Skeleton-Based Transformer for Classification of Errors and Better Feedback in Low Back Pain Physical Rehabilitation Exercises

Aleksa Marusic, Sao Mai Nguyen, Adriana Tapus

Marusic, A., Nguyen, S. M., and Tapus, A. (2025). Skeleton-Based Transformer for Classification of Errors and Better Feedback in Low Back Pain Physical Rehabilitation Exercises. International Conference on Rehabilitation Robotics.

Skeleton-Based Transformer for Classification of Errors and Better Feedback in Low Back Pain Physical Rehabilitation Exercises

Aleksa Marusic[1], Sao Mai Nguyen[1] and Adriana Tapus[1]

***Abstract*— Physical rehabilitation exercises suggested by healthcare professionals can help recovery from various musculoskeletal disorders and prevent re-injury. However, patients' engagement tends to decrease over time without direct supervision, which is why there is a need for an automated monitoring system. In recent years, there has been great progress in quality assessment of physical rehabilitation exercises. Most of them only provide a binary classification if the performance is correct or incorrect, and a few provide a continuous score. This information is not sufficient for patients to improve their performance. In this work, we propose an algorithm for error classification of rehabilitation exercises, thus making the first step toward more detailed feedback to patients. We focus on skeleton-based exercise assessment, which utilizes human pose estimation to evaluate motion. Inspired by recent algorithms for quality assessment during rehabilitation exercises, we propose a Transformer-based model for the described classification. Our model is inspired by the HyperFormer method for human action recognition, and adapted to our problem and dataset. The evaluation is done on the KERAAL dataset, as it is the only medical dataset with clear error labels for the exercises, and our model significantly surpasses state-of-the-art methods. Furthermore, we bridge the gap towards better feedback to the patients by presenting a way to calculate the importance of joints for each exercise.**

I. INTRODUCTION

Physical rehabilitation plays a crucial role in helping patients recover from injuries, surgeries, and various medical conditions that affect their movement and functional abilities. Low back pain (LBP), which is the focus of our research, is a leading cause of disability globally, affecting over 50% of the population at some point in their lives. Consequently, healthcare providers face substantial challenges in managing the increasing number of LBP patients [1]. The success of rehabilitation exercises depends significantly on how accurately and consistently they are performed. Throughout the weeks or even months of a rehabilitation program, patients often have to carry out exercises at home without direct oversight from healthcare professionals. The absence of supervision and timely feedback from healthcare providers diminishes patient engagement in rehabilitation. Additionally, incorrect movements can not only delay the rehabilitation process but also increase the risk of further injuries [2]. Thus, there is a critical need for accurate and automated movement analysis methods to support patients and clinicians in monitoring rehabilitation exercises.

*This work was supported by ENSTA Paris

[1]Autonomous Systems and Robotics Lab, Computer Science and System Engineering (U2IS), ENSTA Paris, Institut Polytechnique de Paris, 828 Blvd des Maréchaux, 91120 Palaiseau, France, name.surname@ensta-paris.fr

The primary objective of automated monitoring systems for physical rehabilitation is to identify the activity being performed, evaluate its quality, and offer detailed information on errors and potential improvements. Current movement analysis methods in rehabilitation typically provide a general quality score that reflects how well an exercise is executed [3], [4], [5] While these scoring systems provide useful feedback, they represent only the initial step toward a fully automated and practical monitoring system.

A more detailed analysis, which not only evaluates the overall quality of movement but also identifies and localizes specific errors, would be highly beneficial for both patients and clinicians. This detailed feedback can help patients adjust their movements more precisely, keep them motivated throughout the rehabilitation sessions, and allow clinicians to more accurately monitor patients' progress during in-home rehabilitation [6], [7].

Human activity analysis is an evolving field that tackles the complex challenge of interpreting body movements based on data collected from various sources such as videos, images, or wearable sensors. While using data from wearable sensors directly as input is the dominant approach, recently, human pose estimation to estimate joint coordinates from videos has become a promising approach to analyze human movement. Given that skeletons can be naturally modeled as graphs, it is not surprising that Graph Convolutional Networks (GCNs) have become a leading method for skeleton-based action recognition. By representing the body as a graph of joints and their spatial-temporal relationships, skeleton-based methods effectively capture movement dynamics and remain resilient to changes in appearance and environmental conditions, while capturing spatial and temporal relationships during exercise performance [8], [9].

Furthermore, Transformer models, originally developed for natural language processing tasks, have shown remarkable success in various applications, such as recognition of daily activities in smart homes [10]. Their ability to model long-range dependencies and capture contextual information makes them well-suited for analyzing sequential data, such as skeleton sequences in rehabilitation exercises [11].

For automated feedback allowing physical rehabilitation patients to improve their performance, this paper offers two key innovations: error classification and movement analysis. Unlike previous methods that focus solely on providing a quality score, our approach requires a more precise model, thus we utilize a skeleton-based transformer model. Through self-attention mechanism, our model is able to better learn spatial and temporal relations between skeleton

joints. Specifically, our model is designed to:

- Classify errors: Identify and categorize different types of errors that may occur during the performance of rehabilitation exercises.
- Biomechanical attention: Identify the most important joints in human body skeleton for a specific movement, thus providing better feedback to the patient.

The remainder of this paper is organized as follows: Section II reviews related work on movement analysis in rehabilitation and skeleton-based action recognition models. After outlining the data we used in Section III, Section **??** describes the architecture of our skeleton-based transformer model and the methods for error classification and localization. Section V details our experimental setup and results, and discusses the implications of our findings. Finally, Section VI summarizes our main contributions, acknowledges the limitations of our approach, and suggests potential directions for future research..

II. RELATED WORK

A. Skeleton-based Action Recognition

Skeleton-based action recognition is a dynamic and rapidly evolving research area. Early studies depended on handcrafted features that utilized relative 3D rotations and translations between joints [12], [13]. However, the field has experienced significant advancements in recent years, largely driven by Deep Learning algorithms [8]. These Deep Learning methods for skeleton-based action recognition can be categorized into three primary groups, based on their approaches to extracting features from skeleton data for classification:

- Recurrent Neural Networks (RNNs), which consider skeleton data primarily as a temporal sequence of continuous features. Some of the examples are [14] that introduced a hierarchical recurrent neural network, and [15] that proposed Context-Aware Attention LSTM Networks.
- Convolutional Neural Networks (CNNs), [16], [17] apply CNNs on pseudo-images obtained from skeleton data [16], [17]. In this way, they capture spatial relations between joints in a frame.
- Graph Neural Networks (GNNs): Skeleton data naturally corresponds to a graph structure, with joints as vertices and bones as edges. As a result, Graph Convolutional Networks have become increasingly popular as they can extract both spatial relationships and be combined with temporal data [18], [19].

B. Physical Rehabilitation Assessment

Qualitative exercise assessment is essential for effective home-based rehabilitation systems, aiming to provide patients with informative feedback and enhance their performance. Early research on exercise evaluation applied traditional machine learning methods for classification, such as Adaboost, K-Nearest Neighbors (KNN), or Bayesian classifiers. Some approaches also utilized distance function-based models [20], [21]. Later studies adopted probabilistic models like Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs) [6], [22]. While these models capture the stochastic nature of human motion and provide a quality score for movement accuracy, they do not fully exploit given information, such as joint or spatial connections between body parts.

Liao et al. [23] introduced a deep neural network model that generates quality scores for movements. They proposed a deep learning architecture for hierarchical spatio-temporal modeling, combining GMMs, CNNs, and LSTMs to compute a quality score. With the advent of Graph Neural Networks (GNNs), it is now possible to exploit spatial information through the skeletal graph structure. The authors in [24] and [25] applied Graph Convolutional Networks (GCNs) to assess physical rehabilitation, achieving state-of-the-art results on popular datasets such as KIMORE and UI-PRMD. Additionally, Yu et al. [26] employed an ensemble of two GCNs, one for position and one for orientation features of the skeletal joints, to further improve performance.

While existing methods primarily focus on quality scoring, a few studies have explored error classification and localization. For example, dynamic time warping (DTW) was used in [27] to compare patient movements with reference movements and identify phases with significant deviations. Devanne et al. [7] added on top of DTW, a GMM model to analyse the likelihood per body part and per time segment to localize the error. However, its accuracy remains modest. On the other hand, our approach leverages the power of Deep Learning and Transformer models to capture the spatial-temporal dynamics of movements and provide actionable feedback with error classification and localization.

In the following sections, we describe our methodology in detail, including the architecture of our skeleton-based transformer model and the approaches for error classification and movement analysis.

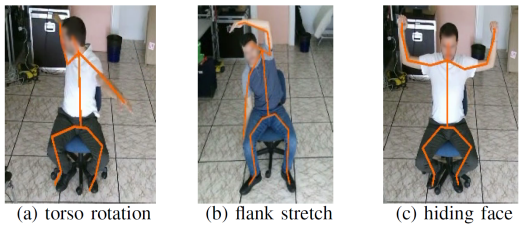
III. DATA PROCESSED

A. Dataset

We evaluate our model on the Keraal dataset [28] collected during a clinical trial [29], as it is the only one of the available rehabilitation datasets that fits into our problem setting and has error labels. TRSP [30] is the only other rehabilitation dataset, to our knowledge, with error labels, but they only use two types of simple reaching motions.

In the Keraal dataset, participants performed each of 3 predefined exercises, and the movements were recorded with Microsoft Kinect V2 and Vicon Recordings are annotated by physiotherapists in detail, with assessment of correctness, recognition of errors, and spatio-temporal localization of errors. In our model we used Kinect data as input, since it is a non-invasive system we would like to focus on.

The three exercises, chosen in agreement with medical experts, are namely torso rotation, flank stretch, and hiding face, which are illustrated in Figure 1. For each exercise therapists identified three most common errors, and each labeled recording was classified either as correct or as one of these 3 errors. As can be seen on Figure 2, the common



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

Fig. 1: The three rehabilitation exercises in the Keraal dataset. Image sourced from [28].

Fig. 2: Errors descriptions for all three exercises in the Keraal dataset. Image sourced from [28].

mistakes for exercise (a) torso rotation are: 1) Arms are not raised enough, 2) The torso's rotation is not sufficient, 3) The body is leaned on the side; for exercise (b) flank stretch: 1) Opposite arm is not along the body, 2) Body is not tilted, 3) The above arm is not bent; and for exercise (c) hiding face: 1) Arms are not raised enough, 2) Arms are not outspread enough, 3) Arms are not raised enough.

The Keraal dataset was recorded during a clinical rehabilitation study that included Low Back Pain patients, aged 18 to 70 years. The dataset includes recordings from both healthy subjects and 12 rehabilitation patients, from three groups of participants:

- Group 1 where patients performed exercises
- Group 2 where healthy participants performed exercises without specific intructions
- Group 3 where trained healthy participants performed exercises while simulating errors.

Groups 1 and 2 were labeled by physiotherapists, while Group 3 was implicitly labeled as participants simulated errors.

B. Data representation

The data of the Keraal dataset recorded exercises at a fixed frame rate, capturing the spatial information of key skeletal joints in the human body, either in two or three dimensions. The data include joint positions and sometimes their orientations (in our case we use only 3D joint positions).

Consider a tracking system that monitors V key points (joints) on a person's body. When someone performs one repetition of an exercise, the system records data across a certain number of time frames. Since that number can vary from recording to recording even for the same exercise type, we interpolate each recording to T timeframes. For each frame, we collect the positions of all V joints in a vector $x(i), i \in [1 \dots T]$. This vector has the dimension $D = V \times C$, where C equals either 2 or 3, depending on whether we are working with 2D or 3D data. When we combine these vectors from all frames, we create a three-dimensional tensor $X \in \mathbb{R}^{T \times V \times C}$, which represents the complete movement data from a single repetition.

C. Evaluation scenarios

Our goal is to classify a performed movement into one of four categories: correct, error1, error2, or error3. To achieve this, we train a separate model for each exercise. The primary metrics used for evaluation are the F1 score and accuracy.

We consider three cases:

- Scenario 1: Initial scenario where we train the models on trained professionals simulating errors (group 3), and evaluate on healthy subjects and patients (groups 2A and 1A). This setup is critical as it demonstrates the model's performance on unseen data, closely simulating real-life application conditions.
- Scenario 2: In the second scenario, data from all three groups are combined and split into training and testing sets, in proportion 80:20. Also, it is important to note that the split is stratified, meaning the proportion of class labels is approximately the same in both training and test splits. This second scenario highlights the model's full potential by showing the level of performance achievable with sufficient data for both training and validation.
- Scenario 3: The scarcity of labels for certain exercises in the first scenario prompted us to design a third scenario. We still wanted to have unseen patient data in the test set to simulate real-life conditions, but we added a portion of healthy participants in the test set as well, so that we better evaluate our model. In this setup, the test set is made from patient data (group 1) and 15% of the combined data from groups 2 and 3, while the training set consists of the remaining 85% of the healthy participants' data (split is stratified in this case as well).

IV. Algorithm

A. Self-attention

Self-attention in Transformers was originally developed to identify and map relationships between words in a text, whether they are close together or far apart in a sentence. This mechanism helps the network prioritize important parts of the input data [31], [32].

We can adapt this same concept to analyze human movement patterns captured through skeletal tracking. Just as Transformers process relationships between words in a text, we can process relationships between body joints in motion. Individual joint positions can be seen as words and each

frame of movement as a sentence. By applying the self-attention mechanism, we can more effectively learn both local (between nearby joints or sequential frames) and distant relationships (between far-apart joints or temporally distant frames) in the motion sequence.

Consider a given input sequence $X = (x_1, x_2, \ldots, x_n)$ where each x_i represents an input token. Each token x_i is transformed into three vectors: Query vector q_i, Key vector k_i, and Value vector v_i. These vectors are obtained by applying learned projection matrices to each token, transforming them into new representations suited for computing attention. Essentially, these projections allow the model to learn more complex spatial relations of the input data.

For each token x_i, its attention score is computed relative to every other token in the sequence x_j, including itself. The attention score is calculated by applying a softmax function on the dot product A_{ij} between the query vector q_i of token i and the key vector k_j of token j:

$$A_{ij} = q_i \cdot k_j^{\top} \tag{1}$$

To capture more complex relationships, an extension called Multi-Head Self-Attention (MHSA) is added [32]. With multiple attention heads, the model learns different kinds of spatial relations between the joints, in parallel.

B. Model architecture

The main processing idea is inspired by a novel algorithm - Hyperformer [33]. It has achieved state-of-the-art results on well-known human action recognition (HAR) benchmarks, proving it can learn complex relations between joints in the movement. However, unlike HAR datasets, we are not interested in distinguishing different types of actions, but different types of errors in one specific action. Thus, we are looking for much more subtle differences in motions during the same actions and trying to classify each motion accordingly. We train one model per exercise (action), compared to HAR algorithms which train one model in total to classify actions.

Furthermore, we have only a few classes (for errors), and even more importantly, very limited medical data to train the model, which emphasizes the need for a better understanding of spatio-temporal relations between the joints. That was another reason to choose Hyperformer as our basis, since it is one of the smallest HAR algorithms available (by number of trainable parameters).

One of the key novelties of Hyperformer is utilizing hypergraphs - dividing the initial skeleton graph into subgraphs in order to obtain more precise relations between the joints. In our paper we propose a different split of the skeleton graph, more tailored to our problem setting. We split 25 joints into 6 different groups: left and right forearm and hand (with wrists and fingers), legs and spine with head, as can be seen on the left part of Figure 3. We give particular importance to arms as they are one of the key parts in all exercises from the dataset used, as opposed to legs which do not move much during the exercises.

The overall model architecture can be seen on the right part of Figure 3 The input data is passed through a certain

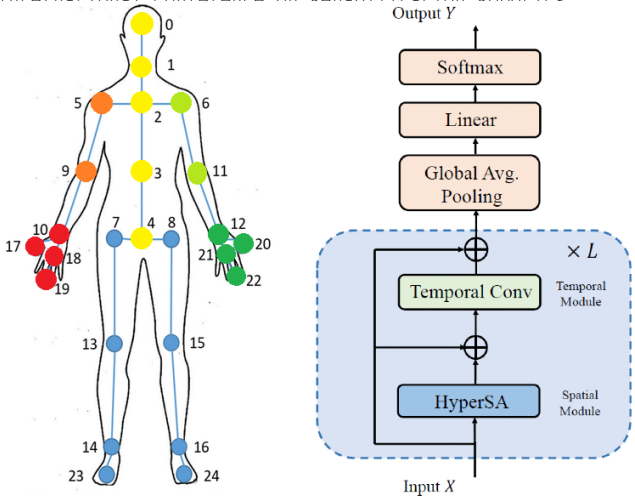


Fig. 3: Groups of skeletal joints (hypergraphs) on the left and model overview on the right. Right part taken from [33]

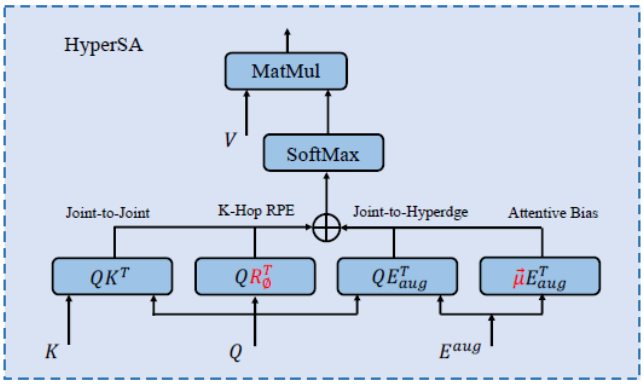


Fig. 4: Overview of the hyper self-attention module [33].

number of layers (10 in our case). Each layer consists of a custom self-attention layer followed by Multi-Scale Temporal Convolution. Global pooling, a linear activation and softmax functions are added at the end. The model is trained via backpropagation through a cross-entropy loss function by comparing $\hat{\mathbf{y}}$ with the true label $\mathbf{y}$ - a class labeled by the physiotherapists.

The main part of the layer is a custom self-attention module, as depicted in Figure 4. It consists of 4 parts :

- Joint-to-joint attention. Projection matrices used to get q, k, v vectors are obtained with 2D convolutions. Attention is calculated as described in IV-A.
- Joint-to-subgraph attention. Learns relations between joints and subgraphs, calculated the same way
- Relative positional embedding. Incorporates structural information of the skeleton, obtained with the Shortest Path Distance between the joints
- Attentive bias. Assigns the same amount of attention to each joint that belongs to a certain subgraph, enforcing independence of the query position.

V. RESULTS & DISCUSSION

In this section, we present the results of comprehensive comparative experiments conducted to assess the perfor-

mance of our model. We provide a comparison between our proposed approach and the current state-of-the-art, as well as some guidelines for further improvement.

A. Implementation details

Python 3.10 and the PyTorch 2.3 + CUDA 12.1 framework were used to develop the model. The system is equipped with an Intel Core i9-9900KF Silver 4215R CPU, 32GB of RAM, and an Nvidia RTX 2080 Ti 11GB. The model was trained with the Adam optimizer for 600 epochs, with a learning rate set to 25×10^{-4}, and using a batch size of 10. Performance measures were recorded for each run and averaged to ensure accuracy. Our code is based on the Hyperformer code repository and is available at `https://github.com/aleksamarusic/hyperphysio/`

B. Comparison with state-of-the-art

Table I shows that out model reaches higher accuracies in Scenario 1, outperforming the models benchmarked in [28].

TABLE I: Accuracies in percentage

Exercise	Ours	LSTM best	LSTM mean	GMM
Torso rotation	**73.17**	64.44	53.89	27.78
Flank stretch	**64.10**	43.04	31.64	25.32
Hiding face	**74.28**	56.19	49.1	33.33

Additionally, we present confusion matrices for the classification of rehabilitation movements, comparing three described scenarios

The confusion matrices for the first case are presented in Figure 5, with true labels on the left and predicted labels on the bottom. The values in the matrices represent the percentage of exercises classified as a certain class. For instance, in Figure 5a, the first row indicates that 71% of "Correct" exercises are classified as correct, 28% as error2, and 1% as error3. Consequently, the sum of any row in these matrices should be 1. Exceptionally, rows with all zeros mean that the corresponding error is not present in the test set. The value in brackets next to the left-side labels denote the frequency of the corresponding class in the test set.

For the torso rotation exercise, it is very hard to evaluate the model performance as we have 72 "correct" labels, only 4 of error1 and 6 of error2, while error3 is not present in the test set.

Similarly, the scarcity of labels for errors 1 and 3 in the flank stretch exercise poses significant challenges for accurately evaluating the model's performance. Error2 exhibits some confusion with correct exercises, which can be explained by the general challenge to determine that the body is not tilted enough solely based on joint positions.

For hiding face, error2 has only 1 example in the test set, and we need more to properly evaluate. Also, it can be noticed that error3 shows misclassification with both correct and error1 categories. These can be also attributed to a general challenge as errors 1 and 3 are quite similar, and both are hard to distinguish from the correct performance as it is only a matter of a few pixels on the image whether they would be classified as one or the other. These issues could potentially be mitigated with the inclusion of additional training and validation data, and with more precise data.

Confusion matrices for the second case can be seen in Figure 6. As explained, we randomly sample 20% of the dataset for testing, keeping the proportion of the labels in the test set. This scenario obtained better performance, which confirms the model's potential to perform significantly better when having more training data.

The confusion matrices for the third case are shown in Figure 7. While the majority of movements are correctly classified, some confusion can be observed between "correct" and "error3" in the flank stretch exercise . This may be attributed to the limited number of sequences labeled as correct by the medical expert and the subtle differences between correct movements and error2. The latter occurs when the body is not sufficiently tilted, a distinction that can be challenging to detect solely through skeleton joint data.

C. Visualization of joints' importance

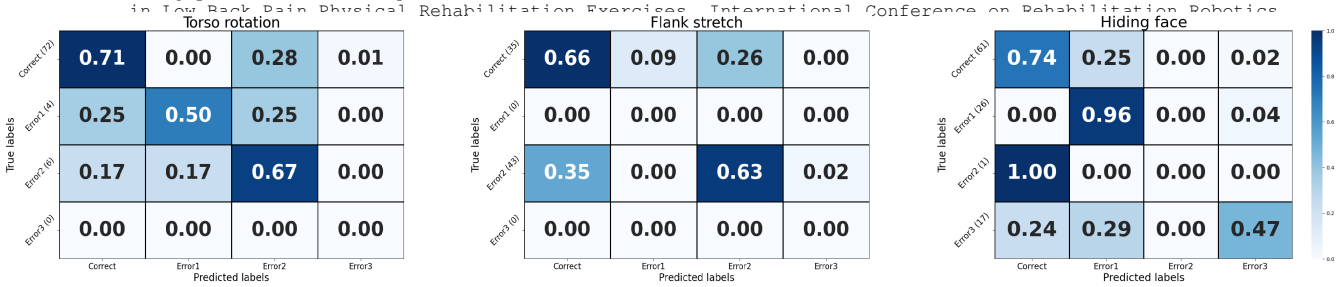
The described visualization aims to analyze and compare the attention weights learned during the training and evaluation phase. By analyzing attention matrices, we can identify important joints in each exercise.

From our model, we can obtain the self-attention weights, which are quite comprehensive as we have weights for every layer, frame, and head (of multi-head attention) from every batch and input sequence. However, by averaging across these dimensions, we can obtain self-attention map that shows the attention weights for each body joint to every other body joint (since we have 25 joints, it is represented as $25 * 25$ matrix). Further, the joint role can be computed by the column-wise summation over that map. Additionally, we also notice that is important to look at differences in self-attention maps obtained from correct and incorrect exercises. This way we can more easily obtain information on which joints play an important role in the exercise.

Looking at Figure 8, we can notice that arm joints play an important role in all exercises, contrarily to the legs, as the exercises relate to the upper body. Additionally, Torso rotation and Hiding face give more importance to the angles of the shoulders and elbows, contrarily to Flank stretch. This is expected as the two former exercises require keeping the upper arm horizontal, while Flank stretch leaves the arm free, focusing more on the flank. These observations align with key joint patterns seen in the exercises (Figure 1), confirming the ability of the model to provide accurate and useful feedback to patients.

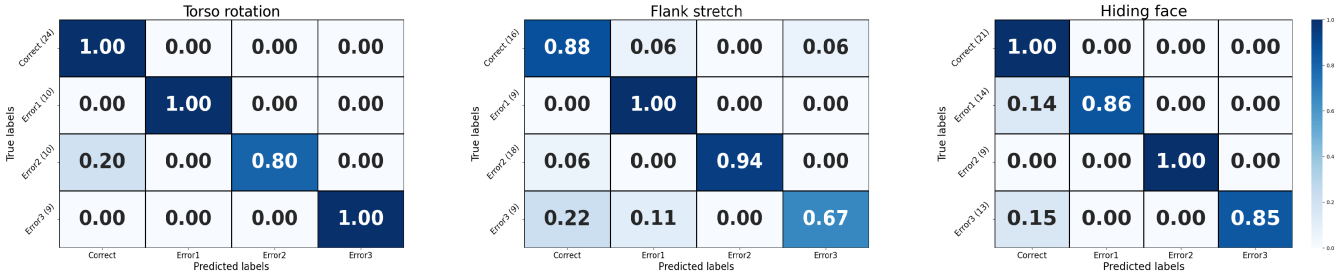
VI. CONCLUSION

In this paper, we present a Transformer-based model for classifying errors in physical rehabilitation exercises. The model takes skeleton joint positions as input and classifies movement as either correct or one of the predefined errors. Inspired by the Hyperformer algorithm, our approach divides the skeleton graph into subgraphs to enhance the learning



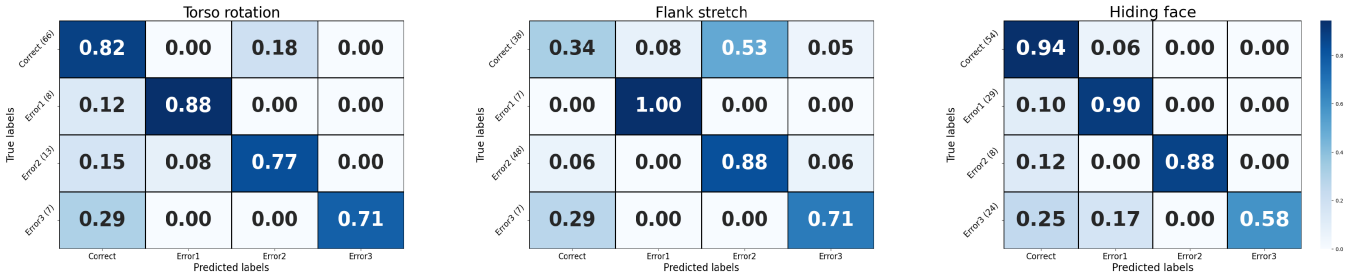
(a) Torso rotation confusion matrix (b) Flank stretch confusion matrix (c) Hiding face confusion matrix

Fig. 5: Confusion matrices for each of 3 exercises for the first case (trained on group3 and tested on groups 2A and 1A). Rows full of 0.00 mean that the corresponding error is not present in the test set.



(a) Torso rotation confusion matrix (b) Flank stretch confusion matrix (c) Hiding face confusion matrix

Fig. 6: Confusion matrices for each of 3 exercises for the second case (all groups combined and then data split into train and test sets).



(a) Torso rotation confusion matrix (b) Flank stretch confusion matrix (c) Hiding face confusion matrix

Fig. 7: Confusion matrices on test for each of 3 exercises for the third case (trained on most 85% healthy (groups 2 & 3) and tested on patients (group 1) and 15% of groups 2 & 3 combined.

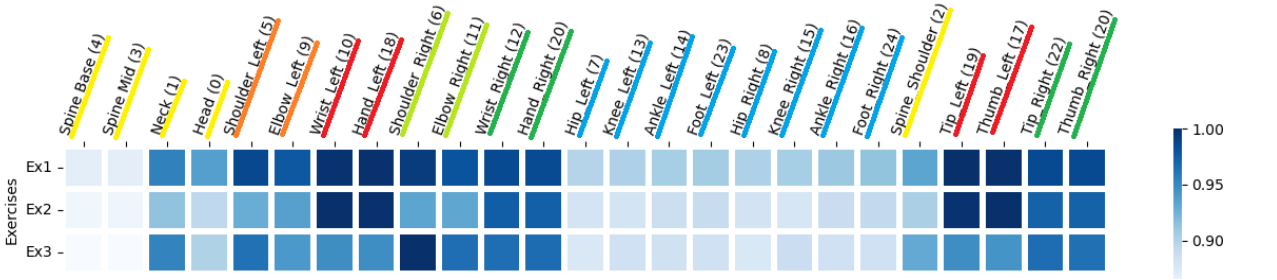


Fig. 8: An illustration of the joints' importance calculated from attention weights. Each row represents importance values for each of 25 skeleton joints during a specific exercise. Ex1 : Torso rotation, Ex2 : Flank stretch, and Ex3 : Hiding face. Each column corresponds to a joint, designated by its name and its joint order (numbers in brackets next to labels), that are underlined with a color matching its group from Fig 3

of complex spatial joint relations using the self-attention mechanism. We evaluate our model on the Keraal dataset, focusing on low back pain rehabilitation, which is the only dataset suitable for this error classification task. We can draw several conclusions:

- Our model achieves significant improvements over benchmark algorithms and sets new state-of-the-art results. Our model better captures spatial and temporal

relations with the self-attention module, and sets a new direction for automated rehabilitation assessment.

- We also calculate the importance of specific joints in performed exercises. This information can be used to detect which joints are wrongly moved, leading to enhanced feedback provided to the patients.
- Since we were interested in subtle differences in motions during the same exercise, it was necessary to train one model for each exercise type. In a scenario in which patients are trained by a coach, the system knows which exercise is being performed and the focus was on the assessment method. A more comprehensive model integrating exercise detection, can be a future research direction.
- One of the key limitations lies in the limited data available. While the Keraal dataset has detailed annotations, medical datasets are unfortunately limited in number.
- The imbalance of the dataset makes it harder to evaluate the model for underrepresented classes. Exploring data augmentation methods could lead to better classification and is a potential research direction for further studies.
- Some of the further research directions are temporal localization of errors, as well as more precise instructions on how to fix errors. These are essential for qualitative feedback to a patient.

REFERENCES

[1] A. Wu, L. March, X. Zheng, J. Huang, X. Wang, J. Zhao, F. M. Blyth, E. Smith, R. Buchbinder, and D. Hoy, "Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the global burden of disease study 2017," *Annals of Translational Medicine*, 2020.

[2] S. F. Bassett and H. Prapavessis, "Home-based physical therapy intervention with adherence-enhancing strategies versus clinic-based management for patients with ankle sprains," *Physical Therapy*, 2007.

[3] S. Sardari, S. Sharifzadeh, A. Daneshkhah, B. Nakisa, S. W. Loke, V. Palade, and M. J. Duncan, "Artificial intelligence for skeleton-based physical rehabilitation action evaluation: A systematic review," *Computers in Biology and Medicine*, 2023.

[4] 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 *2023 European Conference on Mobile Robots (ECMR)*, 2023.

[5] Y. Mourchid and R. Slama, "D-stgcnt: A dense spatio-temporal graph conv-gru network based on transformer for assessment of patient physical rehabilitation," *Computers in Biology and Medicine*, vol. 165, p. 107420, 2023.

[6] M. Devanne and S. M. Nguyen, "Multi-level motion analysis for physical exercises assessment in kinaesthetic rehabilitation," in *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, 2017.

[7] M. Devanne, S. M. Nguyen, O. Remy-Neris, B. Le Gales-Garnett, G. Kermarrec, and A. Thepaut, "A co-design approach for a rehabilitation robot coach for physical rehabilitation based on the error classification of motion errors," in *IEEE International Conference on Robotic Computing (IRC)*, Jan 2018, pp. 352–357.

[8] V.-T. Le, K. Tran-Trung, V. T. Hoang, and H. Chen, "A comprehensive review of recent deep learning techniques for human activity recognition," *Intell. Neuroscience*, Jan. 2022.

[9] J. Shin, N. Hassan, A. S. M. Miah1, and S. Nishimura, "A comprehensive methodological survey of human activity recognition across divers data modalities," 2024.

[10] D. Bouchabou, J. Grosset, S. M. Nguyen, C. Lohr, and X. Puig, "A smart home digital twin to support the recognition of activities of daily living," *Sensors*, 2023.

[11] W. Xin, R. Liu, Y. Liu, Y. Chen, W. Yu, and Q. Miao, "Transformer for skeleton-based action recognition: A review of recent advances."

[12] R. Vemulapalli, F. Arrate, and R. Chellappa, "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.

[13] M. E. Hussein, M. Torki, M. A. Gowayyed, and M. El-Saban, "Human action recognition using a temporal hierarchy of covariance descriptors on 3d joint locations," ser. IJCAI 2013.

[14] Y. Du, W. Wang, and L. Wang, "Hierarchical recurrent neural network for skeleton based action recognition," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[15] J. Liu, G. Wang, L. yu Duan, K. Abdiyeva, and A. C. Kot, "Skeleton-based human action recognition with global context-aware attention lstm networks," *IEEE Transactions on Image Processing*, vol. 27, pp. 1586–1599, 2017.

[16] Q. Ke, Bennamoun, S. An, F. Sohel, and F. Boussaïd, "A new representation of skeleton sequences for 3d action recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[17] M. Liu, H. Liu, and C. Chen, "Enhanced skeleton visualization for view invariant human action recognition," *Pattern Recognition*, 2017.

[18] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," ser. AAAI'18/IAAI'18/EAAI'18. AAAI Press, 2018.

[19] F. Ye, S. Pu, Q. Zhong, C. Li, D. Xie, and H. Tang, "Dynamic gcn: Context-enriched topology learning for skeleton-based action recognition," ser. MM. ACM, 2020.

[20] I. Ar and Y. S. Akgul, "A computerized recognition system for the home-based physiotherapy exercises using an rgbd camera," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 6, pp. 1160–1171, 2014.

[21] R. Houmanfar, M. Karg, and D. Kulić, "Movement analysis of rehabilitation exercises: Distance metrics for measuring patient progress," *IEEE Systems Journal*, vol. 10, 2016.

[22] A. Vakanski, J. Ferguson, and S. Lee, "Mathematical modeling and evaluation of human motions in physical therapy using mixture density neural networks," *Journal of Physiotherapy & Physical Rehabilitation*.

[23] Y. Liao, A. Vakanski, and M. Xian, "A deep learning framework for assessing physical rehabilitation exercises," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2020.

[24] S. Deb, M. F. Islam, S. Rahman, and S. Rahman, "Graph convolutional networks for assessment of physical rehabilitation exercises," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 410–419, 2022.

[25] C. Du, S. A. Graham, C. A. Depp, and T. Q. Nguyen, "Assessing physical rehabilitation exercises using graph convolutional network with self-supervised regularization," *International Conference of the IEEE Engineering in Medicine & Biology Society*.

[26] B. X. Yu, Y. Liu, X. Zhang, G. Chen, and K. C. Chan, "Egcn: An ensemble-based learning framework for exploring effective skeleton-based rehabilitation exercise assessment," in *Proceedings of International Joint Conference on Artificial Intelligence 2022*.

[27] X. Chen, Y. Liu, and Q. Huang, "Real-time error detection and feedback for physical rehabilitation exercises using kinect and dtw," 2016.

[28] S. M. Nguyen, M. Devanne, O. Remy-Neris, M. Lempereur, and A. Thepaut, "A medical low-back pain physical rehabilitation dataset for human body movement analysis," in *International Joint Conference on Neural Networks*, 2024.

[29] 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.

[30] 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," 2017.

[31] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *CoRR*, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:11212020

[32] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017.

[33] Z. Hu, V. Gutiérrez-Basulto, Z. Xiang, R. Li, and J. Z. Pan, "Hyperformer: Enhancing entity and relation interaction for hyper-relational knowledge graph completion," 2023.