LightPRA: A Lightweight Temporal Convolutional Network for Automatic Physical Rehabilitation Exercise Assessment

Sardari, S., Sharifzadeh, S., Daneshkhah, A., Loke, S. W., Palade, V., Duncan, M. J. & Nakisa, B.

DOI 10.1016/j.compbiomed.2024.108382
ISSN 0010-4825
ESSN 1879-0534

Publisher: Elsevier

Computers in Biology and Medicine 173 (2024) 108382

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

LightPRA: A Lightweight Temporal Convolutional Network for Automatic Physical Rehabilitation Exercise Assessment

Sara Sardari [a,b,*], Sara Sharifzadeh [c], Alireza Daneshkhah [a,d], Seng W. Loke [b], Vasile Palade [a], Michael J. Duncan [e], Bahareh Nakisa [b]

[a] Research Centre for Computational Science and Mathematical Modelling, Coventry University, Coventry, UK
[b] School of Information Technology, Faculty of Science Engineering and Built Environment, Deakin University, Geelong, Vic, Australia
[c] Department of Computer Science, Swansea University, Swansea, UK
[d] School of Mathematics and Data Science, Emirates Aviation University, Dubai, United Arab Emirates
[e] Centre for Sport, Exercise and Life Sciences, Coventry University, Coventry, UK

ARTICLE INFO

Keywords:
Activity evaluation
Dilated convolutions
Temporal Convolutional Network
Telerehabilitation
Skeleton data

ABSTRACT

Research evidence shows that physical rehabilitation exercises prescribed by medical experts can assist in restoring physical function, improving life quality, and promoting independence for physically disabled individuals. In response to the absence of immediate expert feedback on performed actions, developing a Human Action Evaluation (HAE) system emerges as a valuable automated solution, addressing the need for accurate assessment of exercises and guidance during physical rehabilitation. Previous HAE systems developed for the rehabilitation exercises have focused on developing models that utilize skeleton data as input to compute a quality score for each action performed by the patient. However, existing studies have focused on improving scoring performance while often overlooking computational efficiency. In this research, we propose LightPRA (Light Physical Rehabilitation Assessment) system, an innovative architectural solution based on a Temporal Convolutional Network (TCN), which harnesses the capabilities of dilated causal Convolutional Neural Networks (CNNs). This approach efficiently captures complex temporal features and characteristics of the skeleton data with lower computational complexity, making it suitable for real-time feedback provided on resource-constrained devices such as Internet of Things (IoT) devices and Edge computing frameworks. Through empirical analysis performed on the University of Idaho-Physical Rehabilitation Movement Data (UI-PRMD) and KInematic assessment of MOvement for remote monitoring of physical REhabilitation (KIMORE) datasets, our proposed LightPRA model demonstrates superior performance over several state-of-the-art approaches such as Spatial–Temporal Graph Convolutional Network (STGCN) and Long Short-Term Memory (LSTM)-based models in scoring human activity performance, while exhibiting lower computational cost and complexity.

1. Introduction

Prescribing physical rehabilitation exercises is essential for physically disabled individuals due to different medical conditions such as strokes, surgeries, and injuries [1], or deterioration of muscle mass and bone density as a consequence of aging [2]. The process of monitoring patients' movement involves tracking the prescribed activities performed by patients and assessing their recovery level by medical experts. The new concept of telerehabilitation [3], which involves providing monitoring services remotely, holds immense promise for both medical experts and patients. It allows the experts to reach patients facing barriers, like those living in rural or under-served areas. In addition, it allows medical experts to efficiently manage their time by conducting remote sessions and providing service for a larger number of patients. Moreover, in a study conducted on aging people [4], travel time and cost are reported as the primary barriers to rehabilitation program participation. The possibility of performing exercises from home can reduce the cost and time of patient treatment [5]. However, the importance of accurate feedback and guidelines following the monitoring cannot be overstated since it directly impacts the success and influence of rehabilitation on individuals [6]. By providing accurate feedback and regular check-ins, the medical experts can guarantee the continuity of the rehabilitation procedure for patients, including those receiving care at home [7,8]. Automation of the monitoring phase plays a significant role in the effectiveness and efficiency of the rehabilitation procedure by assisting the expert in enhancing the quality of feedback [6,9,10]. Therefore, the advantages of automated movement evaluation in the

* Corresponding author at: Research Centre for Computational Science and Mathematical Modelling, Coventry University, Coventry, UK.
E-mail address: sardaris@uni.coventry.ac.uk (S. Sardari).

https://doi.org/10.1016/j.compbiomed.2024.108382
Received 13 November 2023; Received in revised form 22 February 2024; Accepted 24 March 2024
Available online 25 March 2024

context of rehabilitation include enhancing objective action analysis (due to the precise data collection of different sensors rather than only the human eye), providing frequent and consistent assessment for the patients, early detection of suboptimal progress or deviation from correct activity, facilitating telerehabilitation and remote consultation, and finally promoting adherence to a rehabilitation program.

Automatic movement monitoring often involves computer vision techniques and AI-based models for tracking and analyzing the captured data from a patient's movement. The computer vision techniques for action monitoring include different modalities such as skeleton data, InfraRed (IR) sequences, depth data, and RGB images [6]. Skeleton data, among these vision-based data capturing methods, has gained significant interest among researchers due to factors such as privacy preservation and cost effectiveness [6]. The advent of Kinect and Vicon sensors as methods for capturing skeleton data, along with the utilization of diverse Deep Learning (DL)/ Machine Learning (ML) algorithms as the HAE pipelines, have underscored their potential as adjunct decision-making tools for medical experts. The HAE for rehabilitation problem can be addressed in two ways: classification or regression. In the former approach, each activity performed by the patient is classified as correct or incorrect, overlooking the degree of deviation from the reference correct action and possible improvements in performing the activity. The latter approach predicts continuous numerical scores for each action, providing a finer assessment of the degree of correctness. Performing traditional ML algorithms for regression, which utilize hand-crafted feature extraction (such as extracting relative or projected trajectory) [11–13] encounters some drawbacks. They might require an expert's knowledge or problem-specific algorithms, which hinders generalization and increases preprocessing cost [14,15]. Therefore, DL algorithms that encompass automatic feature learning align better with the complexity of the HAE problem and are considered superior solutions. Different DL techniques have been employed in healthcare-related studies illustrating their effective performance within this scope [16–19].

Considering the spatial–temporal nature of both orientational and positional data, different DL models have been proposed, including CNN-based, Graph Convolutional Network (GCN)-based, and Recurrent Neural Network (RNN)-based architectures. These models are designed to capture either spatial or temporal features of the data for activity assessment. Leveraging computational resources, they process and train on skeleton data. Nevertheless, a common challenge faced by these models is finding a balance between scoring performance and computational cost. For instance, when handling temporal data such as skeleton sequences, LSTM-based models are a natural consideration. However, there are several shortcomings in this approach [20,21] such as losing long-term information, and high memory and time costs for training the model. Therefore, there is a high chance of inferior scoring performance, coupled with a substantial sacrifice of computational resources. To address the challenge of suboptimal scoring performance, attention-based models can be employed that focus on the most discriminative features; however, it is worth noting that this may lead to an increase in computational resource usage. With the continuous expansion of resource-constrained technologies such as IoT devices [22] and Edge Computing [23], and the potential deployment of this HAE application on such devices, there is a growing need for an HAE system that effectively manages scoring performance and computational costs. This means that in the future, healthcare providers and patients as users of the system will need accurate and real-time feedback for the activities, they perform from the system employed on their simple device. Given the newness of the HAE issue in rehabilitation and the swift advancement of resource-constrained technologies such as IoT devices, it seems that prior research may have overlooked this problem and its associated remedy.

To address the aforementioned shortcomings (lack of balance between computational complexity and scoring performance) more efficiently, we introduce a dilated causal convolutional architecture that performs automatic feature extraction of the body skeleton using the TCN pipeline. According to Meng et al. [24], TCNs not only have longer memory in sequence modeling compared to LSTMs, but they also perform large-scale parallel processing like CNNs. The parallel nature of convolutions makes TCNs well-suited for real-time applications with time and computational power constraints [25]. Moreover, TCNs are intricately designed to handle sequential data, ensuring commendable performance despite computational constraints. This illustrates their potential to be used on any kind of processor like Central Processing Units (CPU) or even the ones with parallel computational architectures like Graphics Processing Units (GPU) and Field Programmable Gate Arrays (FPGAs) [26]. In the context of deploying diverse DL frameworks for IoT applications, which frequently struggle with limited computational resources and the critical constraints of energy efficiency, TCNs emerge as a fitting solution. Their compatibility with constrained environments aligns well with the requirements established earlier. TCNs on FPGAs use their inherent parallelism to accelerate the computational time and reduce latency. Running the TCN on energy-efficient FPGAs can reduce power consumption by optimizing power usage on battery-powered IoT devices [27]. The fewer trainable parameters and less Random Access Memory (RAM) usage in TCNs make them well-suited for resource management in IoT devices. In general, the characteristics mentioned above add to the computational power of the TCNs in the training and inference phase, which makes them suitable for real-time training and providing feedback on resource-constrained devices, Edge computing, and IoT devices.

As noted by Sardari et al. [6], the domain of physical rehabilitation movement analysis encounters a scarcity of publicly available datasets. This scarcity arises from concerns related to patient privacy and ethical considerations. Due to the low participation rate of elderly and physically disabled people in exercise-related data collection studies, some studies require healthy participants to perform both healthy and simulated unhealthy exercises [28–30]. Among the remaining datasets, there are very few ones targeting a general population of patients with the target of score prediction and regression. Due to these reasons, KIMORE [31] and UI-PRMD [28] datasets have become more popular in recent activity evaluation studies. Two important aspects regarding these datasets are required to be mentioned here. First, the KIMORE and UI-PRMD datasets are captured through distinct sensing technologies, namely Kinect and Vicon, respectively. These sensors yield different levels of accuracy in skeleton data capturing. As detailed by Sardari et al. [6], earlier versions of Kinect sensors (which are used in the KIMORE) are noted for their relatively lower accuracy, producing noisy data compared to Vicon sensors. The analysis provided in our and prior studies indicates a noticeable reduction in scoring accuracy for the data captured by the Kinect sensor. The UI-PRMD dataset [28] includes positions and angles of body joints [14]. The KIMORE [31] includes both 3D joint position and joint orientation data. The joint orientation data captured in the KIMORE dataset is represented as their quaternion rotations with respect to the spine base. According to previous studies [32] The usage of quaternions compared to other approaches of 3D rotation representation such as Cardan and Euler angles is more robust and compact, due to avoiding mathematical singularities and gimbal lock issues. Moreover, the orientational data has spatial angular and rotational information of the joint points during activity and can be considered better than positional data for the following reasons: (1) Joint orientation is robust to changes in body scale, environment, and shifting camera angles. (2) Joint orientation provides richer information about the angle of body joints during dynamic and complex activities [32]. It encapsulates the spatial inter-dependencies between the body joints, encoding them into joint angular information. Therefore, each joint orientation represents a signal inherently embedded with spatial information, leaving the task of learning its temporal features to the model. Therefore, in this study, we investigated the joint orientation and positions as the input data for further comparison and analysis.

The key contributions of this study are summarized as follows:

- In this study, we proposed a novel fully TCN-based architecture, namely LightPRA, specifically tailored for processing the temporal features in sequential joint data. Leveraging the distinctive characteristics of the causal and dilated convolutions, our model excels in automatic action scoring. To the best of our knowledge, this is the first study conducted on applying a fully TCN-based architecture for HAE in the context of rehabilitation applications.
- We conduct comprehensive experiments on different types of data (joint orientation and joint position), to compare the scoring performance of the proposed LightPRA model to state-of-the-art methods. On average, the derived results demonstrate that our proposed model is as effective as the strong graph-based models in the UI-PRMD [28] dataset. Moreover, it surpasses the accuracy of scoring achieved by previous studies in the KIMORE [31] dataset.
- To evaluate the model's affordability, the computational time and complexity are compared to previous approaches. The results show that the proposed LightPRA architecture significantly reduces the training and inference time. This holds significant value in applications involving adapting and training the model for new patients and exercises. In addition, the inference time on the testing experiments demonstrates the superiority of the LightPRA model over other methods in providing real-time feedback. The reduced RAM consumption and time complexity in the proposed TCN-based pipeline highlight its potential in resource-constrained devices and Edge computing environments.

It is worth mentioning that the implementation of the preprocessing phase and the proposed method is publicly available at https://github.com/SaraS92/LightPRA. The remainder of this paper is structured as follows. Section 2 presents the related work. Section 3 includes the proposed method in detail. Section 4 presents the results and discussion. Finally, Section 5 concludes the study and suggests potential future work.

2. Related work

In this section we briefly review approaches related to our, i.e., DL-based work for rehabilitation exercise evaluation, and TCN-based models previously proposed for different applications. According to the previous literature [6] in the domain of rehabilitation exercises for HAE, studies exhibit a wide range of objectives, contributing to significant diversity that complicates comparison. In addition, Because of its novelty, the field presents considerable potential for further investigation and the discovery of unexplored areas. Due to this factor, only a limited number of studies have explored the UI-PRMD [28] and KIMORE [31] datasets for evaluating their DL-based models.

In 2020, Liao et al. [33] proposed a DL framework for HAE validated on the UI-PRMD dataset [28]. The framework includes an autoencoder for dimensionality reduction, a Gaussian mixture model (GMM) for automatic label generating, and, most importantly, a hierarchical CNN–LSTM-based model for extracting spatial–temporal features and movement quality assessment. First, LSTM models are prone to losing long-term information in long temporal sequences, though they were previously introduced to remove the vanishing gradient problem of RNNs. In addition, LSTMs have shown longer training time, resulting in high computation costs due to the sequential nature of the model. In addition, LSTMs are prone to consume a significant amount of memory to keep the partial outputs for their multiple-cell gates. In another study conducted by Deb et al. [14], the authors proposed an STGCN architecture augmented with a self-attention mechanism to provide better scoring performance. However, incorporating LSTM and self-attention layers within this pipeline presents a significant challenge, primarily due to the increased associated computational costs. According to Dao et al. [34], the self-attention module exhibits time and memory complexity that grows quadratically with the length of the sequence. Finally, to solve the problem of time and memory complexity, the self-attention modules need parallel processing, which can typically be handled by GPUs. Training these architectures often experience substantial improvements in performance through GPU acceleration, due to the GPUs' inherent parallel processing capabilities. Their adeptness at managing extensive matrix computations inherent in self-attention mechanisms leads to heightened efficiency, ultimately resulting in accelerated training and inference duration. However, to date, the literature has not yet explored solutions with high scoring performance, tailored for resource-constrained devices lacking high memory and GPU access.

As emphasized in the previous section, considering the efficient resource utilization of TCN and its proven success in the analysis of temporal data, TCN presents itself as a promising alternative. Li et al. [35] utilized a combination model of a TCN and Gated Recurrent Unit for accurate fall detection. They highlighted the resulting architecture as an enriching classifier capable of delivering strong performance even with limited motion information. Several other studies have explored various versions of these networks for tasks such as human action segmentation and detection [36,37]. In the study conducted by Lea et al. [38], the authors utilized a TCN-based Encoder–Decoder for capturing the long-range temporal patterns Following dilated TCNs for action segmentation and detection. They have illustrated that TCNs can capture long-range dependencies, and they are faster than LSTM models in the training phase. In a 2023 study [39] that explored the combination of TCN models and Transformers for temporally aware surgical workflow recognition, the significance of the TCN model in capturing temporal information was emphasized as crucial for the model's superior performance. Sabater et al. [40] introduced a TCN-based model for one-shot activity recognition in the context of therapy for autistic patients, achieving high performance with the lightweight model. Due to the remarkable performance demonstrated by TCNs, we proposed the TCN-based framework for HAE in rehabilitation applications.

3. Proposed method

This section presents the proposed LightPRA pipeline as a human activity evaluation technique for rehabilitation exercises. First, it includes an analysis of the two public rehabilitation datasets of UI-PRMD [28] and KIMORE [31] utilized in the evaluation of different methods in Section 3.1. We discuss the description and preprocessing stages and mention the observed challenges regarding each dataset. Then, in Section 3.2 we offer a detailed description of the LightPRA, our proposed TCN-based model, highlighting the characteristics that make it well-suited for temporal data analysis, particularly tailored for the HAE task.

3.1. Dataset description and preprocessing

UI-PRMD dataset: One of the public datasets utilized in this study is The UI-PRMD dataset [28] published in 2018, which is publicly available at the website provided by the University of Idaho (www.webpages.uidaho.edu/ui-prmd). The dataset includes 10 general rehabilitation exercises with 10 repetitions performed by 10 healthy individuals in both correct and incorrect manner (mimicking patients). The exercises consist of deep squats (E1), hurdle step (E2), inline lunge (E3), side lunge (E4), sit-to-stand (E5), standing active straight leg raise (E6), standing shoulder abduction (E7), standing shoulder extension (E8), standing shoulder internal–external rotation (E9), and standing shoulder scaption (E10). Vicon motion-capturing cameras as accurate sensors for capturing exact joint positions and orientations [6,41] were utilized to capture skeletal data (including joint coordinates and angles). The information captured by Vicon in this dataset illustrates 117 joint displacements of the body skeleton in different time steps. After acquiring the skeleton data, it is essential to perform the preprocessing. In this stage, firstly the spatial and temporal alignment and centering of the skeleton data are performed. After preprocessing the data and

excluding some noisy data, our dataset includes the following: 180 samples for each exercise (90 correct movements and 90 incorrect movements), 240 time-steps (number of frames), and 117 joint displacements, which are reduced to 90 displacements after excluding head joint displacements. The multi-channel signal of human body movement for the first training sample of deep squat action of this dataset is shown in Fig. 1. It illustrates the orientations of the joints in the trunk, left arm, right arm, left leg, and right leg through 240 frames. It can be seen that the left leg and right leg had more changes over time for this activity.

To automatically generate scores (labels) for each movement performed by the participants, the GMM as a statistical model is trained on the correct movements of the specific exercise [33]. Then the negative log-likelihood of the trained model is utilized to measure the performance quality of each movement. Finally, a scoring function is defined, which maps the performance metric values into a series of scores ranging between 0 and 1. This method of scoring is already compared with other methods of scoring like Euclidean distance and Dynamic Time Warping (DTW) [42] distance in the study conducted by Liao et al. [33] and the authors concluded that this scoring method results in better action assessment. All of these preprocessing stages have been considered the same as the previous studies. To enhance comprehension of data, we performed an analysis for the scores distribution of some exercises as depicted in Fig. 2. The plots illustrate that the concentration of scores is primarily in the range of 0.7 to 0.95 despite the expected range of 0 to 1. This observation notes that even for mimicked unhealthy movements, the majority of score distribution is in the higher range. In addition, the score distribution depicts the outliers, which deviate significantly from the rest of the data (see Fig. 2 parts A, B, and C, where samples from the lowest ranges are in the lowest frequency). The problem of outliers can be addressed by considering better evaluation metrics for the models that are less sensitive to outliers. The non-uniform scoring and outliers in some exercises can introduce some inconsistency to the HAE models in scoring accuracy.

KIMORE dataset: The KIMORE dataset introduced by Capecci et al. [31] (available in https://vrai.dii.univpm.it/content/kimore-dataset) is one of the recent studies providing a higher number of participants. The participants include 44 healthy and 34 unhealthy individuals performing 5 repetitions of 5 exercises specific for back pain rehabilitation. The physical exercises include lifting the arms up (EX1), Lateral tilt of the trunk with the arms in extension (EX2), Trunk rotation (EX3), Pelvis rotations on the transverse plane (EX4), Squatting (EX5). Kinect V2 was utilized to capture joint positions and orientations data and we utilized both of them for comparison. The orientation data includes quaternions (X, Y, Z, W) of rotation as a representation of body movement, where (X, Y, Z) represents the vector part of the quaternion (the axis of rotation) and W represents the scalar part (the amount of rotation). For preprocessing the orientation data, the continuous actions (including several repetitions of one action) were segmented into distinct repetitions. For each action and participant, some of the orientation files were not captured, and they were not considered. Finally, three repetitions from the remaining participants are considered as the samples. This resulted in 213, 183, 189, 210, and 207 samples for the exercises EX1, EX2, EX3, EX4, and EX5, respectively. The 3D position data has previously been segmented and captured in studies such as [14], which we leverage in our research. For both position data and orientational data we performed zero-mean preprocessing and a Gaussian filter to feed proper data to the models. The actual activity scores are from 0 to 50, which need rescaling to either [0,1] or [−1,1]. We considered both in the results and discussion section.

To better compare sensor accuracy, we provided the preprocessed signals of the squat action performed in the KIMORE dataset for the trunk, left arm, right arm, left leg, and right leg through 104 frames as illustrated in Fig. 1. Comparing each row in the columns (a) and (b) in Fig. 1 demonstrates the use of noisy sensors such as Kinect V2 (in KIMORE dataset) instead of Vicon (in UI-PRMD [28]) dataset can introduce noise to the data, impacting the performance of the model. Due to the nature of the activity, squats should include more orientation in the leg joints and trunk and less movement around the arm joints. Although this fact is shown perfectly in the samples from the UI-PRMD dataset, the noisy samples in KIMORE can hinder the model from learning discriminative features. In addition, the scoring distribution in the KIMORE dataset may introduce confusion to the model during training. The scores for all of the repetitions of the same activity for the same individual are equal. However, joint movements can involve varying degrees of freedom in the same action performed by the same person due to different reasons, including changes in the environment, or fatigue. This critical factor is illustrated in Fig. 3 (panel B), where the left hip joint movement is shown for all of the repetitions of squats one participant performs. This issue can add challenges to the HAE model in providing generalized scores for similar activities, resulting in inconsistent results for different exercises.

3.2. LightPRA model for action evaluation

In this subsection, we propose a novel deep regression model based on TCNs for automated action evaluation. The model effectively learns the complex relationship between activity data, including body movements and their corresponding scores generated for each movement. TCNs are well-suited for this task as they can learn long-range dependencies in the data. A graphical representation of the proposed model is shown in Fig. 4. In this study, we leveraged the joint orientation and position data, rich with inherently embedded spatial angular and rotational features varying over time. Therefore, it can be processed with a temporal model that intricately considers its temporal characteristics.

Traditional CNN architectures are widely used in different areas, including image processing, where they can effectively extract local features and spatial hierarchies [43]. However, they are not typically considered suitable for temporal data processing due to the size constraint of the convolutional kernel and the disregard of temporal features of adjacent time-steps [44]. To address this problem, a deep TCN architecture is introduced by adapting simple convolutional layers to sequential data in a way that leverages the strengths of both CNN and ResNET (Residual Networks) architectures, which perform special kernels across the time axis. TCN accomplishes this task through the following three distinguishing features: (1) **Causal convolutions**: This ensures that the output at time t only depends on the inputs at time $t-1$ and earlier. This is introduced to prevent the leakage of information from the future into the past. The nature of human action involves a temporal sequence of events where the past influences the present and the future. Causal convolutions enforce a temporal causality constraint, ensuring that each output in the sequence depends on only the present and past and not the future. This is important in HAE where the chronological order of actions is essential for accurate analysis; (2) **Dilated convolutions**: In traditional CNNs, a kernel with fixed size and stride slides over the input tensor, operating the kernel function on only the adjacent elements. On the other hand, TCNs leverage the dilated convolutions that introduce gaps (strides) between the kernel elements. The dilation factor defines how many elements of the input signal are skipped between two filters. By increasing the dilation factor, TCNs can capture information from a larger receptive field, incorporating a broader temporal context. This enables the network to model long-range dependencies and capture patterns over larger time intervals. In addition, this characteristic can contemplate non-linearity by allowing the network to capture and process information at different temporal scales. When a TCN applies dilated convolutions with increasing dilation factors, the receptive field of each layer expands exponentially. This expansion enables the network to capture a broader range of temporal information, including short-term and long-term dependencies. The causal and dilation characteristics of the TCN

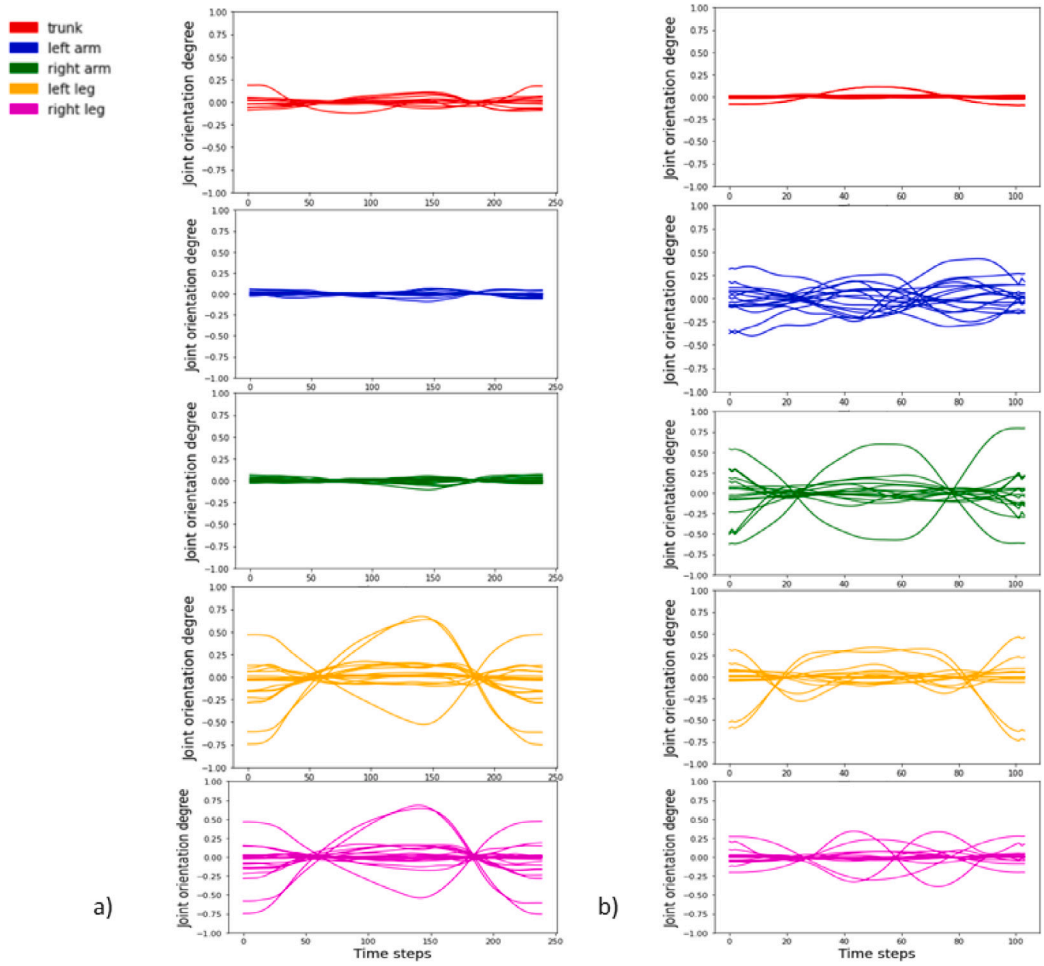

Fig. 1. Training samples of human body movement during deep squat action (a) (E1, UI-PRMD [28] dataset) showing joint orientations of the trunk, left arm, right arm, left leg, and right leg through 240 frames; (b) (EX5, KIMORE dataset [31]) showing joint orientations of the trunk, left arm, right arm, left leg, and right leg through 104 frames.

pipeline are depicted in Fig. 5; (3) **Residual connections**: The TCN architecture further incorporates ResNet using residual connections in its pipeline. It consists of several stacked residual blocks, as illustrated in Fig. 6. The residual connections (jumping or skip connections) bypass one or more temporal convolutional layers within the residual block. The output of the temporal convolutional layers is added element-wise to the original input, which is then fed as input to subsequent layers or residual blocks. The residual connection allows the model to learn the residual or the difference between the input and the transformed output, similar to the concept in ResNet. This helps to stabilize the training of the model and prevent it from vanishing gradients [45]. By introducing residual connections, TCNs with residual blocks can effectively propagate gradients during training, even through deep architectures. This can help overcome the vanishing gradient problem, improve optimization, and enable the training of deeper TCN models.

In this study, we utilized 5 TCN sub-networks for gathering characteristics of human movements from 5 different parts of the body (right arm, left arm, trunk, left leg, and right Leg) by analyzing joint displacements of these body parts. It should be noted that the signals of each body part movement are subsampled over time to smaller lengths of signals based on the baseline study [33] to have multi-scale data representation [46–48]. Therefore, the original input signal and the signals subsampled with the factor of 1/2, 1/4, and 1/8 of temporal length are given to TCN pipelines in TCN sub-networks. These TCN layers have dilation with weights of 1, 2, 4, 8, 16, 32, kernel size of 3, and dropout rate of 0.007. After concatenating the information of different body parts, 3 more TCN layers have been applied to learn the temporal correlations from the learned representation. The first TCN has dilation with weights of 1, 2, 4, and 8; the second one has dilation with weights of 1, 2, and 4; and the third TCN consists of dilation of 1 and 2. Finally, a linear regression layer predicts the movement quality score. This pipeline is designed in a way that it can comprehend the shared temporal patterns for each body part individually, and combine this information in the global representation as suggested by Shahroudy et al. [49]. We would like to highlight the advantages and motivations behind the sub-networks in the following: (1) The sub-networks allow the TCNs to capture information at different temporal scales effectively. This can enhance the discriminative power of the model by processing the data in different scales and capturing intricate details that might be missed in single-scale representation. (2) The different sub-networks are designed to adapt to diverse patterns within the input structures. This adaptability is crucial when dealing with temporal data that exhibits variations across multiple scales. (3) Each sub-network specializes in extracting features from different body parts, contributing to a more comprehensive representation of the whole body after combining the feature vectors. This usage of the sub-networks to process each body part individually with multiple scales is motivated based on the experimental results of prior works [33,50].

4. Results and discussion

This section encompasses the results and discussion. In Section 4.1 we describe the tools and configurations utilized for developing the LightPRA model and experimental results. Section 4.2 presents the

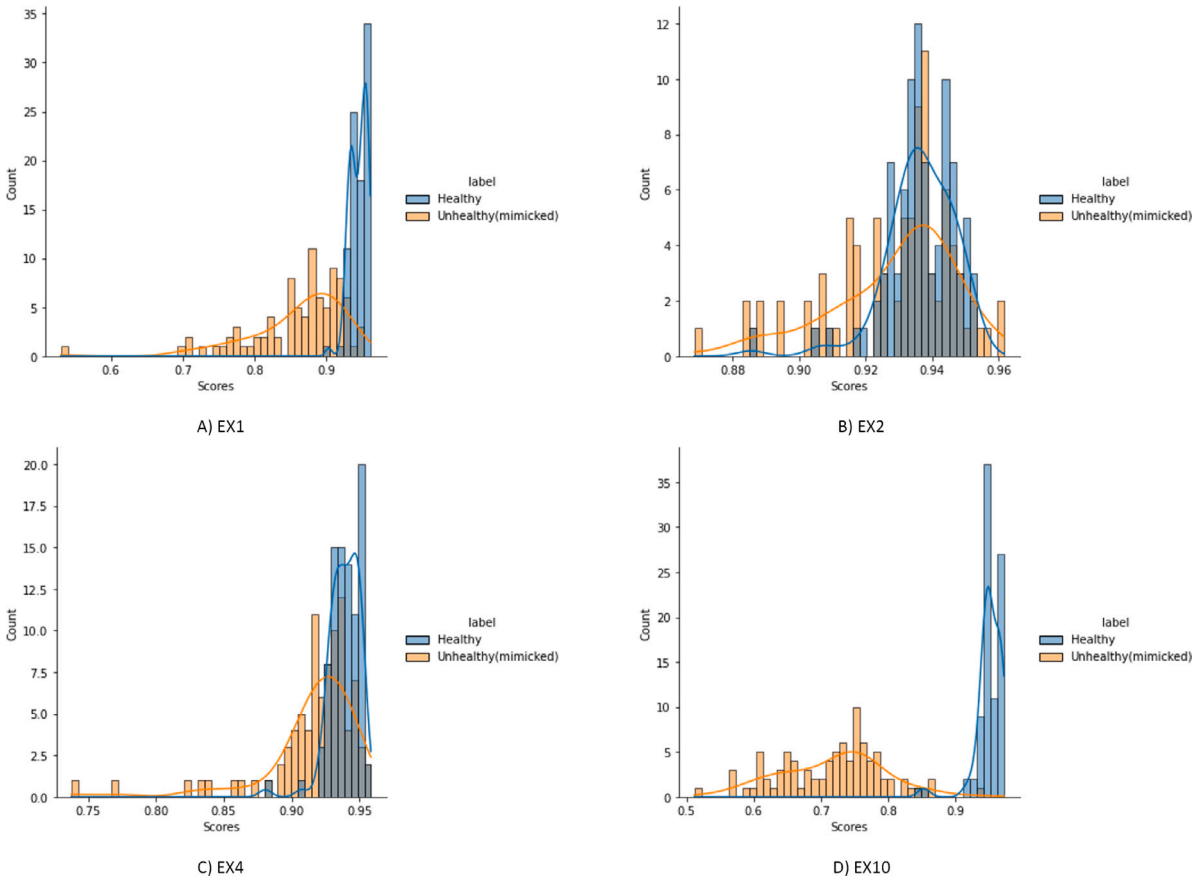

Fig. 2. Score distribution for some of the exercises in the UI-PRMD [28] dataset illustrating outliers in the samples. EX1, EX2, EX4, and EX10, illustrate deep squats, hurdle steps, side lunges, and standing shoulder scaption, respectively.

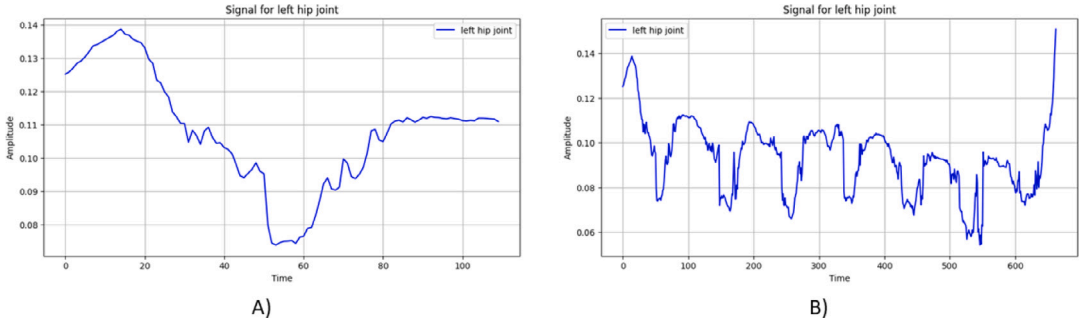

Fig. 3. The raw signal of the left hip joint in EX5 (Squat) of the KIMORE dataset [31]: (A) Illustrates the normalized and segmented signal of the first squat of one participant. (B) Depicts the whole action performed by a participant, which illustrates deviations in each repetition leading to the possible inappropriate scoring method.

overall results for the scoring performance of various methods on the two UI-PRMD and KIMORE datasets. In addition, the ablation analysis is provided in this section. Finally, in Section 4.3, the computational time and resources are presented, and a comprehensive discussion is provided.

4.1. General setup and configuration

The pipeline of the proposed model, including preprocessing, model training, and testing, is performed using *Python 3.8*, *Tensorflow 2.X*. We reported the average evaluation metric for 10 runs for a fair comparison between the LightPRA and the existing methods, similar to the results reported in the study conducted by Deb et al. [14]. For an objective comparison, we followed the same strategy as Deb et al. [14] on splitting the datasets into training, testing, and validation sets. It is worth mentioning that the processing times and computational costs are determined by computer configurations of Core i5 CPU and 8 GB RAM. In addition, different hyperparameter tunings, such as the activation function, dilation rate, dropout rate, batch size, and epoch, are selected by optimizing the error rate. This sub-optimal configuration of hyperparameter settings is performed through a trial-and-error procedure recommended in many other studies [15,51].

4.2. Scoring performance comparison

To assess the scoring performance of the proposed methodology, a comprehensive comparison is conducted using KIMORE [31] and UI-PRMD [28] datasets. The datasets include two distinct sensing technologies namely Kinect and Vicon. This compares two different sensing technologies, namely Vicon and Kinnect. In the following subsection,

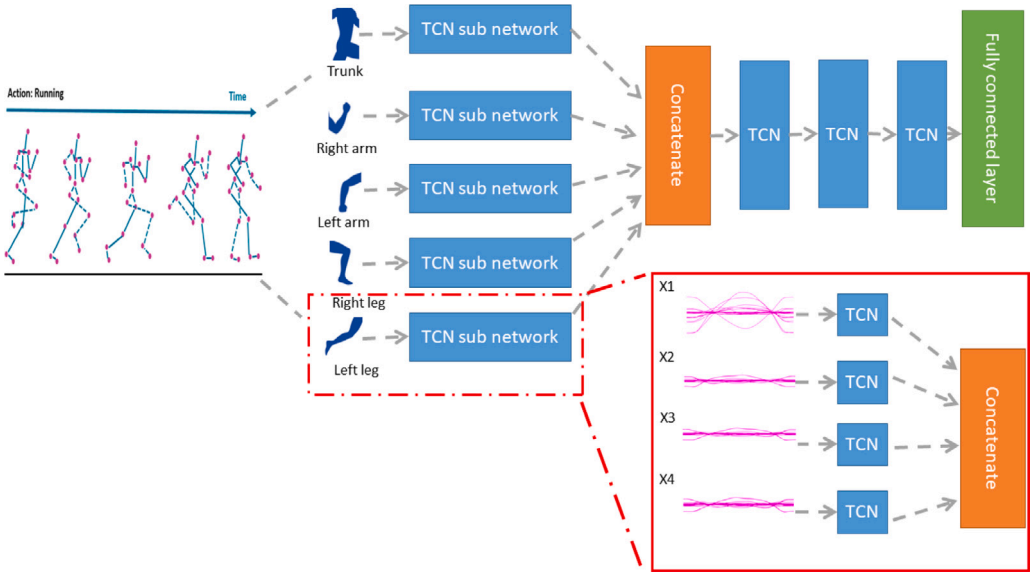

Fig. 4. Proposed LightPRA pipeline for physical rehabilitation exercise evaluation using an end-to-end TCN model. Every body movement comprises a multi-channel signal, where each joint orientation over time represents one channel in the signal. The signal is segmented into five different body parts and then fed into TCN sub-networks. Then in the TCN sub-networks, each original signal (X_1) and subsampled signals with the rate of 1/2 (X_2), 1/4 (X_3), and 1/8 (X_4) are given to TCN models as inputs. Next, the concatenated information of different body parts is processed using 3 TCN modules. Finally, a linear regression layer predicts the movement quality score.

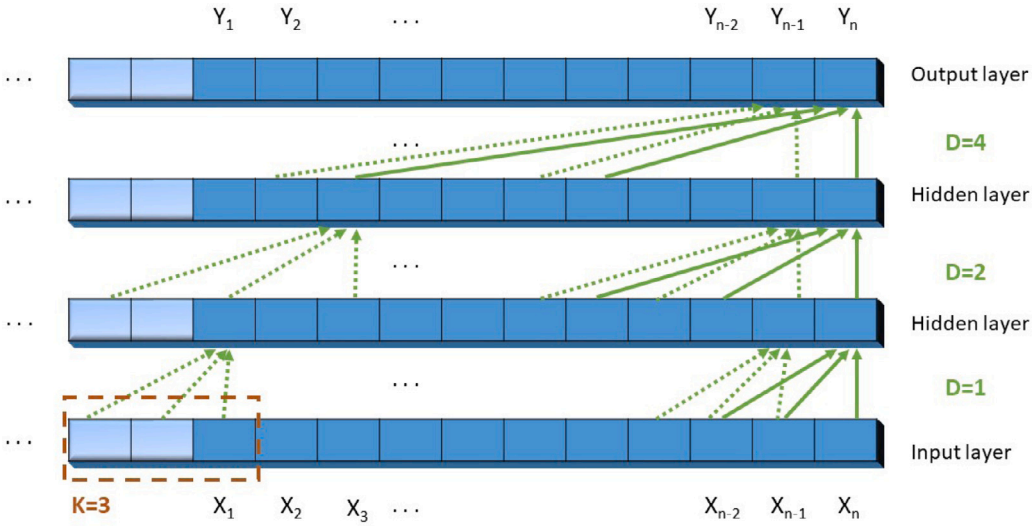

Fig. 5. Illustration of how causal dilated convolutions with dilation factors (shown as D in the figure) of 1, 2, 4, and a kernel (shown as K in the figure) of 3 are functioning. Temporal input (X_t) captures joint orientation at time t, with applied padding for processing early time steps. The output vector (Y) is fed to the next block for further feature extraction or processing. Dilated convolutions introduce the capacity to capture larger receptive fields and skip strides in the kernel. The causal feature prevents information leakage from the future into the past, making the convolutions different from a normal CNN.

we first discuss the specific sub-optimal hyperparameter tuning performed with trial-and-error and comparison with other state-of-the-art models on the UI-PRMD dataset. Then, the same procedure and discussion are performed for the KIMORE dataset to evaluate the model's generalization. As widely adopted for performance comparison in the previous studies, Mean Absolute Deviation (MAD), serves as one of our chosen metrics. The MAD is an average of the absolute deviation between true values and predicted values. In the context of HAE, a lower MAD indicates superior system performance. This metric evaluates the disparity between true and predicted scores in the dataset. MAD is preferred for its robustness to outliers, as observed in the UI-PRMD dataset, making it a more intuitive and interpretable measure compared to alternative metrics [52].

The LightPRA model for the UI-PRMD [28] dataset is trained with an ADAM optimizer and a batch size of 7, for a maximum of 400 epochs with a learning rate of $1e-3$. However, to mitigate overfitting and reduce the number of epochs, we implemented the early stopping callback mechanism. This approach allows training to cease when the loss function for the validation set ceases to improve while retaining the trained parameters associated with the lowest validation error. In our study, we employed early stopping with a minimum change threshold of different values for each exercise in the validation loss, accompanied by a patience value of 75, ensuring optimal training convergence. Fig. 7 illustrates the validation and training loss through the epochs of training the model for EX8 in the UI-PRMD dataset, which depicts no evidence of overfitting.

Previously, Liao et al. [33] proposed a Temporal Pyramid (for dividing and scaling the body movement signals) which includes a CNN–LSTM architecture. Song et al. proposed a multi-stream GCN backbone in which discriminative features are explored that ignore

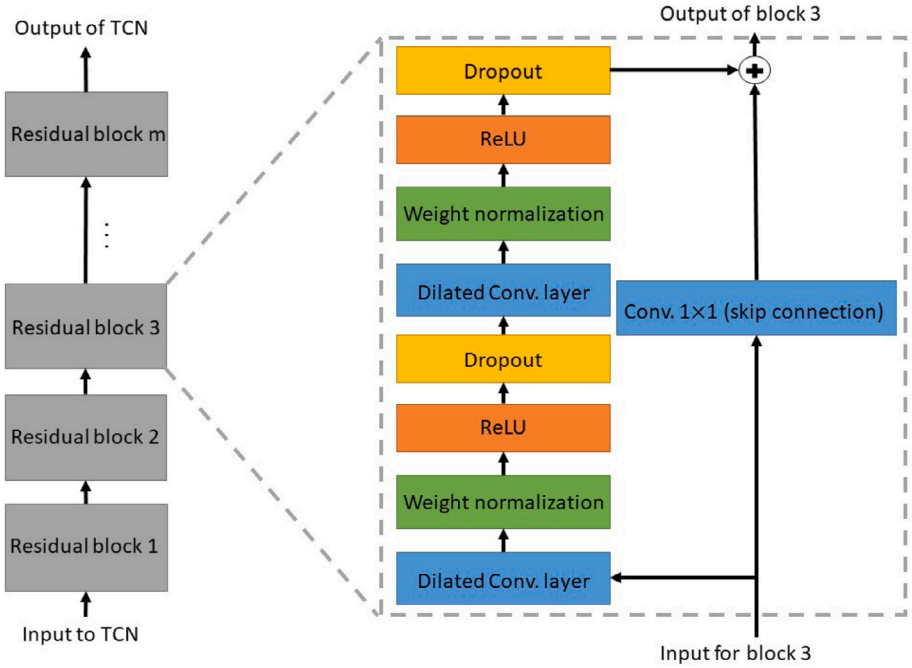

Fig. 6. Stacked residual blocks of TCN in which the output of one residual block is transferred to the input of the next block. The residual block consists of a pair of dilated causal convolutions followed by REctified Linear Units (RELU) as non-linearities. Convolutional filters within the block are subjected to weight normalization, and to prevent overfitting, dropout is employed after each dilated convolution.

redundant information using the activation degree of skeletons. Li et al. [53] introduced an end-to-end convolutional co-occurrence feature learning framework in which the semantic information of both temporal and spatial features are explored. Zhang et al. [54] introduced a semantic-guided network in which joint type and frame index as high-level semantics are fed into the framework with two frame-level and joint-level modules to find frame and joint correlations. However, according to Deb et al. [14], this model fails to extract rich sequential dependencies between different frames. Deb et al. [14] introduced a graph-based model in which an attentional STGCN model is followed by an LSTM for capturing skeleton and sequential data. Table 1 depicts the average results for MAD of scores predicted by different models for 10 exercises in the UI-PRMD [28] dataset. For a better comparison of results, we also developed a dilated 1D CNN-TCN model in which 1D dilated CNN layers are followed by three TCN layers for temporal feature capturing. The comparisons illustrate that LightPRA outperforms all of the previous models in scoring most of the exercises, and for a few of them, the results were close to the best result. The comparison of the TCN and LSTM-based models (such as models introduced by Liao et al. [33] and Deb et al. [14]) proves the superiority of using a TCN architecture in capturing long-term information in a long sequence of data. In the last row of Table 1, we provided the average scoring performance through all exercises for all of the methods. This illustrates that our proposed model on average is outperforming other methods, and performs as effectively as the strong graph-based method like STGCN [14].

To evaluate the generalization of the method on another dataset and another type of sensor, the noisy skeleton information in KIMORE [31] is fed into different models for 5 exercises. Our proposed model is trained with a batch size of 7, and a maximum of 400 epochs with an early stopping of 70 patience degrees. The learning rate is set to $1e-3$. The MAD metric results are depicted in Table 2, for the previously mentioned methods. In the table the columns depicting the performance for orientation (quaternions) data and position data are demonstrated with O and P in parentheses, respectively. The reason behind the big difference in MAD values for KIMORE compared to UI-PRMD [28] can be based on the noisy data and action evaluation metric as discussed before. The LightPRA proposed method illustrates a high difference in performance compared to the CNN–LSTM model [33], demonstrating TCN architecture's comparative advantage. The model performs well for a fairly complex activity like deep squat (EX5), which includes several orientations of several joints and limbs. In addition to MAD, the widely used metric of Root Mean Squared Error (RMSE) for scoring performance is considered. This metric calculates the square root of the average of the squared differences between each predicted score and its corresponding true score. The results are illustrated in Table 3, which shows that for this metric the model also provides good performance. In addition, comparison of the results between the joint orientation and positional data in Tables 2 and 3 shows that the orientation provides more discriminative information for the model. On average, the model performs very well compared to the previous studies, suggesting this model can be used in HAE applications.

In our ablation and sensitivity analysis, we investigated several factors regarding LightPRA's scoring performance. Firstly, we evaluated the scoring sensitivity of LightPRA for the Deep Squat action (EX5) in the KIMORE dataset, considering both MAD and RMSE, for different sets of inputs. We considered the deletion of some body parts as input and fed the rest of the body data to evaluate the changes in the model's performance. Our findings (illustrated in Table 4) showed that excluding data from both legs as input for the deep squat action results in worse performance. This is understandable since the leg movement in this action is significant, and eliminating such data may lead to the loss of vital information for the model.

Additionally, we explored various configurations of the trained LightPRA model on exercise 5 of the KIMORE dataset. Given the considerable number of hyperparameters in our model, we deliberated on including or replacing a few important architecture and parameter aspects that have shown an important role in the initial try and error (as illustrated in Table 5). For instance, we considered the exclusion of two final TCN layers from the architecture, along with testing two different activation functions in the last regression layer. Our analyses illustrate that the model achieved better performance when utilizing three TCNs at the final stage with a linear activation function.

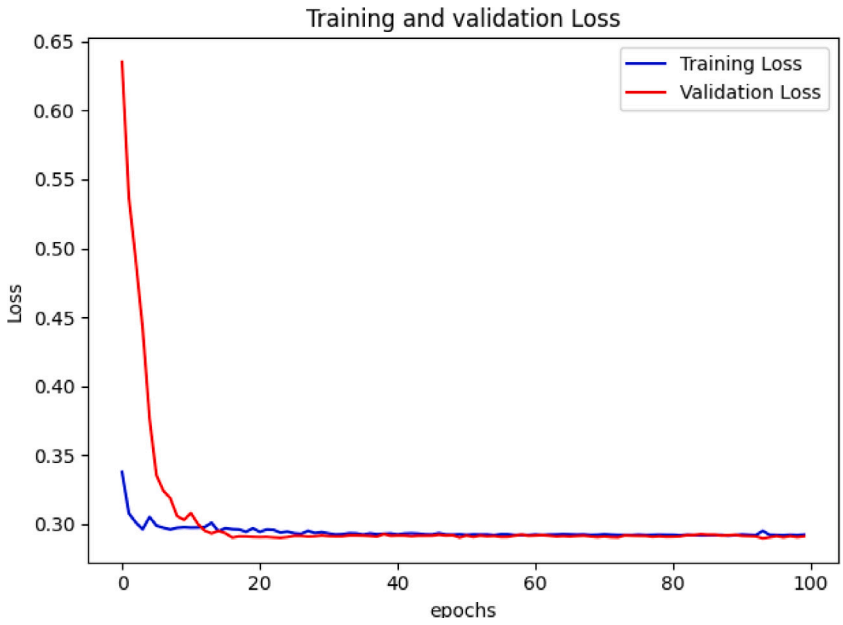

Fig. 7. Training and validation loss over epochs for UI-PRMD [28] dataset for EX8 using the proposed model, indicating absence of overfitting.

Table 1
Performance comparison of different methods for HAE on UI-PRMD [28] dataset based on MAD(↓).

Exercise	Methods						
	LightPRA	Dilated CNN+TCN	Liao et al. [33]	Deb et al. [14]	Song et al. [55]	Zhang et al. [54]	Li et al. [53]
EX1	0.014	0.015	0.011	**0.009**	0.011	0.022	0.011
EX2	0.007	0.008	0.028	**0.006**	**0.006**	0.008	0.029
EX3	0.011	0.013	0.039	0.013	**0.010**	0.016	0.056
EX4	**0.006**	**0.006**	0.012	**0.006**	0.014	0.016	0.014
EX5	**0.008**	**0.008**	0.019	**0.008**	0.013	**0.008**	0.017
EX6	**0.006**	0.007	0.018	**0.006**	0.009	0.008	0.019
EX7	**0.010**	0.011	0.038	0.011	0.017	0.021	0.027
EX8	**0.011**	**0.011**	0.023	0.016	0.017	0.025	0.025
EX9	**0.008**	0.009	0.023	**0.008**	**0.008**	0.027	0.027
EX10	0.038	0.041	0.042	**0.031**	0.038	0.066	0.047
Avg	0.0119	0.0129	0.0253	**0.0114**	0.0143	0.0217	0.0272

Table 2
Performance comparison of different methods for HAE on KIMORE dataset [31] based on MAD(↓). The (O), and (P) designations for each method mean that the model is trained on orientation and positional data, respectively.

Exercise	Methods					
	LightPRA(O)[a]	LightPRA(O)[b]	LightPRA (P)[a]	Liao et al. (O)[a] [33]	Liao et al. (P) [33]	Deb et al. (P) [14]
EX1	**0.20**	0.40	0.25	0.24	1.14	0.80
EX2	**0.27**	0.57	0.28	0.31	1.53	0.77
EX3	**0.21**	0.39	0.25	0.23	0.85	0.37
EX4	**0.28**	0.48	0.30	**0.28**	0.47	0.35
EX5	**0.25**	0.47	0.28	0.29	0.85	0.62
Avg	**0.24**	0.46	0.27	0.27	0.97	0.58

[a] Scoring label is normalized in the range of 0 to 1.
[b] Scoring label is normalized in the range of −1 to 1.

4.3. Computational time and resource management

In this subsection, we provide the comparative analysis of the computational time and computational resource utilized in the proposed LightPRA model, LSTM-based architecture [33], and especially the STGC-LSTM model [14] which demonstrated high scoring performance in the previous subsection. We considered several factors for analyzing the computational efficiency of these methods, aiming to determine whether it is more important to build a model with high-scoring performance that requires considerable computational time or to construct and apply models that provide a balance between scoring performance and computational efficiency for telerehabilitation applications. It should be noted that in this study, we selected previous studies with available source code for fair and objective comparisons under equal computational conditions while maintaining computational resources on CPUs. Some Transformer-based approaches explored in previous studies, lack source code and rely on GPU resources, potentially impacting computational fairness. First, we analyzed training time (in seconds) for the UI-PRMD [28] dataset, illustrated in Table 6. The number of trainable parameters for the LightPRA model, Dilated CNN+TCN, and CNN+LSTM are 122 557, 2 105 341, and 5 688 081, respectively. The values for time in Table 6 illustrate an average of time for finding the sub-optimal solution based on the validation set and early stopping. These values depict the low computational time for training the model

Table 3
Performance comparison of different methods for HAE on KIMORE dataset [31] based on RMSE(↓). The (O), and (P) designations for each method mean that the model is trained on orientation and positional data, respectively.

Exercise	Methods				
	LightPRA(O)[a]	LightPRA(O)[b]	Liao et al. (O) [33]	Liao et al. (P) [33]	Deb et al. (P) [14]
EX1	**0.25**	0.46	0.26	2.53	2.02
EX2	**0.32**	0.57	0.33	3.74	2.12
EX3	**0.19**	0.43	0.21	1.56	0.55
EX4	0.30	0.51	**0.27**	0.79	0.64
EX5	**0.27**	0.52	**0.27**	1.91	1.18
Avg	**0.27**	0.50	**0.27**	2.11	1.3

[a] Scoring label is normalized in the range of 0 to 1.
[b] Scoring label is normalized in the range of −1 to 1.

Table 4
Sensitivity analysis of LightPRA on exercise 5 of KIMORE dataset [31] based on MAD, and RMSE(↓) for exclusions of body part as input.

Deleted body part	MAD	RMSE
Trunk	0.28	0.29
Left arm	0.3	0.27
Right arm	0.26	0.27
Left leg	0.27	0.28
Right leg	0.26	0.3
Both legs	0.38	0.47
Both arms	0.29	0.26

Table 5
Ablation study of LightPRA on exercise 5 of KIMORE dataset [31] based on MAD, and RMSE(↓) based on exclusion of final TCNs and change of activation function.

Has final 3 TCNs	Activation function	MAD	RMSE
No	Sigmoid	0.29	0.34
Yes	Sigmoid	0.26	0.31
No	Linear	0.27	0.30
Yes	Linear	**0.25**	**0.27**

in pure TCN-based architecture. This can be due to the characteristics of TCN and LSTM and the lower number of trainable parameters for the TCN backbone.

For a better analysis of computational power, we performed the same analysis on the KIMORE dataset [31], adding the model proposed by Deb et al. [14] for comparison. The number of trainable parameters for the proposed model, CNN+LSTM [33], and STGCN [14] are 121 261, 5 613 841, and 712 209, respectively. The computational time for training the models is depicted in Table 7. It should be noted that these values are provided for the number of epochs found by early stopping for the LightPRA and CNN+LSTM [53] models and 30 epochs for the STGCN model [14]. This is because it is clearly mentioned in the study conducted by Deb et al. [14] that the model needs to be trained for 1500 epochs, and based on other studies, this can take up to 72 h of training [56] for this architecture on GPUs. Comparing the results in Table 7 proves that not only the LightPRA model performs better than LSTM-based models, but it also performs better than graph-based attentional models. Even though the number of parameters for the STGCN [14] is fairly normal, the attentional architecture in the pipeline in the model adds to the time complexity and makes it not suitable for real-time applications.

For a fair comparison, we included step-wise training time in Fig. 8. In the context of training a model,*"step"* refers to one iteration for one gradient update. This includes processing a batch of data, computing the gradients of the model's parameters concerning the loss condition, and updating the model's weights using the optimizer. Evaluating the results illustrates that the LightPRA model significantly reduces the step-wise training time across all of the exercises in the KIMORE dataset [31]. This notable decrease in the step-wise training time for the LightPRA model can be attributed to the fewer trainable parameters in the model while being able to swiftly learn distinctive features in a parallel way. The training time plays an important role in real-world applications where the model needs to be trained on a new patient's activity and requires fine-tuning for new exercises. In Fig. 8, the LightPRA model illustrates remarkably fewer inference times on the testing set for which it illustrated similar scoring performance to STGCN [14] in the preceding subsection. Inference time holds significant relevance to real-world healthcare applications where the LightPRA model excels by prioritizing real-time feedback without compromising scoring accuracy. This characteristic is crucial in applications such as HAE requiring immediate feedback without sacrificing the scoring accuracy.

In Fig. 9, the plot illustrates the RAM usage in MegaBytes (MB) for training the models across all of the exercises in the KIMORE dataset [31]. The impact of a diminished number of training parameters and epochs for training is clearly reflected in the computational resource usage. These findings emphasize that in more extensive data and complex problems, these values can drastically escalate to the scales of GigaBytes (GB) or TeraByte (TB), highlighting the promising potential of TCN models for real-world problems. In general, the results underscore the LightPRA model's ability to maintain a balance between achieving high scoring performance and relatively low computational time and complexity.

In summary, the comparative analysis of different architectures sheds light on different factors affecting their scoring and computational performance. (1) **Parallel computation**: Using Causal convolutions in the TCN allows the model to consider a long-term sequence as a whole in both the training and inference stages. This assists the model to compute multiple temporal dependencies simultaneously. In contrast, the LSTM-based models relying on several input, forget, and output gates, process data in a sequential way, with computations in each time step dependent on prior outputs. (2) **Shorter Memory path**: The dilated convolutions in the TCN assist the model to have more receptive fields without significantly increasing the number of parameters, making them more efficient in capturing long-term dependencies. However, the complex structure of the LSTM (having many gates and recurrent connections adds to the number of trainable parameters in these models making them computationally complex. The straightforward structure of TCN compared to the attentional graph-based model which often includes intricate computation and recurrent iterations results in faster computation in TCN. (3) **Data representation**: the graph-based attentional structure includes graph representation of the data for which the computationally expensive attention calculations over the nodes and edges escalate the time complexity and scoring performance of the model. However, the straightforward data representation in TCN and the performing convolutions across the sequences add to the flexibility and simplicity of TCN-based models, making the proper learning of discriminative features efficient.

5. Conclusion

This paper introduces the novel LightPRA model for automatic physical rehabilitation exercise assessment which focuses on learning temporal features from joint orientation data with rich spatial information. To evaluate the computational and scoring performance of the

Table 6
Model training time comparison in seconds(↓) for different HAE methods on UI-PRMD [28] dataset.

Exercise	Methods		
	LightPRA	Dilated CNN+TCN	CNN+LSTM [33]
EX1	**480**	503	1560
EX2	**462**	920	1257
EX3	**784**	820	2057
EX4	**480**	637	2281
EX5	**452**	558	2367
EX6	**432**	569	3084
EX7	**484**	743	2001
EX8	**487**	523	1338
EX9	**448**	687	3624
EX10	**485**	768	5469

Table 7
Model training time comparison in seconds(↓) for different HAE methods on KIMORE [31] dataset.

Exercise	Methods		
	LightPRA	CNN+LSTM [33]	STGCN+LSTM+ATT [14]
EX1	**722**	866	6295
EX2	**1500**	1535	6223
EX3	**1200**	1832	6452
EX4	**432**	492	5942
EX5	**420**	546	5401

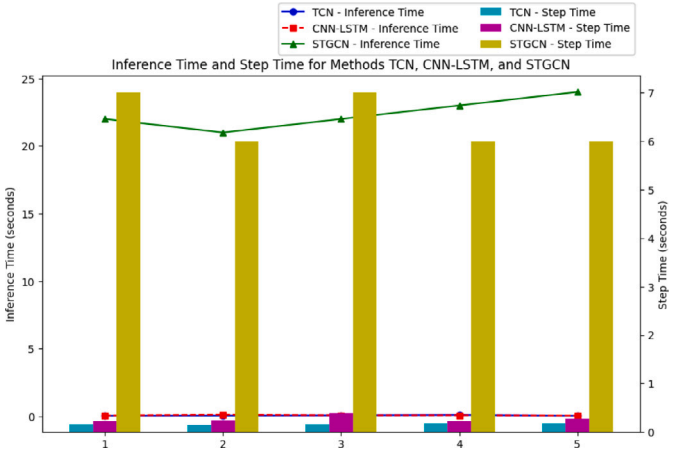

Fig. 8. The inference time (left side vertical axis, illustrated with lines) and training time per step (right side vertical axis, illustrated with bars) in seconds(↓) for different HAE-models of proposed LightPRA (TCN) model, STGCN [14] and CNN–LSTM [33] for KIMORE [31] dataset.

proposed method, two public datasets are utilized, namely UI-PRMD and KIMORE. The comparisons of the proposed method and the previously proposed LSTM-based algorithms and graph-based techniques suggest that on average the proposed LightPRA methodology outperforms them in scoring performance in the KIMORE dataset (especially for a complex action like deep squat). Additionally, on average it performs as effectively as the strong attentional graph-based model in the UI-PRMD dataset. In addition, this method depicts significantly reduced training and inference computational time and memory resources compared to the state-of-the-art approaches, marking it as a promising approach suitable for devices commonly found in households, as well as resource-limited devices, Edge computing setups, and IoT devices. In the future, our focus will involve exploring the implementation of the proposed method on resource-constrained devices. In addition, further exploration of statistical action scoring (labeling) is needed to address the challenges associated with the scoring method of the UI-PRMD and KIMORE datasets.

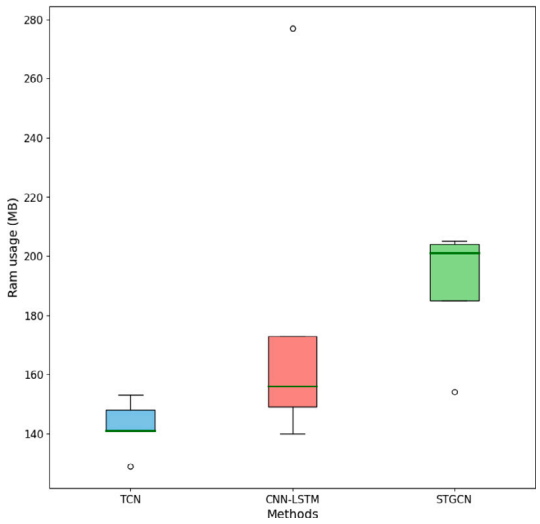

Fig. 9. The RAM usage in MegaBytes(↓) for building and training of different HAE methods including proposed LightPRA (TCN) model, STGCN [14] and CNN–LSTM [33] for KIMORE [31] dataset.

It is worthwhile noting that while our LightPRA architecture excels in skeleton-based human action evaluation, its use in broader motion analysis tasks like action recognition or prediction may lead to information loss. The architecture, relying solely on skeleton data and convolutional operations, may not effectively capture crucial contextual details such as location, occasion, motivation, and facial expressions, vital for tasks like action prediction in surveillance scenarios. In addition, TCNs primarily operate on raw temporal sequences and may not explicitly incorporate information such as interactions, purpose, and emotion of the actions. The inclusion of this information especially in complex series of actions can enhance the interpretability and performance of the HAE system. In this study, determining metrics like floating-point operations (FLOPs) poses a challenge due to the complicated nature of the multi-input complex architecture employed. To enhance the comparability and evaluation of the models considered in this study, we recommend further research to explore the development of specialized functions tailored to address the complexities of such architectures. This will help us compare models more accurately and meaningfully in future studies.

CRediT authorship contribution statement

Sara Sardari: Formal analysis, Methodology, Writing – original draft. **Sara Sharifzadeh:** Supervision, Writing – review & editing. **Alireza Daneshkhah:** Supervision, Writing – review & editing. **Seng W. Loke:** Writing – review & editing, Supervision. **Vasile Palade:** Supervision, Writing – review & editing. **Michael J. Duncan:** Supervision, Writing – review & editing. **Bahareh Nakisa:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Acknowledgments

The authors would like to thank Coventry University and Deakin University for jointly funding this Ph.D. project titled "Activity Recognition Using Digital Frame Streams for Monitoring Rehab Period".

References

[1] I.D. Cameron, S.E. Kurrle, 1: Rehabilitation and older people, Med. J. Australia 177 (7) (2002) 387–391.
[2] R. Trumpf, L.E. Schulte, H. Schroeder, R.T. Larsen, P. Haussermann, W. Zijlstra, T. Fleiner, Physical activity monitoring-based interventions in geriatric patients: a scoping review on intervention components and clinical applicability, Eur. Rev. Aging Phys. Activity 20 (1) (2023) 1–17.
[3] S. Sharififar, H. Ghasemi, C. Geis, H. Azari, L. Adkins, B. Speight, H.K. Vincent, Telerehabilitation service impact on physical function and adherence compared to face-to-face rehabilitation in patients with stroke: A systematic review and meta-analysis, PM&R (2023).
[4] L. Desveaux, R. Goldstein, S. Mathur, D. Brooks, Barriers to physical activity following rehabilitation: Perspectives of older adults with chronic disease, J. Aging Phys. Activity 24 (2) (2016) 223–233.
[5] B. Debnath, M. O'brien, M. Yamaguchi, A. Behera, A review of computer vision-based approaches for physical rehabilitation and assessment, Multimedia Syst. 28 (1) (2022) 209–239.
[6] S. Sardari, S. Sharifzadeh, A. Daneshkhah, B. Nakisa, S.W. Loke, V. Palade, M.J. Duncan, Artificial intelligence for skeleton-based physical rehabilitation action evaluation: A systematic review, Comput. Biol. Med. (2023) 106835.
[7] A.Y. Gelaw, B. Janakiraman, B.F. Gebremeskel, H. Ravichandran, Effectiveness of home-based rehabilitation in improving physical function of persons with stroke and other physical disability: A systematic review of randomized controlled trials, J. Stroke Cerebrovasc. Dis. 29 (6) (2020) 104800.
[8] H. Manjunatha, S. Pareek, S.S. Jujjavarapu, M. Ghobadi, T. Kesavadas, E.T. Esfahani, Upper limb home-based robotic rehabilitation during COVID-19 outbreak, Front. Robotics AI 8 (2021) 612834.
[9] S.C. Cramer, L. Dodakian, V. Le, A. McKenzie, J. See, R. Augsburger, R.J. Zhou, S.M. Raefsky, T. Nguyen, B. Vanderschelden, et al., A feasibility study of expanded home-based telerehabilitation after stroke, Front. Neurol. 11 (2021) 611453.
[10] S. Stephenson, R. Wiles, Advantages and disadvantages of the home setting for therapy: views of patients and therapists, Br. J. Occup. Ther. 63 (2) (2000) 59–64.
[11] N. Eichler, H. Hel-Or, I. Shmishoni, D. Itah, B. Gross, S. Raz, Non-invasive motion analysis for stroke rehabilitation using off the shelf 3d sensors, in: 2018 International Joint Conference on Neural Networks, IJCNN, IEEE, 2018, pp. 1–8.
[12] S.H. Chowdhury, M. Al Amin, A.M. Rahman, M.A. Amin, A.A. Ali, Assessment of rehabilitation exercises from depth sensor data, in: 2021 24th International Conference on Computer and Information Technology, ICCIT, IEEE, 2021, pp. 1–7.
[13] M.H. Lee, D.P. Siewiorek, A. Smailagic, A. Bernardino, S.B.i. Badia, Learning to assess the quality of stroke rehabilitation exercises, in: Proceedings of the 24th International Conference on Intelligent User Interfaces, 2019, pp. 218–228.
[14] S. Deb, M.F. Islam, S. Rahman, S. Rahman, Graph convolutional networks for assessment of physical rehabilitation exercises, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 410–419.
[15] S. Sardari, B. Nakisa, M.N. Rastgoo, P. Eklund, Audio based depression detection using convolutional autoencoder, Expert Syst. Appl. 189 (2022) 116076.
[16] I. Ahmad, A. Merla, F. Ali, B. Shah, A.A. AlZubi, M.A. AlZubi, A deep transfer learning approach for COVID-19 detection and exploring a sense of belonging with diabetes, Front. Public Health 11 (2023).
[17] S.P. Praveen, P.N. Srinivasu, J. Shafi, M. Wozniak, M.F. Ijaz, ResNet-32 and FastAI for diagnoses of ductal carcinoma from 2D tissue slides, Sci. Rep. 12 (1) (2022) 20804.
[18] G.E. Rao, B. Rajitha, P.N. Srinivasu, M.F. Ijaz, M. Woźniak, Hybrid framework for respiratory lung diseases detection based on classical CNN and quantum classifiers from chest X-rays, Biomed. Signal Process. Control 88 (2024) 105567.
[19] G. Alfian, M. Syafrudin, M.F. Ijaz, M.A. Syaekhoni, N.L. Fitriyani, J. Rhee, A personalized healthcare monitoring system for diabetic patients by utilizing BLE-based sensors and real-time data processing, Sensors 18 (7) (2018) 2183.
[20] Z. Wang, Y. Ma, Z. Liu, J. Tang, R-transformer: Recurrent neural network enhanced transformer, 2019, arXiv preprint arXiv:1907.05572.
[21] R. Mehta, S. Sharifzadeh, V. Palade, B. Tan, A. Daneshkhah, Y. Karayaneva, Deep learning techniques for radar-based continuous human activity recognition, Mach. Learn. Knowl. Extraction 5 (4) (2023) 1493–1518.
[22] M. Kumar, S. Verma, A. Kumar, M.F. Ijaz, D.B. Rawat, et al., ANAF-IoMT: A novel architectural framework for IoMT-enabled smart healthcare system by enhancing security based on RECC-VC, IEEE Trans. Ind. Inform. 18 (12) (2022) 8936–8943.
[23] I. Zakariyya, M.O. Al-Kadri, H. Kalutarage, A. Petrovski, Reducing Computational Cost in IoT Cyber Security: Case Study of Artificial Immune System Algorithm, SciTePress, 2019.
[24] C. Meng, X.S. Jiang, X.M. Wei, T. Wei, A time convolutional network based outlier detection for multidimensional time series in cyber-physical-social systems, IEEE Access 8 (2020) 74933–74942.
[25] S. Bai, J.Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, 2018, arXiv preprint arXiv:1803.01271.
[26] M. Nan, M. Trăscău, A.M. Florea, C.C. Iacob, Comparison between recurrent networks and temporal convolutional networks approaches for skeleton-based action recognition, Sensors 21 (6) (2021) 2051.
[27] M. Carreras, G. Deriu, L. Raffo, L. Benini, P. Meloni, Optimizing temporal convolutional network inference on FPGA-based accelerators, IEEE J. Emerg. Sel. Top. Circuits Syst. 10 (3) (2020) 348–361.
[28] A. Vakanski, H.p. Jun, D. Paul, R. Baker, A data set of human body movements for physical rehabilitation exercises, Data 3 (1) (2018) 2.
[29] A. Paiement, L. Tao, S. Hannuna, M. Camplani, D. Damen, M. Mirmehdi, Online quality assessment of human movement from skeleton data, in: British Machine Vision Conference, BMVA Press, 2014, pp. 153–166.
[30] L. Tao, A. Paiement, D. Damen, M. Mirmehdi, S. Hannuna, M. Camplani, T. Burghardt, I. Craddock, A comparative study of pose representation and dynamics modelling for online motion quality assessment, Comput. Vis. Image Underst. 148 (2016) 136–152.
[31] M. Capecci, M.G. Ceravolo, F. Ferracuti, S. Iarlori, A. Monteriu, L. Romeo, F. Verdini, The KIMORE dataset: Kinematic assessment of movement and clinical scores for remote monitoring of physical rehabilitation, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (7) (2019) 1436–1448.
[32] J.H. Challis, Quaternions as a solution to determining the angular kinematics of human movement, BMC Biomed. Eng. 2 (1) (2020) 5.
[33] Y. Liao, A. Vakanski, M. Xian, A deep learning framework for assessing physical rehabilitation exercises, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (2) (2020) 468–477.
[34] T. Dao, D. Fu, S. Ermon, A. Rudra, C. Ré, Flashattention: Fast and memory-efficient exact attention with io-awareness, Adv. Neural Inf. Process. Syst. 35 (2022) 16344–16359.
[35] Y. Li, Z. Zuo, J. Pan, Sensor-based fall detection using a combination model of a temporal convolutional network and a gated recurrent unit, Future Gener. Comput. Syst. 139 (2023) 53–63.
[36] S.J. Li, Y. AbuFarha, Y. Liu, M.M. Cheng, J. Gall, Ms-tcn++: Multi-stage temporal convolutional network for action segmentation, IEEE Trans. Pattern Anal. Mach. Intell. (2020).
[37] Y.A. Farha, J. Gall, Ms-tcn: Multi-stage temporal convolutional network for action segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3575–3584.
[38] C. Lea, M.D. Flynn, R. Vidal, A. Reiter, G.D. Hager, Temporal convolutional networks for action segmentation and detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 156–165.
[39] B. Zhang, B. Goel, M.H. Sarhan, V.K. Goel, R. Abukhalil, B. Kalesan, N. Stottler, S. Petculescu, Surgical workflow recognition with temporal convolution and transformer for action segmentation, Int. J. Comput. Assist. Radiol. Surg. 18 (4) (2023) 785–794.
[40] A. Sabater, L. Santos, J. Santos-Victor, A. Bernardino, L. Montesano, A.C. Murillo, One-shot action recognition in challenging therapy scenarios, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2777–2785.
[41] C.Y. Chang, B. Lange, M. Zhang, S. Koenig, P. Requejo, N. Somboon, A.A. Sawchuk, A.A. Rizzo, Towards pervasive physical rehabilitation using microsoft kinect, in: 2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops, IEEE, 2012, pp. 159–162.
[42] H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, IEEE Trans. Acoust., Speech, Signal Process. 26 (1) (1978) 43–49.
[43] Y. Liu, X. Li, L. Yang, G. Bian, H. Yu, A CNN-transformer hybrid recognition approach for sEMG-based dynamic gesture prediction, IEEE Trans. Instrum. Meas. (2023).
[44] Z. Zhang, J. Liu, S. Pang, M. Shi, H.H. Goh, Y. Zhang, D. Zhang, General short-term load forecasting based on multi-task temporal convolutional network in COVID-19, Int. J. Electr. Power Energy Syst. 147 (2023) 108811.
[45] C.F.G.D. Santos, J.P. Papa, Avoiding overfitting: A survey on regularization methods for convolutional neural networks, ACM Comput. Surv. 54 (10s) (2022) 1–25.
[46] C. Zhao, A.K. Thabet, B. Ghanem, Video self-stitching graph network for temporal action localization, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 13658–13667.
[47] M.E. Swift, W. Ayers, S. Pallanck, S. Wehrwein, Visualizing the passage of time with video temporal pyramids, IEEE Trans. Vis. Comput. Graphics 29 (1) (2022) 171–181.
[48] C. Yang, Y. Xu, J. Shi, B. Dai, B. Zhou, Temporal pyramid network for action recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 591–600.

[49] A. Shahroudy, J. Liu, T.T. Ng, G. Wang, Ntu rgb+ d: A large scale dataset for 3d human activity analysis, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1010–1019.

[50] Y. Du, W. Wang, L. Wang, Hierarchical recurrent neural network for skeleton based action recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1110–1118.

[51] J.M. Ortíz Rodríguez, M.d.R. Martínez Blanco, J.M. Cervantes Miramontes, H.R. Vega Carrillo, et al., Robust Design of Artificial Neural Networks Methodology in Neutron Spectrometry, IntechOpen, 2013.

[52] T.O. Hodson, Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not, Geosci. Model Dev. 15 (14) (2022) 5481–5487.

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

[54] P. Zhang, C. Lan, W. Zeng, J. Xing, J. Xue, N. Zheng, Semantics-guided neural networks for efficient skeleton-based human action recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 1112–1121.

[55] Y.F. Song, Z. Zhang, C. Shan, L. Wang, Richly activated graph convolutional network for robust skeleton-based action recognition, IEEE Trans. Circuits Syst. Video Technol. 31 (5) (2020) 1915–1925.

[56] Y. Mourchid, R. Slama, D-STGCNT: A dense spatio-temporal graph conv-GRU network based on transformer for assessment of patient physical rehabilitation, Comput. Biol. Med. 165 (2023) 107420.