204, 2020.
- [26] R. A. Güler, N. Neverova, and I. Kokkinos, "Densepose: Dense human pose estimation in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7297–7306.
- [27] J. Zhen, Q. Fang, J. Sun, W. Liu, W. Jiang, H. Bao, and X. Zhou, "SMAP: Single-shot multi-person absolute 3d pose estimation," in *Computer Vision – ECCV 2020*. Springer International Publishing, 2020, pp. 550–566. [Online]. Available: https://doi.org/10.1007%2F978-3-030-58555-6_33
- [28] A. Benzine, F. Chabot, B. Luvison, Q. C. Pham, and C. Achard, "Pandant: Anchor-based single-shot multi-person 3d pose estimation," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020, pp. 6855–6864.
- [29] D. Leightley, M. H. Yap, J. Coulson, Y. Barnouni, and J. S. McPhee, "Benchmarking human motion analysis using kinect one: An open source dataset," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*. IEEE, 2015, pp. 1–7.
- [30] G. A. Nishiwaki, Y. Urabe, and K. Tanaka, "Emg analysis of lower extremity muscles in three different squat exercises," *Journal of the Japanese Physical Therapy Association*, vol. 9, no. 1, pp. 21–26, 2006.
- [31] I. Ar and Y. Akgul, "A computerized recognition system for the home-based physiotherapy exercises using an rgb-d camera," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, 05 2014.
- [32] 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, Jan 2018. [Online]. Available: <http://dx.doi.org/10.3390/data3010002>
- [33] E. Dolatabadi, Y. X. Zhi, B. Ye, M. Coahran, G. Lupinacci, A. Mihailidis, R. Wang, and B. Taati, "The toronto rehab stroke pose dataset to detect compensation during stroke rehabilitation therapy," in *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, ser. PervasiveHealth '17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 375–381. [Online]. Available: <https://doi.org/10.1145/3154862.3154925>
- [34] A. D. Miron, N. M. Sadawi, W. Ismail, H. Hussain, and C. Grosan, "Intellirehabs (irds) - a dataset of physical rehabilitation movements," *Data*, vol. 6, p. 46, 2021.
- [35] M. S. H. Aung, S. Kaltwang, B. Romera-Paredes, B. Martinez, A. Singh, M. Cella, M. Valstar, H. Meng, A. Kemp, M. Shafizadeh, A. C. Elkins, N. Kanakam, A. de Rothschild, N. Tyler, P. J. Watson, A. C. d. C. Williams, M. Pantic, and N. Bianchi-Berthouze, "The automatic detection of chronic pain-related expression: Requirements, challenges and the multimodal empain dataset," *IEEE Transactions on Affective Computing*, vol. 7, no. 4, pp. 435–451, Oct 2016.
- [36] M. Capecchi, M. G. Ceravolo, F. Ferracuti, S. Iarlori, A. Monteriù, L. Romeo, and F. Verdini, "The kimore dataset: Kinematic assessment of movement and clinical scores for remote monitoring of physical rehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 7, pp. 1436–1448, July 2019.
- [37] C. Wu, J. Zhang, S. Savarese, and A. Saxena, "Watch-n-patch: Unsupervised understanding of actions and relations," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 2015.
- [38] H. S. Koppula and A. Saxena, "Learning spatio-temporal structure from rgb-d videos for human activity detection and anticipation," in *Int. Conf. on Machine Learning*, 2013.
- [39] A. Shahroudy, J. Liu, T.-T. Ng, and G. Wang, "Ntu rgb+ d: A large scale dataset for 3d human activity analysis," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 2016, pp. 1010–1019.
- [40] C. Liu, Y. Hu, Y. Li, S. Song, and J. Liu, "Pku-mmd: A large scale benchmark for continuous multi-modal human action understanding," *arXiv preprint arXiv:1703.07475*, 2017.
- [41] S. Stein and S. J. McKenna, "User-adaptive models for recognizing food preparation activities," in *Proceedings of the 5th international workshop on Multimedia for cooking & eating activities*, 2013, pp. 39–44.
- [42] M. Tenorth, J. Bandouch, and M. Beetz, "The TUM Kitchen Data Set of Everyday Manipulation Activities for Motion Tracking and Action

- Recognition,” in *IEEE International Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS)*, in conjunction with ICCV2009, 2009.
- [43] D. Roggen, A. Calatroni, M. Rossi, T. Holleczeck, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha *et al.*, “Collecting complex activity datasets in highly rich networked sensor environments,” in *2010 Seventh international conference on networked sensing systems (INSS)*. IEEE, 2010, pp. 233–240.
- [44] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, “Realtime multi-person 2d pose estimation using part affinity fields,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7291–7299.
- [45] A. Marusic, S. M. Nguyen, and A. Tapus, “Evaluating kinect, openpose and blazepose for human body movement analysis on a low back pain physical rehabilitation dataset,” in *Companion of the ACM/IEEE International Conference on Human-Robot Interaction*. New York, NY, USA: Association for Computing Machinery, March 2023, pp. 587–591.
- [46] K. Krippendorff, *Content Analysis An Introduction to Its Methodology (4th Edition)*. SAGE Publications, Inc, 2018.
- [47] J. Cohen, “A coefficient of agreement for nominal scales,” *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960. [Online]. Available: <https://doi.org/10.1177/001316446002000104>
- [48] J. Jost, *Riemannian Geometry and Geometric Analysis*, ser. Springer Universitat texts. Springer, 2005.
- [49] M. J. Zeestraten, I. Havoutis, J. Silvério, S. Calinon, and D. G. Caldwell, “An approach for imitation learning on riemannian manifolds,” *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1240–1247, 2017.
- [50] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [51] S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao, “Independently recurrent neural network (indrnn): Building a longer and deeper rnn,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5457–5466.
- [52] J. Liu, G. Wang, P. Hu, L.-Y. Duan, and A. C. Kot, “Global context-aware attention lstm networks for 3d action recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1647–1656.
- [53] A. Marusic, L. Annabi, S. M. Nguyen, and A. Tapus, “Analyzing data efficiency and performance of machine learning algorithms for assessing low back pain physical rehabilitation exercises,” in *European Conference on Mobile Robots*, ACM/IEEE, Ed., 2023.

APPENDIX : ARTICULATED FIGURES

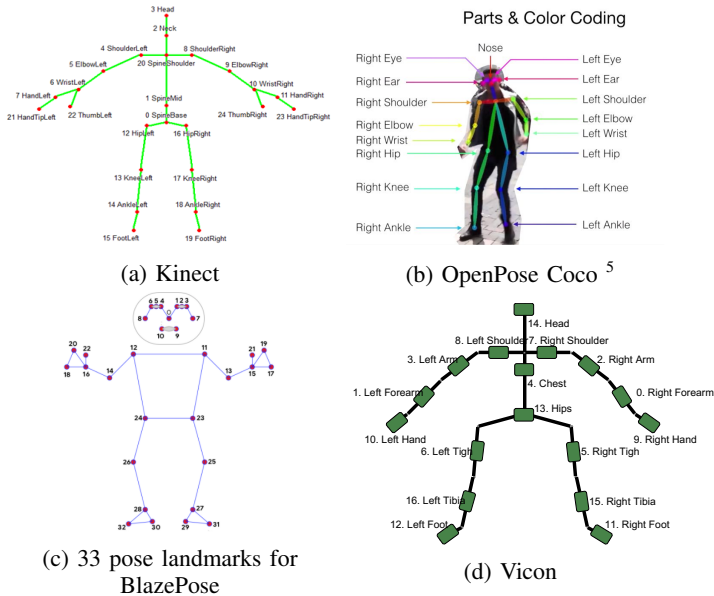


Fig. 9: Pose (Skeleton) output formats of the Kinect, OpenPose , BlazePose and Vicon.