Journal of Universal Computer Science, vol. 31, no. 12 (2025), 1386-1413
submitted: 10/11/2024, accepted: 6/5/2025, appeared: 28/10/2025 CC BY 4.0

Enhancing Home-Based Rehabilitation Exercises with a Temporal Conditional Generative Adversarial Network

Fouzi Lezzar
(Abdelhamid MEHRI Constantine 2 University - LIRE laboratory, Constantine, Algeria
https://orcid.org/0000-0002-9520-4924, fouzi.lezzar@univ-constantine2.dz)

Seif Eddine Mili
(Ecole Normale Supérieure, Constantine, Algeria, LISCO laboratory, Annaba, Algeria
https://orcid.org/0000-0002-8424-3354, seifmili@gmail.com)

Abstract: Physical rehabilitation is essential for restoring motor function; however, traditional methods often require therapist supervision, which can be costly and inaccessible. Home-based rehabilitation offers a practical alternative, but without real-time guidance, patients may develop incorrect movement patterns that impede progress. Existing approaches typically provide feedback only after exercises are completed, reducing their effectiveness. To overcome this limitation, we propose a Temporal Conditional Generative Adversarial Network (TCGAN)-based motion generation system that delivers real-time skeletal guidance tailored to each patient's body structure and positioning. By detecting key anatomical landmarks and generating adaptive motion sequences, the system ensures precise movement execution, minimizing errors and improving rehabilitation outcomes. Patients can mimic these movements, enabling them to perform exercises with greater accuracy and confidence. Quantitative and qualitative evaluations confirm the effectiveness of the generated exercises, thanks to an optimized architecture, an improved loss function, a refined training process, and fine-tuned TCGAN hyperparameters. Experimental results demonstrate a high degree of similarity between generated and real movements, with a Fréchet Inception Distance (FID) score of 0.89 and strong temporal alignment, as shown by Dynamic Time Warping (DTW) scores ranging from 2.9 to 5.6 across nine rehabilitation exercises. These metrics underscore the system's realism and reliability.

Keywords: Home-based rehabilitation; real-time exercise guidance; TCGAN; skeletal movements; movement generation.
Categories: I.2, I.4
DOI: 10.3897/jucs.141304

1 Introduction

Physical rehabilitation encompasses a set of methods aimed at enabling individuals affected by disabling conditions to recover, either partially or fully. It is most commonly applied to patients with muscular or joint disorders that result in functional impairments. Following an acute illness or accident, many individuals prefer to pursue rehabilitation at home rather than consult a specialized practitioner or be admitted to a rehabilitation facility. This preference also applies to those with chronic conditions requiring maintenance rehabilitation, which can be challenging due to limitations on the number of sessions or the distance to specialized centres. The potential benefits extend beyond cost savings; home-based rehabilitation can be fragmented and repeated in a manner better suited to each individual's availability and capabilities. Its

effectiveness is further enhanced when performed with sufficient regularity and rigor [Kramer et al., 2019]. Nevertheless, despite the widespread adoption of home-based rehabilitation, challenges persist regarding patient motivation and adherence to prescribed exercise routines. These issues often lead to prolonged treatment durations and increased healthcare costs [Jack et al., 2010]. Several factors contribute to reduced patient motivation and engagement, with a primary concern being the lack of timely feedback and real-time supervision from healthcare professionals in a home environment [Miller et al., 2017]. To address these challenges, rehabilitation monitoring systems—devices and methods designed to enhance and track patients' rehabilitation processes—have been developed. These systems include motion-tracking devices to support physical therapy, cable-driven parallel robots for lower limb rehabilitation, and pulley-based exercise systems to regulate exercise execution, among others. Such technologies aim to provide more precise and objective monitoring of patients' progress while enhancing the efficiency and quality of rehabilitation [Giggins et al. 2013, Tamburella et al., 2019, Bowman et al., 2021].

In this study, we propose a deep learning system based on Temporal Conditional Generative Adversarial Networks (TCGAN) [Saito et al., 2017] to address a critical challenge in physical rehabilitation: the lack of real-time, patient-specific guidance during exercise execution. Traditional diagnostic methods, which predominantly rely on classification techniques, merely categorize movements as correct or incorrect after completion. These approaches fail to provide immediate corrective feedback, often leading patients to repeatedly perform erroneous movements, which can delay recovery, reinforce poor habits, and exacerbate injuries. Our system tackles this problem by generating skeleton-based rehabilitation data and dynamically superimposing it onto the patient's live video feed, offering a real-time visual guide tailored to his unique body structure and positioning. Unlike existing methods that depend on post-exercise analysis, our approach enables patients to instantly visualize and replicate precise movements, reducing the reliance on delayed or ambiguous feedback. The advantages of this system lie in its ability to minimize execution errors, prevent the reinforcement of incorrect techniques, and enhance patient engagement through an interactive experience. By providing continuous, adaptive support, it promotes safer, more effective, and more efficient recovery outcomes compared to conventional classification-based or manual supervision methods.

The remainder of this paper is organized as follows. A state-of-the-art review of rehabilitation systems is discussed in Section 2. In Section 3, we define our approach, dataset preparation, and model training. Results are explained in detail in Section 4. Section 5 provides a comparative study, comparing our method with existing solutions in the literature. The conclusion and future work are presented in Section 6.

2 Related Work

Rehabilitation systems based on Artificial Intelligence (AI) have been the focus of numerous scientific studies, with several research papers exploring their development, applications, and effectiveness [Kuncan et al., 2022a, Kuncan et al., 2022b, Miao et al., 2024]. Below are some key areas and related scientific works in the field of artificial intelligence-based rehabilitation systems.

2.1 Classification Techniques

In this section, we present classification-based methods used in physical rehabilitation. These methods primarily focus on analysing movements after they have been performed by the user. Once the movement is completed, the system classifies it and provides feedback based on the accuracy and quality of the execution. The goal of these approaches is to help users understand whether their movements were performed correctly and to offer guidance on how to improve or correct any detected errors. In the literature, we find essentially the following techniques:

2.1.1 Human Motion Analysis and Recognition

Research often focuses on the use of AI for human motion analysis and recognition, including the detection and classification of specific movements during rehabilitation exercises. For example, studies may employ computer vision techniques and machine learning algorithms to analyse body movements and provide real-time feedback. For instance, in [Lee et al., 2023], the authors developed an AI-driven real-time motion feedback system for patients with spinal cord injuries (SCI) to support rehabilitation, aiming to increase their engagement and motivation. The system's effectiveness in enhancing upper-limb muscle strength during TheraBand exercises was assessed. To facilitate this, a motion analysis program was developed using MediaPipe, which tracks exercise repetition counts and calorie consumption, focusing on three primary movements: chest press, shoulder press, and arm curl for upper extremity exercises. In [Maskeliūnas et al., 2023], the authors introduce BiomacVR, a rehabilitation system that integrates a Virtual Reality (VR) physical training monitoring environment with upper limb rehabilitation technology to enable accurate interactions and enhance patient engagement during training. The system employs a deep learning motion identification model known as the Convolutional Pose Machine (CPM), which utilizes a stacked hourglass network. This model is trained to accurately detect key points on the human body by analysing image sequences captured by depth sensors, enabling the identification of correct and incorrect movements while evaluating the effectiveness of physical training based on presented scenarios. In another study, [Biebl et al., 2021], the authors developed an application based on computer vision pose estimation techniques for knee osteoarthritis rehabilitation. Another method to assess physical rehabilitation exercises is proposed in [Du et al., 2021], where the authors suggested a deep learning framework using a graph convolutional network (GCN) with self-supervised regularization. In a different study on home-based rehabilitation for patients with stable chronic obstructive pulmonary disease (COPD), specific normalization techniques are applied by the authors to address positional discrepancies due to variations in body shape or camera distance [Guo et al., 2021]. The dynamic time warping (DTW) algorithm compensates for movement lag caused by reduced mobility in COPD patients. Additionally, a support vector machine (SVM) algorithm is used to develop an assessment model aligned with clinical expert evaluations. The paper [Bijalwan et al., 2023] presents an automated system for detecting and recognizing upper limb exercises using an RGB-depth camera, providing patients with real-time feedback during physiotherapy exercises. The method in [Miao et al., 2024] employs a monocular camera to capture video data of patients' upper limb rehabilitation exercises, using Faster R-CNN and HRNet to detect human body positions and key joint points

of the upper limb. A Long Short-Term Memory (LSTM) neural network model, enhanced with the ProbSparse Self-Attention mechanism, is then built to evaluate the rehabilitation training movements. In [Soni et al., 2023b], deep neural networks were used to train models for predicting joint contact forces using only joint angles as input features, employing an Explainable Convolutional Model (XCM) architecture as the classifier. In [Yu et al., 2024], the authors present an advanced framework, EGCN++ (Ensemble-based Graph Convolutional Network), for skeleton-based exercise assessment. Building on EGCN, they introduce a new fusion strategy called MLE-PO, which integrates fusion at both the data and model levels. To validate their approach, they conduct extensive cross-validation experiments and analyse the consistency between machine and human evaluations across three datasets. In another study using skeletal data [He et al., 2024], an Expert-knowledge-based Graph Convolutional approach is proposed to automate the assessment of physical rehabilitation exercise quality. This approach leverages expert knowledge to enhance the spatial feature extraction capabilities of the Graph Convolutional module and incorporates a Gated pooling module for feature aggregation. Additionally, a Transformer module is employed to capture long-range temporal dependencies in the movements. The attention scores and weight matrix obtained through this approach can serve as interpretability tools, helping therapists understand the assessment model and guiding patients in improving their exercises.

2.1.2 Wearable Devices for Rehabilitation

Wearable devices equipped with AI capabilities are explored for monitoring and assisting individuals during rehabilitation. Research in this area may involve the development of smart sensors, wearable AI systems, and the analysis of data collected from these devices to track and improve rehabilitation outcomes [Qiu et al., 2022]. For example, the study in [Bavan et al., 2019] aims to assess the effectiveness of a fully automated classification system designed to recognize analytical shoulder rehabilitation activities performed by patients with subacromial shoulder pain. Supervised machine learning models are used to classify the shoulder movements. In another work, [Espinoza et al., 2021], the authors developed personalized classification algorithms, including those based on neural networks, to categorize four rehabilitation exercises performed by two stroke patients attending a week-long therapy camp in Jamaica. Accelerometery-based wearable sensors were attached to each upper and lower limb to collect movement data throughout the therapy sessions. A wearable system that utilizes lightweight stretch sensors to estimate the spinal posture of patients performing physical therapy exercises is proposed in [Chen et al., 2021]. One primary function of this system is to detect single-axis spinal movements based on the sensor readings. In [Burns et al., 2021], a machine learning approach was employed to evaluate overall physiotherapy participation using inertial data from a smartwatch. Similarly, researchers in [Lee et al., 2024] aimed to classify 11 types of shoulder rehabilitation exercises in patients with shoulder pain using an artificial intelligence (AI) algorithm. Patients wore an IMU-based sensor to collect data during exercises, enabling the calculation of exercise classification accuracy. The study in [Nair and Sakthivel, 2023] presents a deep learning-based system designed to assist patients in identifying the completion status of twelve hand and arm rehabilitation exercises. This task is formulated as a multi-class classification problem, with limb movement data collected

for each exercise using accelerometers. In [Kuncan et al., 2019], the authors proposed a novel feature extraction approach for the Human Activity Recognition (HAR) system based on mobile sensor signals and a two-stage Down Sampling One-Dimensional Local Binary Pattern (DS-1D-LBP) method. Authors in [Soni et al., 2023a] conducted research in the field of HAR using mobile devices. Their study proposes a deep neural network (DNN) architecture that combines a bidirectional gated recurrent unit (BiGRU) with two convolutional layers. The raw data collected from smartphone sensors, is first processed through a two-layer BiGRU, followed by two CNN layers. The model's performance is evaluated using two publicly available datasets.

2.1.3 Virtual Reality and Gamification

Researchers investigate how AI algorithms enhance the immersive experience of virtual environments, making rehabilitation exercises more engaging and motivating for patients. In [Martinez-Martin and Cazorla, 2019], a computer-based virtual reality rehabilitation platform called the Medical Interactive Rehabilitation Assistant (MIRA) is proposed. A Classification and Regression Tree (CART) method is used to predict optimal input settings for patients playing MIRA games. This approach improves the efficiency of standard default prediction methods in determining ideal exergame settings to enhance patient performance. Another system is proposed in [Macintosh et al., 2020], where movement-based video games provide engaging practice for repetitive therapeutic gestures, aiming to improve manual ability in youth with cerebral palsy. A random forest classification technique is utilized for in-game calibration and classification procedures to select the most discriminating, person-specific features. A motion rehabilitation platform that employs games and biophysical sensors is presented in [Morando et al., 2018]. Key features of the system include motion tracking using Microsoft Kinect V2 and Leap Motion, with comparisons to alternative solutions. Additionally, the patient's emotional state is assessed through heart rate monitoring and electrodermal activity using Microsoft Band 2, all recorded during functional exercises prescribed by the therapist. In another study, [Alhwoaimel et al., 2024], the authors present a game-based rehabilitation system designed for upper-limb cerebral palsy, featuring three interactive exercises and a computerized assessment method. These exercises aim to engage participants in physical movements such as shoulder flexion, horizontal abduction/adduction, and right arm adduction. The system employs a Kinect sensor to track the participant's skeletal joints, enabling interaction with the game-based rehabilitation program. The study by [Castillo et al., 2024] introduces a playtesting methodology to enhance the design of a VR exergame developed using a user-centred approach for upper limb rehabilitation in stroke survivors. Across four playtesting sessions, stroke survivors engaged with early game versions using VR headsets, providing valuable feedback to refine game content and mechanics. Additionally, a pilot study involving 10 stroke survivors collected data through VR-related questionnaires to evaluate key game design aspects, including mechanics, assistance, user experience, motion sickness, and immersion.

2.2 Use of Generative Systems

In addition to classification techniques, several studies utilize generative AI methods for data generation in the field of physical rehabilitation. These approaches primarily

aim to augment existing datasets to enhance training for classification models. By generating synthetic data that closely replicates real-world rehabilitation scenarios, they address challenges such as limited or imbalanced datasets. This data augmentation improves the model's generalizability across diverse patient populations and exercise variations, ultimately enhancing the accuracy and robustness of movement recognition systems in rehabilitation. In [Boukhennoufa et al., 2023], the authors tackled the issue of mode collapse in GAN-based (Generative Adversarial Network) data augmentation techniques for post-stroke assessment. They applied a GAN to generate synthetic data from two post-stroke rehabilitation datasets. Results show that the proposed solution achieves a significant increase in classification accuracy (from 35.2% to 42.07%) for both selected datasets. The study in [Hadkey et al., 2024] utilizes Conditional Generative Adversarial Networks (CGANs) to generate synthetic kinematic data based on a publicly available dataset, closely replicating the reaching movements observed in stroke survivors. This method not only captures the intricate temporal dynamics and typical movement patterns seen post-stroke but also significantly enriches the training dataset. Models incorporating synthetic data attained an overall accuracy of 80.2%, substantially higher than the 63.1% observed in models trained solely with real data. The authors in [Bicer et al., 2022] employed a GAN to augment datasets for deep learning-based biomechanical models. While these models perform best when trained on large datasets, collecting such data in gait laboratories can be challenging, and existing data augmentation techniques often have limitations. To address this issue, the study proposes a data augmentation approach using generative deep learning to enhance biomechanical datasets. A Conditional GAN was developed to synthesize biomechanical data during gait. In another study, [Kárason et al., 2023], a CGAN was used to synthesize biomechanical data during gait. The study in [Peppes et al., 2023] focuses on a data augmentation solution employing GANs, using a freezing of gait (FoG) symptom dataset as input. In [Hadley and Pulliam, 2024], the authors used GANs to create synthetic accelerometry data representing both typical and freezing gait patterns for Parkinson's disease. [Mennella et al., 2023] introduced a new synthetic dataset for rehabilitation exercises, utilizing pose-guided person image generation through conditioned diffusion models. By analysing a pre-labelled dataset of class movements for six rehabilitation exercises, the proposed method produces realistic images of elderly individuals performing home-based exercises.

While many physical rehabilitation methods provide valuable insights into movement recognition, they often have a significant drawback: feedback is typically provided only after the movement has been performed. As a result, patients may repeat incorrect movements multiple times because the system does not demonstrate the correct way to perform the exercise. In contrast, although the existing literature on GANs primarily focuses on their application for augmenting datasets, our approach diverges by utilizing GANs to enhance the quality of rehabilitation exercises. Rather than simply increasing the volume of data, we aim to generate tailored movements specifically adapted to each patient's unique needs. By generating accurate and context-aware exercise patterns, our methodology enhances rehabilitation effectiveness, helping patients perform movements correctly and safely. This targeted approach not only accelerates recovery but also improves overall outcomes, ensuring that individuals receive personalized care aligned with their rehabilitation goals.

3 Methodology

Our approach employs a Temporal Conditional Generative Adversarial Network (TCGAN) to generate a dynamic scene depicting a physical rehabilitation exercise through a skeletal model. The patient positions himself in front of a camera, which captures him, and the keypoints are extracted. Using this data, the TCGAN generates a real-time visualization of the exercise movements, adapted to the patient's size and position. This generated skeletal model (keypoints) is then superimposed onto the patient's live image on a monitor. As the skeletal model executes the movements of a rehabilitation exercise, the patient can mirror these actions, ensuring proper execution. This interactive approach enhances the rehabilitation process. To achieve this result, our system follows these steps (Figure 1):

1- Keypoints extraction (Using MoveNet[1]),
2- Generation of synthetic movements (using TCGAN),
3- Superimposing the skeleton onto real-time patient video and displaying it on a monitor.

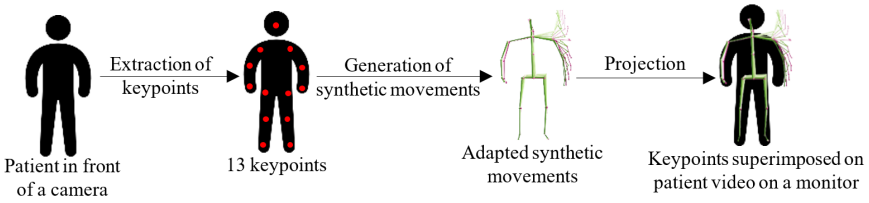

Figure 1: Proposed approach

3.1 Dataset

To validate our approach, we created a dataset in collaboration with a physiotherapist. The dataset comprises skeletal data collected from twenty (20) individuals of varying heights and ages (10 to 70 years) in different locations and lighting conditions. This diversity ensures a broad representation of rehabilitation movements performed either while standing or sitting. The data were acquired using an RGB camera, and keypoints were extracted with MoveNet (Figure 2). Each participant performed multiple repetitions of nine rehabilitation exercises: left elbow flexion, right elbow flexion, left shoulder flexion, right shoulder flexion, left shoulder abduction, right shoulder abduction, anterior shoulder elevation, left lateral touch, and right lateral touch.

After cleaning and removing incomplete data, the final dataset contains 1869 gesture sequence samples (Table 1), each consisting of an average of 50 frames (where a sequence represents one repetition of an exercise). However, in our rehabilitation study, we opted not to use eye keypoints. This decision was based on the fact that eye movements do not play a crucial role in the rehabilitation process we are analysing. By focusing on other joints and body segments, we can more effectively assess the movements and postures essential for the physical recovery of patients. Consequently,

[1] https://www.tensorflow.org/hub/tutorials/movenet

each sample includes 13 keypoints out of the 17 provided by MoveNet (Figure 2), with each represented by two coordinates, x and y.

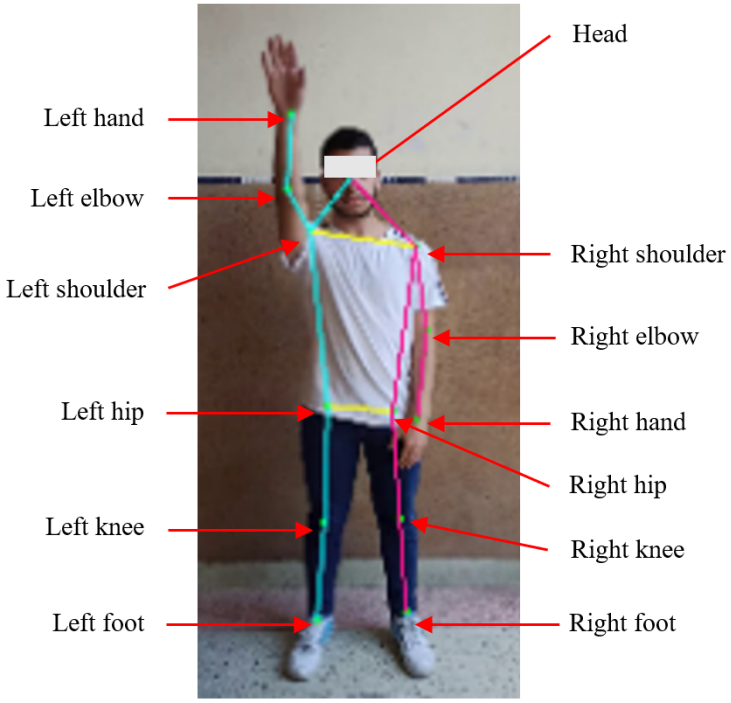

Figure 2: Used keypoints

Movement	**Number of sequences**	**Percentage**
Left elbow flexion	219	11.72%
Right elbow flexion	191	10.22%
Left shoulder flexion	213	11.40%
Right shoulder flexion	220	11.77%
Left shoulder abduction	199	10.65%
Right shoulder abduction	207	11.08%
Anterior shoulder elevation	202	10.81%
Left lateral touch	217	11.61%
Right lateral touch	201	10.75%
Total	1869	100%

Table 1: Number of sequences by exercise

To enhance the training stability of our TCGAN, we applied a normalization technique to our data. This process involves dividing the coordinates of the keypoints

by the dimensions of the image, yielding normalized values within the range [0, 1]. Mathematically, this normalization can be represented by the following equations:

$$X_{norm} = \frac{x}{width} \quad (1)$$

$$Y_{norm} = \frac{y}{height} \quad (2)$$

In these equations, X_{norm} and Y_{norm} represent the normalized coordinates of the keypoints relative to the image's width and height. This normalization provides several benefits, including improving the stability of the learning algorithm's convergence and reducing the risk of overfitting.

3.2 Keypoint Extraction

After using the MoveNet model for dataset preparation, we apply it again in the first step of our process to generate adapted rehabilitation exercises. MoveNet is an advanced pose detection model that accurately extracts keypoints from an image of a person positioned in front of a camera. When an image is captured, MoveNet analyses the body's position and orientation by identifying keypoints such as the shoulders, elbows, hips, and knees. Utilizing deep learning, MoveNet can precisely detect these keypoints, even in challenging lighting conditions or complex backgrounds. As noted in the previous paragraph, we use 13 keypoints to represent the person's body. These keypoints serve as inputs for our TCGAN, which then generates new keypoints corresponding to various movements for the chosen rehabilitation exercise.

3.3 TCGAN Architecture

Our TCGAN (Figure 3) utilizes two neural networks and an adversarial training procedure. Typically, in a CGAN, a generator G and a discriminator D are trained simultaneously to engage in the following min-max game [Saito et al., 2017]:

$$min_G\ max_D\ V(D,G) = E_{x\sim pdata}[log\ D(x|y) + E_{z\sim pz}\ [log(1 - D(G(z|y)|y))] \quad (3)$$

In this formulation, the generator G ($G_0 + G_1$) creates synthetic data samples based on a latent vector z_0 drawn from a Gaussian distribution and a conditional label y. The output G(z|y) is a rehabilitation exercise scene conditioned on y, allowing G to produce data relevant to the current patient. The discriminator D is responsible for differentiating between real data samples and those generated by G, outputting the probability that a given sample x is real, conditioned on y, represented as D(x|y). The first term, $E_{x\sim pdata}[logD(x|y)]$, calculates the expected logarithm of the discriminator's output for real samples, encouraging D to assign high probabilities to actual data and improving its accuracy in identifying real samples. Conversely, the second term, $E_{z\sim pz}[log(1-D(G(z|y)|y)]$, evaluates D's output for synthetic samples, aiming to reduce the likelihood that the discriminator recognizes these samples as real. This adversarial setup constitutes a minimax game where D seeks to maximize its performance in correctly classifying samples, while G aims to minimize this success by generating increasingly realistic data. Ultimately, this objective function encapsulates the

adversarial dynamics between G and D, facilitating the generation of high-fidelity, conditionally relevant data.

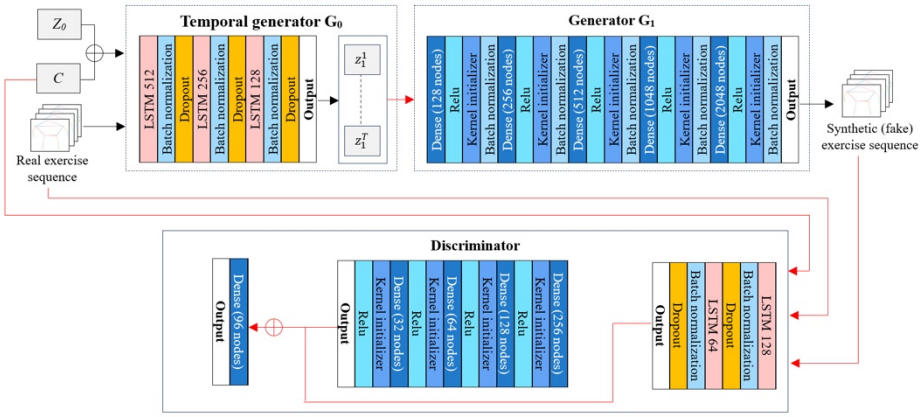

Figure 3: Illustration of the proposed TCGAN. The exercise sequence generator consists of two generators: the temporal generator, G_0, and the image generator, G_1. The temporal generator yields a set of latent variables, z_1^t (t = 1, ..., T), from z_0 and from a vector of conditions. The image generator transforms these latent variables, z_1^t (t = 1, ..., T), and z_0 into sequence data that has T frames. The discriminator consists of two parts (temporal aspect validation and keypoints validation); it evaluates whether these frames are from the dataset or from the generator.

In what follows, we detail the different components of our TCGAN. We begin by presenting the architecture of the temporal generator, followed by the generator and finally the discriminator.

3.3.1 Temporal Generator

The proposed temporal generator in this study is a crucial component of our TCGAN model, specifically designed to produce sequences of movements represented by bodily keypoints. The temporal generator G_0 in our TCGAN architecture generates a set of latent variables z_1^t(for t=1,…,T) from an initial latent vector z_0, where T represents the number of frames in the exercise sequence. The inputs to G_0 include three components: the latent vector z_0 with a shape of 100, a sequence of frames shaped (T,13×2) where each frame consists of 13 keypoints with (x,y) coordinates, and a condition vector with a shape of 8 (condition_dim = 8, encompassing the exercise number and coordinates of the head and feet to determine the patient's size and position in the video). This sequence of real skeletal keypoints provides essential temporal context for generating future movements.

To capture temporal dependencies in the skeletal data, we use LSTM (Long Short-Term Memory) layers. LSTMs are effective for this task because they can learn long-range dependencies in sequential data, which are critical for accurately modelling human movements over time. Given a sequence of skeletal keypoints $X=(x_1,x_2,\ldots,x_t)$, where $x_t \in \mathbb{R}^{13}$ represents the 13 keypoints at time step *t,* the LSTM learns a mapping

where h_t is the hidden state at time t, F is the function defined by the LSTM gates (forget, input, and output), and θ represents the LSTM parameters (weights and biases). The input sequence spans T frames, each consisting of 13 skeletal keypoints representing human body joints at a specific time step. The function F can be expressed as follows:

$$h_t = F(h_{t-1}, x_t; \theta) \tag{4}$$

where h_{t-1} is the hidden state from the previous time step and x_t is the input at the current time step. The temporal generator processes these inputs through three LSTM layers with 512, 256, and 128 units, respectively, followed by batch normalization and dropout layers to ensure regularization and training stability. The output of the LSTM layers has a shape of (T,128), a 2D array representing the hidden state for each time step. After the LSTM layers, the output is flattened using a Flatten layer, converting the 2D output (T,128) into a 1D vector of shape (T×128). These latent variables are then passed to the image generator G_1, which produces a complete sequence of generated skeletal keypoints across T frames. This process ensures that the generated keypoints are temporally consistent and accurately represent the movements for the specified rehabilitation exercise. The final output is a smooth sequence of keypoint-based frames with a shape of (T, 13×2) that guides patients in performing the correct movements during rehabilitation.

3.3.2 Generator

The data generator G_1 in our TCGAN serves a critical function by synthesizing realistic movement sequences represented by skeletal keypoints. This is particularly significant for applications in physical rehabilitation, where accurate motion patterns are essential for patient recovery. The generator's architecture is carefully designed to produce context-aware and temporally coherent sequences that mimic real human motion. By transforming latent vectors into structured outputs, the generator enables the creation of personalised exercises tailored to individual rehabilitation needs.

The inputs to the generator include the outputs from the temporal generator, specifically a sequence of latent variables, z_1^t (t = 1, ..., T), with a shape of (T, 128). The generator comprises a series of fully connected (dense) layers. Each dense layer is designed to learn complex representations of the input data through weighted linear combinations followed by a non-linear activation function. The output of each layer is computed as follows:

$$h = f(Wx + b) \tag{5}$$

Where h is the output, W is the weight matrix, x is the input to the layer, b is the bias term, and f is the activation function, which is a Rectified Linear Unit (ReLU). The ReLU activation function is defined as:

$$f(x) = max(0, x) \tag{6}$$

This function introduces non-linearity into the model, allowing it to learn complex mappings from the input to the output space. We favoured ReLU for its simplicity and

efficiency, helping to mitigate the vanishing gradient problem often encountered in deep networks. The architecture of G_0 includes multiple dense layers, configured with 128, 256, 512, 1024, and 2048 units, respectively. Each of these layers is crucial for progressively transforming the features extracted from the previous layer into the desired output representation of skeletal keypoints. The dense layers enable the generator to capture and refine intricate relationships within the data, ensuring that the final output accurately reflects realistic motion patterns. To enhance training stability and performance, Batch Normalization layers are integrated between the dense layers. Batch Normalization standardizes the inputs to each layer, stabilizing the learning process and enabling higher learning rates. The formula for Batch Normalization can be expressed as:

$$\hat{x} = (x - \mu) / sqrt(\sigma^2 + \varepsilon) \tag{7}$$

Where μ is the mean, σ^2 is the variance of the input batch, and ε is a small constant to prevent division by zero. This normalization helps the network learn more effectively by reducing internal covariate shift. The generator also incorporates kernel initializers, which define the initialization of the weights in the dense layers. Our choice is the "He" initializer, particularly suitable for layers using ReLU activations. The “He” initializer sets the weights according to:

$$W \sim N(0, 2 / n) \tag{8}$$

where n is the number of inputs to the layer. This initialization strategy ensures that the weights are appropriately scaled, facilitating effective learning in the early stages of training. The output of the final dense layer is reshaped to produce a sequence of coordinates representing the skeletal keypoints over time. This transformation ensures that the generator outputs sequences conforming to the expected dimensionality, which is crucial for the rehabilitation context.

In summary, the generator in this TCGAN framework is designed to synthesize realistic sequences of skeletal movements effectively. Its architecture leverages LSTM layers to preserve temporal dependencies between frames and dense layers to capture intricate relationships within the data. Additionally, techniques such as Batch Normalization and kernel initialization enhance training stability and performance.

3.3.3 Discriminator

The TCGAN discriminator integrates spatial and temporal pathways to assess the authenticity of generated skeletal sequences, conditioned on specific movement attributes. Its input consists of the condition vector alongside real and synthetic data. The spatial discriminator utilizes a series of dense layers structured as follows: a 256-neuron layer, followed by layers with 128, 64, and 32 neurons. Each dense layer employs the ReLU activation function, introducing non-linearity to help the model learn complex spatial relationships between keypoints in each frame. To enhance training efficiency, a kernel initializer optimizes the initial weight distribution. The temporal discriminator captures dynamic dependencies within the sequence using two LSTM layers: the first with 128 units and the second with 64 units, processing the skeletal sequence’s temporal aspect. Each LSTM layer is followed by batch

normalization to stabilize training by normalizing layer outputs, improving convergence. Additionally, a dropout layer mitigates overfitting by randomly deactivating a subset of neurons during training.

The fused output from the spatial and temporal pathways is passed through a fully connected layer (96 neurons = 64 + 32) with a sigmoid activation function, producing a probability score that indicates whether the sequence is real or generated. This integrated architecture enables the TCGAN to effectively evaluate both spatial accuracy and temporal consistency, ensuring realistic and high-quality exercise generation for rehabilitation applications.

3.4 Loss Function

To ensure the generated skeletal sequences exhibit smooth transitions over time in our TCGAN, we introduce a smoothness regularization term:

$$L_{smooth} = \sum_{t=1}^{T-1} |x_{t+1} - x_t|^2 \tag{9}$$

xi are the skeletal keypoints at consecutive time steps. This term guarantees fluid and continuous movements in the generated temporal sequences. Additionally, to handle imbalances in the dataset, a weighted cross-entropy loss is applied to the discriminator:

$$L_D = -\alpha \, . \, E_{x\sim pdata}[\log D(x, y)] - (1 - \alpha) \, . \, E_{z\sim pz}[\log(1 - D(G(z, y)))] \tag{10}$$

where α adjusts the importance of real versus generated samples, and y represents the conditional label that defines the action class. This allows the discriminator to distinguish between real and generated sequences while incorporating in the given action label. To further stabilize TCGAN training, we apply a gradient penalty as used in Wasserstein GANs with Gradient Penalty (WGAN-GP):

$$L_{GP} = \lambda \, . \, E_{\hat{x}\sim P\hat{x}}[(|\nabla_{\hat{x}} D(\hat{x}, y)|_2 - 1)^2] \tag{11}$$

∇ denotes the gradient operator, which computes the derivative of a function with respect to its input. Specifically, $\nabla_{\hat{x}} D(\hat{x}, y)$ calculates the gradient of the discriminator's output $D(\hat{x}, y)$ with respect to the input $\hat{x}$, where $\hat{x}$ is either a real or generated skeletal movement sequence. y refers to the condition vector, and λ is a hyperparameter that controls the strength of the gradient penalty. It scales the penalty term, ensuring that the gradient norm of the discriminator's output with respect to $\hat{x}$ is close to 1, which stabilizes the training process. This acts as a regularization term to avoid excessive gradients, ensuring the discriminator does not overfit. It enforces a Lipschitz constraint on the discriminator, improving the stability of the training process and ensuring more realistic sequence generation. For the generator, its goal is to produce sequences that can both fool the discriminator and adhere to the conditional label y. The adversarial loss for the generator is defined as:

$$L_G = E_{z\sim pz}[\log D(G(z, y))] \tag{12}$$

where:

- z is the latent noise vector sampled from a Gaussian distribution $z \sim N(0,1)$,
- $D(G(z,y))$ represents the discriminator's prediction on the generated sequence conditioned on y.

The generator works to minimize this loss to ensure the generated skeletal sequences correspond to the specified action class and are indistinguishable from real data. Finally, the total loss function for training the TCGAN combines all these components, ensuring smooth temporal transitions, balanced learning for real and fake samples, and stable TCGAN training as presented below:

$$L_{total} = L_D + L_G + \lambda_{smooth} L_{smooth} + \lambda_{GP} L_{GP} \tag{13}$$

Here:

- L_D is the discriminator loss,
- L_G is the generator's adversarial loss,
- λ_{smooth} and λ_{GP} are regularization coefficients for smoothness and gradient penalty, respectively.

This loss formulation establishes a balanced mechanism for learning temporally consistent and realistic skeletal sequences while conditioning on specific action labels to effectively guide the generation process.

3.5 Training

The algorithm for training a TCGAN (Algorithm 1) focuses on generating realistic skeletal sequences for physical rehabilitation exercises. It begins by initializing the generator and discriminator networks with random weights and defining critical hyperparameters, including the learning rate, batch size, and sequence length. During each training epoch, the dataset is shuffled, and batches of real skeletal sequences and their labels are sampled. The generator then creates synthetic data by generating latent vectors from a Gaussian distribution and concatenating them with conditional labels, producing fake skeletal sequences. The discriminator is trained to distinguish between real and fake data, with a mechanism to temporarily disable its training if the loss falls below a threshold, thus preventing overfitting. Singular value clipping is applied to enhance stability, followed by weight updates based on computed gradients. Afterward, the generator is trained to minimize its loss, enhanced by smoothness regularization to ensure continuity in movements. Upon completion of the training epochs, the trained generator and discriminator are returned, enabling the generation of tailored skeletal movements that significantly improve the rehabilitation process by offering personalized exercise guidance.

Algorithm 1: Pseudo algorithm of TCGAN training process

Inputs: Training dataset of skeletal sequences + labels *Y*.

Output: Trained generator **G** and discriminator **D**

Data (Hyperparameters):

Learning rate $\boldsymbol{\eta}$ for both generator and discriminator (the same value for G and D)

Batch size **B,** Number of training epochs **E,** Sequence length **T**

Regularization coefficients $\boldsymbol{\lambda_{smooth}}$ and $\boldsymbol{\lambda_{GP}}$

Maximum singular value threshold $\boldsymbol{\tau}$

Discriminator loss threshold $\boldsymbol{threshold_D}$ to temporarily disable training.

Step 1: Initialize G and D

Initialize the generator **G** and discriminator **D** networks with random weights.

Set the maximum singular value threshold $\boldsymbol{\tau}$.

for each epoch *e* from 1 to *E* **do**

  Shuffle the training dataset **D**.

  for each batch *(X, Y)* in *D* **do**

    a. Sample Real Data

    Randomly sample a batch of real skeletal sequences X_{real} and their corresponding labels Y_{real}.

    b. Generate Synthetic Data

    Sample latent vectors z_0 from a Gaussian distribution.

    Concatenate z_0 with the conditional labels Y_{fake}.

    Generate fake skeletal sequences with the generator: $X_{fake}=G(z_0, Y_{fake})$.

    c. Train the Discriminator

    c.1.Temporarily disable discriminator training if the loss is very low:

      Calculate discriminator outputs for real and fake data:

        $D_{real}=D(X_{real},Y_{real})$

        $D_{fake}=D(X_{fake},Y_{fake})$

      Compute the discriminator loss:

        $L_D=-\alpha\cdot E[\log D_{real}]-(1-\alpha)\cdot E[\log(1-D_{fake})]$

      if $L_D < threshold_D$ **then**:

        Disable discriminator training for this batch

      Else:

        Enable discriminator training

    c.2. Singular Value Clipping:

      for each weight matrix *W* in *D* **do**

        Perform SVD: $W=U\Sigma V^T$

        Clip singular values: $\Sigma'=clip(\Sigma,0,\tau)$

        Reconstruct weights: $W'=U\Sigma' V^T$

    c.3. Update discriminator weights: $W \leftarrow W-\eta\nabla L_D$

    d. Train the Generator

      Compute discriminator output for fake data: $D_{fake} = D(X_{fake},Y_{fake})$

      Compute generator loss*:* $L_G= -E[\log D_{fake}]$

      Apply smoothness regularization (if applicable):

        $L_{smooth} = \sum_{t=1}^{T-1} ||x_{t+1}-x_t||^2$

      Update generator weights: $W \leftarrow W-\eta\nabla(L_G+\lambda_{smooth})$

4 Results

This section provides an analysis of the data obtained from the experimental procedures, examining the numerical results, measurements, and statistical findings derived from our study.

4.1 Quantitative Model Evaluation

To evaluate the performance of our TCGAN, we utilize the Fréchet Inception Distance (FID), which measures the similarity between the distributions of real and generated images. We also apply four machine learning algorithms to classify the synthetic data in relation to the real data. These algorithms evaluate the quality of the generated samples using metrics such as accuracy and F1 score, providing a comprehensive validation of the TCGAN's output. Additionally, we employ Dynamic Time Warping (DTW) to test the sequences, assessing the temporal alignment and similarity between real and generated motion patterns.

4.1.1 Frechet Inception Distance

The Fréchet Inception Distance (FID) [Dowson and Landau, 1982] is a measure of similarity between two sets of images. It is used to evaluate the quality of generated data by the proposed TCGAN. FID measures how similar the feature distributions of real and generated data are. FID is expressed by the following equation, which has been adapted for numerical data (keypoints):

$$FID = ||\mu 1 - \mu 2||2 + Tr\,(C1 + C2 - 2 \times \sqrt{C1 \times C2}) \tag{14}$$

where μ_1 and μ_2 represent the features means of the real and generated data, respectively, and C_1 and C_2 are the covariance matrices of the feature vectors of the real and generated data, respectively. Tr denotes the trace function in linear algebra [Solanki et al., 2021]. The interpretation of FID values is essential for evaluating the quality of data generated by models like TCGANs. An FID score close to zero indicates that the generated data are nearly indistinguishable from real data. In the context of our TCGAN for generating synthetic skeletal sequences, a variety of hyperparameters must be established beforehand to define both the model architecture and the training process. These hyperparameters (Table 2) are critical, as they significantly affect the performance, stability, and overall effectiveness of the TCGAN in generating realistic and context-sensitive rehabilitation exercises. The careful selection and fine-tuning of these parameters are vital for enhancing the model's accuracy and its ability to generalize across different rehabilitation scenarios. Furthermore, a detailed experimental approach and a systematic hyperparameter search were necessary to identify the optimal settings for our specific task. Below, we present a comprehensive list of the key hyperparameters employed in our experimentation, along with their respective definitions and chosen values:

Count of LSTM Layers: The model architecture incorporates three LSTM layers for the generator (G) and two for the discriminator (D), chosen for their capacity to effectively capture temporal dependencies in sequential data.
Dropout Rate: Set at **0.3**, this hyperparameter is employed to prevent overfitting by

randomly dropping a fraction of neurons during training. A moderate dropout rate helps maintain the model's generalization capability while ensuring it learns robust representations from the training data.

Hyperparameter	**Range**	**Setup value**
Count of LSTM Layers for G	**1, 2, 3, 4, 5**	**3**
Count of LSTM Layers for D	**1, 2, 3, 4, 5**	**2**
Dropout Rate	**0.2, 0.3, 0.4, 0.5, 0.6**	**0.3**
Learning rate η ((for G and D)	**0.1, 0.01, 0.001, 0.0001, 0.0002**	**0.0002**
Batch size	**32, 64, 128, 256**	**64**
Epochs	**3000, 3500, 4000, 4500, 5000**	**4000**
Discriminator loss threshold $threshold_D$	0.01, 0.05, 0.1, 0.2, 0.3, 0.4	**0.2**
Maximum singular value threshold τ	**0.05, 0.1, 0.2, 0.3, 0.4, 0.5**	**0.1**
Smoothness regularization coefficient λ_{smooth}	0.05, 0.1, 0.2	**0.1**
Gradient penalty coefficient λ_{GP}	**5, 10, 15**	**10**

Table 2: Hyperparameters description with their values used in experimentation

Learning Rate: A learning rate of **0.0002** is applied to both the generator and the discriminator. This value strikes a balance between convergence speed and stability, facilitating effective training while minimizing the risk of overshooting during optimization.
Batch Size: The batch size is set to **64**, enabling the model to process a substantial number of samples per iteration. This enhances training stability and provides more reliable gradient estimates, ultimately contributing to improved performance.
Epochs: Training is conducted over **4000** epochs, allowing the model to learn complex data relationships. This extended duration is crucial for achieving high-quality outputs in a generative framework.
Discriminator Loss Threshold: A threshold of **0.2** is utilized to temporarily disable discriminator training when its loss drops below this value, preventing the discriminator from overpowering the generator. This mechanism is critical for maintaining balanced training dynamic between both components.
Maximum Singular Value Threshold: set at **0.1** to ensure discriminator training stability. By clipping the singular values of the weight matrices, the model enforces Lipschitz continuity, which is crucial for effective TCGAN training.
Smoothness Regularization Coefficient: A value of **0.1** is chosen to promote smoothness in the generated movements. This regularization helps prevent abrupt transitions, enhancing the realism of the generated skeletal sequences.
Gradient Penalty Coefficient: The gradient penalty coefficient is set to **10**, playing a vital role in stabilizing the discriminator's training.

Careful selection and tuning of these hyperparameters have significantly improved our TCGAN's performance. Notably, these optimizations yielded the lowest Fréchet Inception Distance (FID) score of 0.89. This low FID value indicates that the generated

skeletal sequences closely resemble the real data distribution, demonstrating the model's effectiveness in producing high-quality, realistic movements for rehabilitation exercises.

4.1.2 Machine Learning Algorithms

Evaluating TCGAN performance using the Fréchet Inception Distance (FID) is an effective method for assessing the quality of generated data. However, FID alone is insufficient to fully ensure the overall quality of TCGAN. To comprehensively validate our model's capabilities, we employed four machine learning classification algorithms: K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Decision Trees (DT), and Random Forests (RF). These algorithms allow us to evaluate how closely the synthetic data generated by our TCGAN resembles real data in terms of quality. To train them, we used the same real dataset employed in our TCGAN training.

Dataset size	Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
1000	SVM	99.33	99.30	99.41	99.35
	KNN	99.77	100	100	100
	DT	99.12	99.50	99.67	99.58
	RF	98.87	99.22	99.37	99.29
10000	SVM	99.61	99.47	99.45	99.46
	KNN	99.85	99.63	100	99.81
	DT	99.29	99.10	99.23	99.16
	RF	99.00	99.26	99.33	99.29
100000	SVM	99.79	98.59	99.49	99.04
	KNN	99.53	99.54	99.66	99.60
	DT	99.62	99.09	99.23	99.16
	RF	99.22	99.65	99.70	99.67
500000	SVM	99.89	99.41	99.66	99.53
	KNN	99.91	100	100	100
	DT	99.76	99.59	99.71	99.65
	RF	99.62	99.28	99.39	99.33

Table 3: Test results on datasets with different sizes

Afterward, we tested the algorithms on synthetic datasets of varying sizes, generated by our TCGAN. The evaluation relied on standard machine learning metrics, including accuracy, precision, recall, and F1 score, each providing unique insights into classification performance on the generated data. Using classification models to assess our generative TCGAN across datasets ranging from 1000 to 500000 samples is crucial for evaluating its adaptability and ability to generate high-quality data at scale. This approach not only verifies that our TCGAN can effectively produce datasets of different

sizes but also helps identify potential limitations. Testing across varying dataset sizes ensures robustness and reliability, making the model highly applicable to real-world scenarios where data diversity is critical.

Table 3 presents the classification results for datasets of 1000 (1K), 10000 (10K), 100000 (100K), and 500000 (500K) samples. The machine learning algorithms consistently achieved high accuracy in distinguishing synthetic from real data. However, this strong performance does not indicate overfitting, as we employed rigorous validation techniques, including cross-validation and testing on independent sub-datasets. Instead, the high accuracy highlights the TCGAN's ability to generate realistic samples.

Overall, the analysis confirms that the generated data closely mirrors real data, with accuracy exceeding 99%. These findings demonstrate the exceptional effectiveness of our TCGAN in producing high-quality synthetic skeletal data, reinforcing its applicability to rehabilitation exercises and beyond.

4.1.3 Dynamic Time Warping Score

Unlike the previous section, which used machine learning algorithms to assess the quality of generated data by analysing individual frames, this section evaluates entire sequences produced by the TCGAN. These generated sequences are compared to those in the real dataset using the Dynamic Time Warping (DTW) score [Berndt and Clifford, 1994]. DTW is a method designed to measure the similarity between two temporal sequences that may vary in speed or length. In our case, it compares sequences of keypoints—each consisting of an average of 50 frames, with each frame containing 13 keypoints × 2 coordinates—to quantify variability within real sequences (real vs. real) and differences between real and synthetic sequences (real vs. synthetic).

DTW identifies an optimal alignment between two sequences, S1 = [$S_{1,1}$, $S_{1,2}$, …, $S_{1,n}$] and S2 = [$S_{2,1}$, $S_{2,2}$, …, $S_{2,m}$], by minimizing the cumulative distance between them. The basic DTW formula is computed using dynamic programming:

- **Distance between two points**: For each frame i and j, the Euclidean distance is calculated over the 26 values (13 keypoints × (x, y)):

$$d(s_{1,i,k}\text{-}s_{2,j,k}) = \sqrt{\sum_{k=1}^{26}\left(s_{1,i,k}\text{-}s_{2,j,k}\right)^2} \tag{15}$$

- **Cost matrix:** A matrix *DTW[i, j]* is constructed recursively:

$$DTW[i,j]=d(s1,i,s2,j)+min(DTW[i-1,j],DTW[i,j-1],DTW[i-1,j-1])DTW[i,j] \tag{16}$$

with DTW[0,0]=0, and the edges initialized to infinity.

- **Final score:** *DTW [n,m]* represents the total DTW distance, indicating the minimal dissimilarity after alignment.

To ensure the robustness and reliability of our results, we generated five synthetic datasets, each matching the size of the real dataset and containing the same number of sequences per class. This approach allowed for a direct and realistic evaluation of the model's fidelity. By analysing multiple datasets, we gained insights into variations in the generation process, ensuring the generator's stability and generalizability. The

generator's performance was then assessed by computing average metrics across all runs. Averaging these performance metrics mitigates potential biases that could arise from relying on a single synthetic dataset. Additionally, as our generator is conditional, it enables precise control over the number of sequences generated for each class.

Table 4 presents the DTW average scores for nine rehabilitation movements, comparing real sequences (real vs. real) with synthetic sequences (real vs. synthetic) generated by the GAN model. The real vs. real scores (1.8 to 3.5) confirm low inter-sequence variability, indicating high consistency within the real dataset. These values establish a baseline for evaluating synthetic data quality, where real vs. synthetic DTW scores (2.9 to 5.6) reflect the model's ability to replicate movement dynamics.

While most movements exhibit minimal DTW increases, left shoulder abduction (DTW 3.5 vs. 5.6) and right shoulder abduction (DTW 3.4 vs. 5.4) show the largest deviations, suggesting the GAN struggles with complex, high-amplitude movements. Standard deviations indicate greater consistency in real data (0.4 to 0.9) and slightly higher variability in synthetic sequences (0.8 to 1.5), likely due to latent noise in generation. Simpler movements, such as anterior shoulder elevation (DTW 1.8 vs. 2.9), show near-perfect reproduction, confirming the GAN's effectiveness in capturing low-variability movements.

Overall, these results demonstrate the efficiency of our TCGAN, as the generated synthetic sequences exhibit low DTW differences from real data, accurately capturing movement patterns with only minor deviations in high-amplitude exercises.

Movement	**DTW real vs. real**	**Standard deviation**	**DTW real vs. synthetic**	**Standard deviation**
Left elbow flexion	2.0	0.5	3.2	0.9
Right elbow flexion	2.1	0.5	3.3	0.9
Left shoulder flexion	2.0	0.5	3.1	0.8
Right shoulder flexion	1.9	0.4	3.0	0.8
Left shoulder abduction	3.5	0.9	5.6	1.5
Right shoulder abduction	3.4	0.8	5.4	1.4
Anterior shoulder elevation	1.8	0.4	2.9	0.8
Left lateral touch	2.1	0.5	3.4	0.9
Right lateral touch	2.2	0.5	3.5	0.9

Table 4: Comparison of DTW scores and standard deviations for real and synthetic rehabilitation movement sequences

4.2 Qualitative Model Evaluation

To evaluate the quality of sequences generated by our TCGAN, we conducted an experiment with six participants in diverse environments and varying lighting conditions to assess the model's robustness. The participants, differing in size and body type, wore different outfits for each session.

Table 5 presents a physiotherapist's evaluation of 162 rehabilitation exercises across nine movement types, comparing exercises performed with and without our TCGAN-based real-time guidance system. In this approach, patients replicate AI-generated movements, which are seamlessly superimposed onto their live video feed, providing real-time visual guidance for precise execution. The results indicate a significant improvement in physiotherapist approval of executed exercises, rising from 89.11% to 95.08% with TCGAN assistance. The highest gains are observed in complex movements, such as left and right shoulder abduction, where approval rates rise from 87.0% to 93.0% and 87.5% to 93.5%, respectively. This suggests that our system effectively enhances movement accuracy, particularly for exercises with greater spatial complexity and amplitude variations. Even simpler movements, such as anterior shoulder elevation—which initially had a high approval rate (91.0%)—improved further to 97.0%, demonstrating the robustness of our guidance approach. These findings highlight TCGAN's potential as an intelligent corrective assistant, ensuring proper execution of rehabilitation exercises and reinforcing its applicability in real-world physiotherapy

During a rehabilitation exercise, the patient waited an average of only 0.35 seconds for the system to generate a guiding exercise sequence. Each sequence consists of 50 frames, which the patient repeats multiple times. All tests were performed on a computer equipped with an 11th-generation Intel i5 processor and an NVIDIA RTX 3050 graphics card. While this setup is relatively modest for vision-based applications, it is capable of rendering (Superimposing) over 60 frames per second (FPS) on a 1080p video, surpassing the minimum requirement of 25 FPS for our project. This showcases the efficiency of our solution, proving that it does not demand high-performance or costly hardware to achieve optimal results.

Movement	**Number of Exercises**	**Physio. Opinion (Without TCGAN)**	**Physio. Opinion (With TCGAN)**
Left elbow flexion	18	Positive for 89.5%	Positive for 95.5%
Right elbow flexion	18	Positive for 89.0%	Positive for 95.0%
Left shoulder flexion	18	Positive for 90.0%	Positive for 96.0%
Right shoulder flexion	18	Positive for 90.5%	Positive for 96.5%
Left shoulder abduction	18	Positive for 87.0%	Positive for 93.0%
Right shoulder abduction	18	Positive for 87.5%	Positive for 93.5%
Anterior shoulder elevation	18	Positive for 91.0%	Positive for 97.0%
Left lateral touch	18	Positive for 89.0%	Positive for 95.0%
Right lateral touch	18	Positive for 88.5%	Positive for 94.2%
Total/Average	162	Positive for 89.11%	Positive for 95.08%

Table 5: Evaluation of training exercises

To further qualitatively evaluate the performance of our TCGAN, we present two figures illustrating a sequence of images for a right shoulder abduction exercise. This

movement involves raising the right arm laterally while keeping it straight. The first sequence (Figure 4), generated by our TCGAN, visually demonstrates the model's ability to produce realistic movements that accurately follow the intended trajectory and posture. The second sequence (Figure 5), extracted from a real dataset of human subjects performing the same motion, serves as a reference to assess the authenticity and alignment of our generated sequences with actual physical movement. By comparing both sequences, we observe that the TCGAN-generated sequence effectively captures the key aspects of the real movement, confirming the model's ability to generate anatomically accurate rehabilitation exercises.

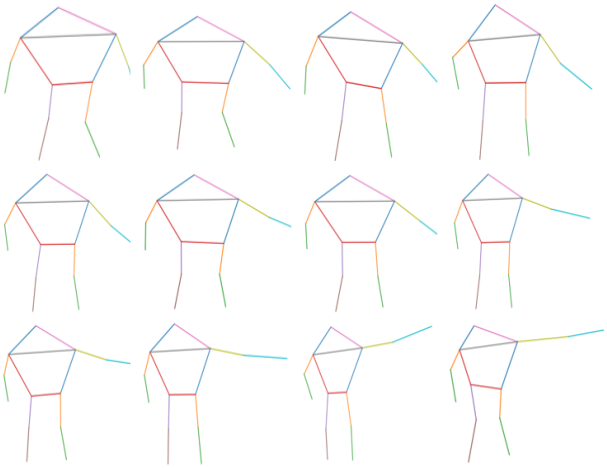

Figure 4: Sequence of shoulder abduction right exercise generated by our TCGAN

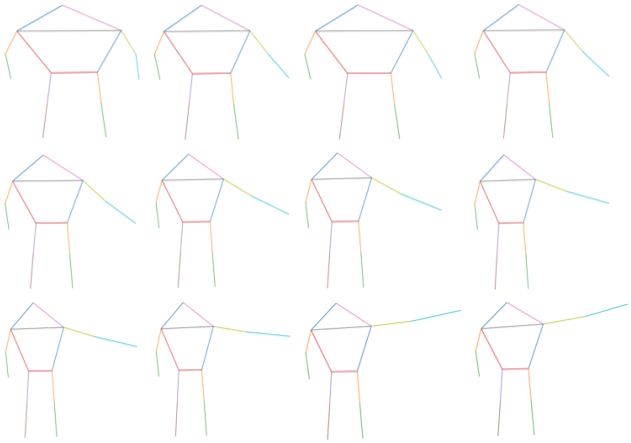

Figure 5: Sequence of shoulder abduction right exercise from the dataset

5 Comparative study

The comparative analysis in Table 6 highlights the key distinctions between our proposed approach and existing rehabilitation solutions. As indicated in the 'Field' column, prior methods are predominantly categorized into vision-based, wearable-based, and game-based systems. Most of these approaches, as shown in the 'Used Technique' column, rely on classification methods such as CNN, SVM, RF, and ANN (Artificial Neuronal Network), providing feedback only after the completion of exercises. While this enables post-exercise performance evaluation, it does not prevent patients from repeatedly performing incorrect movements, which could delay recovery or even exacerbate complications.

In contrast, our proposed method, detailed in the 'Method' column, provides real-time skeletal representation by generating tailored exercises for the patient to mimic, ensuring continuous guidance throughout execution. This real-time capability allows for immediate correction of improper postures and compensatory movements, reducing the risk of long-term impairments. Unlike traditional methods that classify movements as correct or incorrect only after execution, such as those in [Biebl et al., 2021, Guo et al., 2021, Nair and Sakthivel, 2023], our system dynamically adjusts to the patient's posture in real time, ensuring alignment with the prescribed rehabilitation exercise.

Additionally, our approach introduces a personalized and adaptive rehabilitation framework, in contrast to the static classification-based models seen in prior works like [Bijalwan et al., 2023, Miao et al., 2024, Espinoza et al., 2021]. By leveraging LSTM and ANN, our method tailors exercises to each patient's unique characteristics, enhancing both efficiency and effectiveness. Unlike conventional solutions, which do not adapt to individual differences, our model evolves alongside the patient's recovery, optimizing exercise execution.

Another key advantage is the precision of our skeletal representation. While some vision-based models like [Bijalwan et al., 2023, Miao et al., 2024] utilise CNNs and LSTMs for movement classification, they lack interactive correction mechanisms. Our approach overcomes this limitation by offering real-time adaptability and guidance, significantly improving rehabilitation outcomes.

Reference	**Field**	**Used Technique**	**Method**
[Biebl et al., 2021]	Vision	Classification	CNN
[Du et al., 2021]	Vision	Classification	GCN
[Guo et al., 2021]	Vision	Classification	DT, RF, NB, KNN, SVM
[Bijalwan et al., 2023]	Vision	Classification	CNN, LSTM
[Lee et al., 2023]	Vision	Classification	MediaPipe
[Maskeliūnas et al., 2023]	Vision	Classification	RF, CNN
[Miao et al., 2024]	Vision	Classification	Faster R-CNN, HRNe, LSTM
[Bavan et al., 2019]	Wearable	Classification	DT, SVM, KNN, RF
[Burns et al., 2021]	Wearable	Classification	Fully CNN
[Chen et al., 2021]	Wearable	Classification	SVM
[Espinoza et al., 2021]	Wearable	Classification	SVM, ANN, RF

[Nair and Sakthivel, 2023]	Wearable	Classification	DT, SVM, NB
[Lee et al., 2024]	Wearable	Classification	ANN
[Morando et al., 2018]	Game	Classification	Logistic Regression
[Martinez-Martin and Cazorla, 2019]	Game	Classification	SOM, CART
[Macintosh et al., 2020]	Game	Classification	SVM, RF
[Alhwoaimel et al., 2024]	Game	Classification	SVM
Proposed approach	**Vision**	**Creation of tailored exercise to be mimicked by the patient**	**LSTM + ANN**

Table 6: Comparison of the suggested approach with the state-of-the-art (Abbreviations: NB: Naive Bayesian, HRNe: Hierarchical Recurrent Neural Encoder, SOM: Self-Organized Map).

6 Conclusion

In this study, we proposed a TCGAN-based motion generation system for physical rehabilitation exercises. Unlike conventional classification-based methods that provide only post-exercise evaluations, our approach introduces real-time skeletal motion generation, ensuring continuous guidance and correction throughout rehabilitation sessions. By dynamically generating motion sequences tailored to each patient's characteristics—such as height, posture, and positioning—our system enables precise and effective rehabilitation while minimizing incorrect movements that could hinder recovery.

Experimental results confirm the high accuracy and realism of the generated skeletal sequences, achieving an FID score of 0.89 and strong similarity between real and generated temporal sequences, as evidenced by DTW scores ranging from 2.9 to 5.6 across different movement classes. This demonstrates a robust alignment between synthetic and real motion data. Furthermore, our comparative analysis with state-of-the-art rehabilitation techniques highlights the superiority of our approach in real-time correction and personalized movement adaptation, leading to more effective rehabilitation outcomes. Evaluations using machine learning classifiers further validate the quality of the generated exercises, showing a close resemblance to real human movements. The reduction in incorrect movements observed during patient trials underscores the practical benefits of our system in ensuring proper rehabilitation execution.

Future work will focus on expanding the system to support a wider range of rehabilitation exercises, improving adaptability to diverse patient profiles, enhance fidelity for complex exercises, and integrating additional patient data, such as EMG signals, to further enhance motion accuracy. Extensive clinical validation through large-scale patient trials will be crucial to assessing the system's long-term effectiveness, usability, and safety in real-world rehabilitation settings. Additionally, refining real-time feedback mechanisms and introducing adaptive difficulty levels will

ensure that rehabilitation exercises evolve with the patient's progress, making the system more responsive to individual needs.

References

[Alhwoaimel et al., 2024] Alhwoaimel, N. A., Hughes, A.-M., Warner, M., Alenazi, A. M., Alshehri, M. M., Alqahtani, B. A., Alhowimel, A. S., Wagland, R., Brown, S., Turk, R.: Video Game-Based Trunk Exercises for Rehabilitation in Chronic Stroke Survivors: A Mixed-Methods Feasibility Study, Sensors, 24, 21, (2024), 6830. DOI : https://doi.org/10.3390/s24216830

[Bavan et al., 2019] Bavan, L., Surmacz, K., Beard, D., Mellon, S., Rees, J.: Adherence monitoring of rehabilitation exercise with inertial sensors: A clinical validation study, Gait & Posture, 70, (2019), 211-217. DOI: 10.1016/j.gaitpost.2019.03.008

[Berndt and Clifford, 1994] Berndt, D. J., Clifford, J. Using dynamic time warping to find patterns in time series. In Proceedings of the 3rd international conference on knowledge discovery and data mining, 1994, 359-370.

[Bicer et al., 2022] Bicer, M., Phillips, A. T. M., Melis, A., McGregor, A. H., Modenese, L.: Generative deep learning applied to biomechanics: A new augmentation technique for motion capture datasets, Journal of Biomechanics, 144, (2022), 111301. DOI: https://doi.org/10.1016/j.jbiomech.2022.111301

[Biebl et al., 2021] Biebl, J. T., Rykala, M., Strobel, M., Kaur Bollinger, P., Ulm, B., Kraft, E., Huber, S., Lorenz, A.: App-based feedback for rehabilitation exercise correction in patients with knee or hip osteoarthritis: prospective cohort study, Journal of Medical Internet Research, 23, 7, (2021), e26658. DOI: 10.2196/26658

[Bijalwan et al., 2023] Bijalwan, V., Semwal, V. B., Singh, G., Mandal, T. K.: HDL-PSR: Modelling spatio-temporal features using hybrid deep learning approach for post-stroke rehabilitation, Neural Processing Letters, 55, 1, (2023), 279-298. DOI: https://doi.org/10.1007/s11063-022-10744-6

[Boukhennoufa et al., 2023] Boukhennoufa, I., Jarchi, D., Zhai, X., Utti, V., Sanei, S., Lee, T. K. M., Jackson, J., McDonald-Maier, K. D.: A novel model to generate heterogeneous and realistic time-series data for post-stroke rehabilitation assessment, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31, (2023), 2676-2687. DOI: 10.1109/TNSRE.2023.3283045

[Bowman et al., 2021] Bowman, T., Gervasoni, E., Arienti, C., Lazzarini, S. G., Negrini, S., Crea, S., Cattaneo, D., Carrozza, M. C.: Wearable devices for biofeedback rehabilitation: a systematic review and meta-analysis to design application rules and estimate the effectiveness on balance and gait outcomes in neurological diseases, Sensors, 21, 10, (2021), 3444. DOI: 10.3390/s21103444

[Burns et al., 2021] Burns, D., Boyer, P., Razmjou, H., Richards, R., Whyne, C.: Adherence patterns and dose response of physiotherapy for rotator cuff pathology: longitudinal cohort study, JMIR Rehabilitation and Assistive Technologies, 8, 1, (2021), e21374. DOI: 10.2196/21374

[Castillo et al., 2024] Castillo, J. F. V., Vega, M. F. M., Cardona, J. E. M., Lopez, D., Quiñones, L., Gallo, O. A. H., & Lopez, J. F. Design of virtual reality exergames for upper limb stroke rehabilitation following iterative design methods: usability study. JMIR Serious Games, 12, 1, (2024), e48900. DOI: doi:10.2196/48900

[Chen et al., 2021] Chen, J., Caviedes, J., and Li, B.: Classification of Single-Axis Spinal Motion Using a Wearable System of Stretch Sensors for At-Home Physical Therapy; 43rd Annual

International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2021, 7404-7407. DOI: 10.1109/EMBC46164.2021.9630663

[Dowson and Landau, 1982] Dowson, D. and Landau, B.: The fréchet distance between multivariate normal distributions, Journal of Multivariate Analysis, 12, (1982), 450-55. DOI : https://doi.org/10.1016/0047-259X(82)90077-X

[Du et al., 2021] Du, C., Graham, S., Depp, C., and Nguyen, T.: Assessing Physical Rehabilitation Exercises Using Graph Convolutional Network with Self-Supervised Regularization; 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2021, 281-285. DOI: 10.1109/EMBC46164.2021.9629569

[Espinoza et al., 2021] Espinoza Bernal, V. C., Hiremath, S. V., Wolf, B., Riley, B., Mendonca, R. J., Johnson, M. J.: Classifying and tracking rehabilitation interventions through machine-learning algorithms in individuals with stroke, Journal of Rehabilitation and Assistive Technologies Engineering, 8, (2021), 20556683211044640. DOI: doi: 10.1177/20556683211044640

[Giggins et al., 2013] Giggins, O. M., Persson, U. M., Caulfield, B.: Biofeedback in rehabilitation, Journal of Neuroengineering and Rehabilitation, 10, (2013), 1-11. DOI: 10.1186/1743-0003-10-60

[Guo et al., 2021] Guo, L., Zhang, B., Yang, Z., Wang, X., and Wang, J.: Motor Function Evaluation System Based on Machine Vision for COPD Patients; IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT), 2021, 1136-1139. DOI: 10.1109/ICCASIT53235.2021.9633685

[Hadkey et al., 2024] Hadley, A. J., Pulliam, C. L.: Enhancing activity recognition after stroke: Generative adversarial networks for kinematic data augmentation, Sensors, 24, 21, (2024), 6861. DOI : https://doi.org/10.3390/s24216861

[Hadley and Pulliam, 2024] Hadley, A. J., Pulliam, C.: Simulating Accelerometer Signals of Parkinson's Gait Using Generative Adversarial Networks, TechRxiv, 30, (2024). DOI: 10.36227/techrxiv.171707302.26429714/v1

[He et al., 2024] He, T., Chen, Y., Wang, L., Cheng, H.: An Expert-Knowledge-Based Graph Convolutional Network for Skeleton-Based Physical Rehabilitation Exercises Assessment, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 32, (2024), 1916-1925. DOI: https://doi.org/10.1109/TNSRE.2024.3400790

[Jack et al., 2010] Jack, K., McLean, S. M., Moffett, J. K., Gardiner, E.: Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review, Manual Therapy, 15, 3, (2010), 220-228. DOI: 10.1016/j.math.2009.12.004

[Kárason et al., 2023] Kárason, H., Ritrovato, P., Maffulli, N., Tortorella, F.: Generative Data Augmentation of Human Biomechanics; International Conference on Image Analysis and Processing, Udine (2023), 482-493.

[Kramer et al., 2019] Kramer, S. F., Hung, S. H., Brodtmann, A.: The impact of physical activity before and after stroke on stroke risk and recovery: a narrative review, Current Neurology and Neuroscience Reports, 19, (2019), 1-9. DOI: 10.1007/s11910-019-0949-4

[Kuncan et al., 2019] Kuncan, F., Kaya, Y., & Kuncan, M.: A Novel Approach for Activity Recognition with Down-Sampling 1D Local Binary Pattern, Advances in Electrical and Computer Engineering, 19, 1, (2019), 35-44. DOI: https://doi.org/10.4316/AECE.2019.01005

[Kuncan et al., 2022a] Kuncan, F., Kaya, Y., Yiner, Z., & Kaya, M.: A New Approach for Physical Human Activity Recognition from Sensor Signals Based on Motif Patterns and Long-

Short Term Memory, Biomedical Signal Processing and Control, 78, (2022), 103963. DOI: https://doi.org/10.1016/j.bspc.2022.103963

[Kuncan et al., 2022b] Kuncan, F., Kaya, Y., Tekin, R., & Kuncan, M.: A New Approach for Physical Human Activity Recognition Based on Co-Occurrence Matrices, The Journal of Supercomputing, 78, 1, (2022), 1048-1070, DOI: https://doi.org/10.1007/s11227-021-03921-2

[Lee et al., 2023] Lee, H. J., Jin, S. M., Kim, S. J., Kim, J. H., Kim, H., Bae, E. K., Yoo, S. K., Kim, J. H.: Development and Validation of an Artificial Intelligence-Based Motion Analysis System for Upper Extremity Rehabilitation Exercises in Patients with Spinal Cord Injury: A Randomized Controlled Trial, Healthcare, 12, 1, (2023), 7. DOI : https://doi.org/10.3390/healthcare12010007

[Lee et al., 2024] Lee, K., Kim, J.-H., Hong, H., Jeong, Y., Ryu, H., Kim, H., Lee, S.-U.: Deep learning model for classifying shoulder pain rehabilitation exercises using IMU sensor, Journal of NeuroEngineering and Rehabilitation, 21, 1, (2024), 42. DOI : https://doi.org/10.1186/s12984-024-01343-8

[Macintosh et al., 2020] Macintosh, A., Vignais, N., Desailly, E., Biddiss, E., Vigneron, V.: A classification and calibration procedure for gesture specific home-based therapy exercise in young people with cerebral palsy, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29, (2020), 144-155. DOI: 10.1109/TNSRE.2020.3038370

[Martinez-Martin and Cazorla, 2019] Martinez-Martin, E., Cazorla, M.: A socially assistive robot for elderly exercise promotion, IEEE Access, 7, (2019), 75515-75529. DOI: 10.1109/ACCESS.2019.2921257

[Maskeliūnas et al., 2023] Maskeliūnas, R., Damaševičius, R., Blažauskas, T., Canbulut, C., Adomavičienė, A., Griškevičius, J.: BiomacVR: A virtual reality-based system for precise human posture and motion analysis in rehabilitation exercises using depth sensors, Electronics, 12, 2, (2023), 339. DOI : https://doi.org/10.3390/electronics12020339

[Mennella and al., 2023] Mennella, C., Maniscalco, U., De Pietro, G., Esposito, M.: Generating a novel synthetic dataset for rehabilitation exercises using pose-guided conditioned diffusion models: A quantitative and qualitative evaluation, Computers in Biology and Medicine, 167, (2023), 107665. DOI: 10.1016/j.compbiomed.2023.107665

[Miao et al., 2024] Miao, S., Liu, Z., Wang, D., Shen, X., Shen, N.: Applying Hybrid Deep Learning Models to Assess Upper Limb Rehabilitation, IEEE Access, 12, (2024), 154337 – 154348. DOI: 10.1109/ACCESS.2024.3482115. DOI: 10.1109/ACCESS.2024.3482115

[Miller et al., 2017] Miller, K. K., Porter, R. E., DeBaun-Sprague, E., Van Puymbroeck, M., Schmid, A. A.: Exercise after stroke: patient adherence and beliefs after discharge from rehabilitation, Topics in Stroke Rehabilitation, 24, 2, (2017), 142-148. DOI: 10.1080/10749357.2016.1200292

[Morando et al., 2018] Morando, M., Ponte, S., Ferrara, E., Dellepiane, S.: Definition of motion and biophysical indicators for home-based rehabilitation through serious games, Information, 9, 5, (2018), 105. DOI : https://doi.org/10.3390/info9050105

[Nair and Sakthivel, 2023] Nair, B. B., Sakthivel, N. R.: A deep learning-based upper limb rehabilitation exercise status identification system, Arabian Journal for Science and Engineering, 48, 2, (2023), 1237-1271.

[Peppes et al., 2023] Peppes, N., Tsakanikas, P., Daskalakis, E., Alexakis, T., Adamopoulou, E., Demestichas, K.: FoGGAN: Generating realistic Parkinson's disease freezing of gait data using GANs; Sensors, 23, 19, (2023), 8158. DOI : https://doi.org/10.3390/s23198158

[Qiu et al., 2022] Qiu, S., Zhao, H., Jiang, N., Wang, Z., Liu, L., An, Y. Zhao, H., Miao, X., Liu, R., Fortino, G.: Multi-sensor Information Fusion Based on Machine Learning for Real Applications in Human Activity Recognition: State-of-the-Art and Research Challenges, Information Fusion, 80, (2022), 241-265. DOI: https://doi.org/10.1016/j.inffus.2021.11.006

[Saito et al., 2017] Saito, M., Matsumoto, E., Saito, S.: Temporal generative adversarial nets with singular value clipping, Proceedings of the IEEE International Conference on Computer Vision, Cambridge (2017), 2830-2839. DOI : https://doi.org/10.48550/arXiv.1611.06624

[Solanki et al., 2021] Solanki, A. Nayyar, A. Naved, M.: Generative adversarial networks for image-to-image translation; Elsevier, (2021). DOI : https://doi.org/10.1016/C2020-0-00284-7

[Soni et al., 2023a] Soni, V., Jaiswal, S., Semwal, V. B., Roy, B., Choubey, D. K., & Mallick, D. K.: An Enhanced Deep Learning Approach for Smartphone-Based Human Activity Recognition in IoHT, In Machine Learning, Image Processing, Network Security and Data Sciences: Select Proceedings of 3rd International Conference on MIND 2021, Singapore, Springer Nature Singapore, 2023, 505-516.DOI: https://doi.org/10.1007/978-981-19-5868-7_37

[Soni et al., 2023b] Soni, V., Yadav, H., Semwal, V. B., Roy, B., Choubey, D. K., & Mallick, D. K.: A Novel Smartphone-Based Human Activity Recognition Using Deep Learning in Health Care, In Machine Learning, Image Processing, Network Security and Data Sciences: Select Proceedings of 3rd International Conference on MIND 2021, Singapore, Springer Nature Singapore, 2023, 493-503. DOI: https://doi.org/10.1007/978-981-19-5868-7_36

[Tamburella et al., 2019] Tamburella, F., Moreno, J. C., Herrera Valenzuela, D. S., Pisotta, I., Iosa, M., Cincotti, F., Mattia, D., Pons, J. L., Molinari, M.: Influences of the biofeedback content on robotic post-stroke gait rehabilitation: electromyographic vs joint torque biofeedback, Journal of Neuroengineering and Rehabilitation, 16, (2019), 1-17. DOI : https://doi.org/10.1186/

[Yu et al., 2024] Yu, B. X. B., Liu, Y., Chan, K. C. C., Chen, C. W.: EGCN++: A New Fusion Strategy for Ensemble Learning in Skeleton-Based Rehabilitation Exercise Assessment, IEEE Transactions on Pattern Analysis and Machine Intelligence, 46, 9, (2024), 6471-6485. DOI: https://doi.org/10.1109/TPAMI.2024.3378753