ITC 3/54
Information Technology and Control
Vol. 54 / No. 3/ 2025
pp. 864-884
DOI 10.5755/j01.itc.54.3.40634

A TCGAN-Based Real-Time Personalized Motion Guidance System to Reduce Compensatory Movements in Post-Stroke Rehabilitation

Received 2025/02/24 | Accepted after revision 2025/07/15

HOW TO CITE: Lezzar, F., Benmerzoug, D., Berkane, M. L., Boudouda, S., Eddine, M. S. (2025). A TCGAN-Based Real-Time Personalized Motion Guidance System to Reduce Compensatory Movements in Post-Stroke Rehabilitation. *Information Technology and Control*, 54(3), 684-884. https://doi.org/10.5755/j01.itc.54.3.40634

A TCGAN-Based Real-Time Personalized Motion Guidance System to Reduce Compensatory Movements in Post-Stroke Rehabilitation

Fouzi Lezzar*, Djamel Benmerzoug, Mohamed Lamine Berkane, Souheila Boudouda
LIRE Laboratory, Faculty of New Technologies of Information and Communication, University of Abdelhamid Mehri Constantine 2, Constantine 25016, Algeria

Mili Seif Eddine
Ecole Normale Supérieure, Constantine, Engineering Laboratory for Complex Systems (LISCO), Annaba University, Algeria

Corresponding author: fouzi.lezzar@univ-constantine2.dz

Stroke rehabilitation is essential for motor function recovery, yet traditional methods require therapist supervision, which can be costly and inaccessible. Home-based rehabilitation offers an alternative, but without real-time guidance, patients may develop compensatory movements, hindering progress. Existing approaches provide feedback only after exercises are completed, limiting their effectiveness. To address this, we propose a Temporal Conditional Generative Adversarial Network (TCGAN)-based motion generation system that provides 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, reducing errors and improving rehabilitation outcomes. Both qualitative and quantitative evaluations confirm the effectiveness of the generated exercises, benefiting from the proposed architecture, improved loss function, optimized training process, and TCGAN hyperparameter tuning. Experimental re-

sults show a high degree of similarity between generated and real movements, with a Fréchet Inception Distance (FID) score of 0.87, demonstrating the system's realism and reliability. This approach enhances patient autonomy and recovery efficiency, offering a more interactive and adaptive rehabilitation experience.

KEYWORDS: Home-based rehabilitation, Compensation assessment, Real-time exercise guidance, TCGAN, Post-stroke recovery.

1. Introduction

Stroke has become the second leading cause of death and a major cause of acquired disability. Up to 80% of stroke survivors develop limb impairments that significantly reduce their ability to perform daily activities [8] and diminish their overall quality of life [2]. Many survivors rely on compensatory movements, such as excessive thoracic rotation/scapular rotation or hip hiking during the swing phase of walking. Since these adaptations may be beneficial initially, they can increase the risk of poor recovery in the long term [7]. Studies show that post-stroke physical problems can be successfully addressed in the majority of patients by reducing these compensations through the proper use of the dysfunctional limb [10]. In high-impact activities, the affected side should be directed toward the right target to enhance performance and improve accuracy [31]. Repetitive practice of a large number of correct movements has been shown to be effective for home rehabilitation, as prescribed by therapists [24, 19]. However, maintaining motivation and correct form without professional guidance remains a challenge, as low adherence and poor technique can negatively impact recovery [27]. Nowadays, sensor-based [11], camera-based [26] and virtual reality-based [18] systems used to detect compensatory movements primarily rely on classification algorithms and analyse movement patterns after exercises have been completed. These methods such as [9] cause patients to perform more incorrect repetitions before achieving the correct injury. Patients often lack immediate real-time feedback, highlighting the urgent need for mechanisms that dynamically monitor and correct movements during exercise to ensure a safe and effective rehabilitation process. To overcome these drawbacks, we have developed a deep learning architecture based on Temporal Conditional Generative Adversarial Networks (TCGAN) that generates skeleton-based rehabilitation data to help patients perform exercises correctly. The proposed approach differs from current methodologies such as [4, 1], which use classification techniques to indicate whether the executed movement is correct or incorrect. Our system, however, achieves superior performance by dynamically superimposing the generated skeleton onto the patient's video, acting as a visual reference for the movements they should follow. By allowing patients to visually observe and replicate proper rehabilitation exercises, they can better understand the correct technique without form of the exercise, which can delay recovery and even aggravate their relying exclusively on verbal or binary feedback. Continuous visual guidance minimizes errors in movement execution, thereby reducing the probability of practicing faulty techniques, such as compensatory movements. This system will therefore not only enhance the rehabilitation process by making it more interactive and engaging but also promote safer, more functional, and efficient recovery, as patients will benefit from real-time visual guidance. Additionally, the TCGAN-based model can be deployed without cumbersome sensors or specialized equipment, making it more suitable for home use. In this work, we aim to develop an economical and practical solution that empowers recovering patients by enabling them to perform safe and effective exercises independently at home.

The rest of this paper is structured as follows. Section 2 discusses related work in the field, providing an overview of existing rehabilitation approaches and highlighting their limitations. Section 3 presents our proposed approach, detailing the dataset preparation, model architecture, and training process. Section 4 presents the experimental results, providing both qualitative and quantitative evaluations of the generated skeletal sequences. Sections 5 and 6 provide respectively an ablation study and a comparative study, comparing our method with existing solutions in the literature. Finally, Section 7 concludes the paper and discusses potential future research directions.

2. Related Work

Automated exercise assessment using Artificial Intelligence (AI) can be regarded essentially in the literature as a classification task, categorizing a movement into correct or incorrect. Rehabilitation exercises can be classified based on the degree of precision with which they are executed. As reported in the literature, the most frequently used methodologies rely on feature engineering. For instance, in [5] authors proposed a Graph Convolutional Network (GCN) for assessing physical rehabilitation exercises. This model represented the skeleton data of a human as a graph, and took its movement as input to predict the quality of the performed exercise compared to the prescribed version. In their work, Lee et al. [15] tested several hybrid models, including Neural Networks (NNs) and Support Vector Machines (SVM), and determined that NNs achieved the highest effectiveness. Study in [21] aimed to investigate the potential of predicting a treatment's outcome using a deep learning prognosis model developed for another treatment. The data used were gathered from different sources: clinical measurement, biomechanical measurement, and electroencephalography (EEG) measurement. In another study [14], authors discuss the challenge of developing self-rehabilitation systems while formulating an accurate video-based assessment of motor skills. They present a deep learning model for automating motion analysis and create a mobile application based on this model. The proposed method can estimate the upper limb function of stroke survivors using only video data without any other sensors.

Another sub-area of research focuses on applying deep learning to recognize compensatory movements during rehabilitation. Authors in [30] analysed how technology-based methods have been applied to assess and detect compensation during stroke upper extremity rehabilitation. Various Machine Learning (ML) algorithms were applied to train the classification model for compensation recognition. In [22], authors presented a virtual rehabilitation system (VRS) that can detect compensatory movements and improve the outcome of upper extremity rehabilitation in community-dwelling older adults with stroke. Kaku et al. [12] also faced difficulties with their Fully Connected Neuronal Network models and obtained an accuracy rate of only 70%. On the contrary, [23] has shown excellent outcomes in the actual settings. On the other hand, [3] used SVM classifier to enable real-time monitoring of compensatory movements during activities. The system proved helpful in correcting movements by delivering the required force during the stroke exercise. Kashi et al. [13] built a machine-learning-based automated model that gives patients accurate information on the compensatory movements that they perform. They used the random-forest (RF) algorithm for training this classification model on a local dataset.

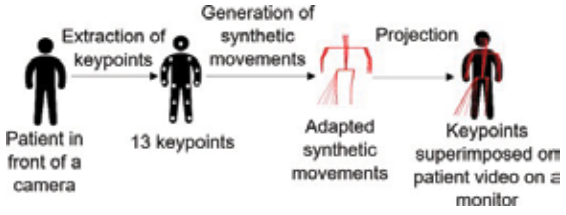
What we have noticed is that all existing methods rely on post-exercise feedback mechanisms, meaning errors are detected only after movement execution. This limitation can delay the correction process and reinforce incorrect motor patterns, ultimately affecting rehabilitation outcomes. Our proposed approach addresses this limitation by providing real-time guidance, allowing immediate corrections during exercise execution. A more detailed comparative study with the relevant related work is provided in Section 6, highlighting the key distinctions and advantages of our method.

3. Proposed Solution

Our approach (Figure 1) represents rehabilitation exercises as dynamic scenes using a skeletal model and employs a TCGAN to generate them. A camera positioned in front of the patient captures keypoints (landmarks) from the body, which are then extracted. The TCGAN generates real-time exercise movements tailored to the patient's size and position. This skeletal model is overlaid onto the live video feed of the patient, allowing him to mimic the movements accurately without compensations or errors. This interactive method supports the patient throughout his recovery process. The system operates as follows:

1. Keypoints extraction (Using MoveNet: https://www.tensorflow.org/hub/tutorials/m ovenet),
2. Generation of synthetic movements using TCGAN,
3. Overlay of the skeletal model onto the real-time patient video, displayed on a monitor.

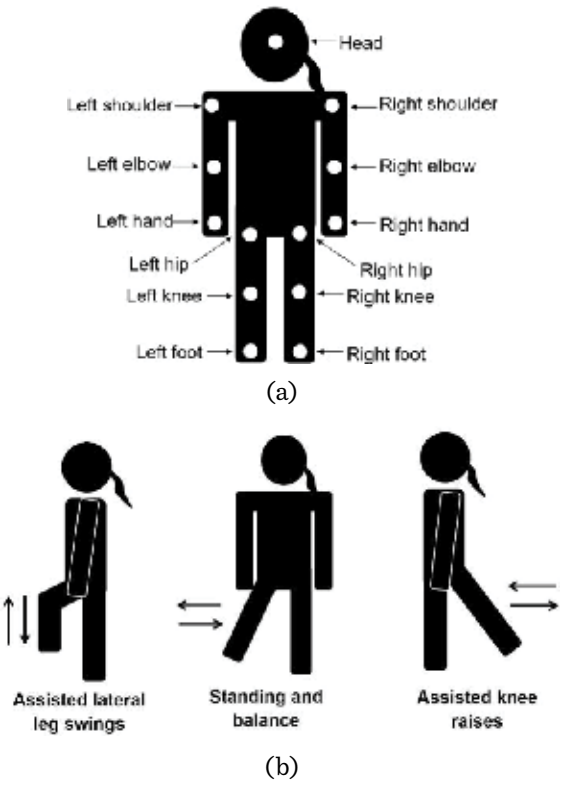
Figure 1
The proposed approach.



3.1. Dataset and Preprocessing

To test our method, we created a dataset with the assistance of a physiotherapist. The dataset comprises skeletal data from six healthy individuals of varying heights and ages, ensuring diversity in the profiles performing different rehabilitation movements. An RGB camera was used to capture the data, and keypoints were extracted using MoveNet (Figure 2(a)). All participants performed multiple repetitions of exercises, including standing and balance, assisted lateral leg swings, and assisted knee raises, as illustrated in Figure 2(b). None of the participants had experienced a stroke, as the goal of this project is not to classify movements based on stroke-related impairments. Instead, the objective is to generate accurately executed rehabilitation movements to prevent compensations and errors during exercise performance.

Figure 2
Examples of the three movements.



After cleaning and removing unusable data, the final dataset consists of 417 gesture sequence samples, each containing 50 frames. In our rehabilitation study, we chose not to include keypoints for the eyes and ears. This decision was made based on the fact that eye and ear movements are not highly relevant to the rehabilitation process. By focusing on other joints and body parts, we can accurately assess the movements and postures essential for the physical restoration of patients. Consequently, for each sample, we utilized 13 of the 17 keypoints proposed by MoveNet (Figure 2(a)). Each keypoint is represented by two coordinates, x and y. To enhance the robustness and stability of the TCGAN, we applied a normalization process to the data. The keypoint coordinates were divided by the image dimensions to obtain normalized values within the range [0, 1].

3.2. Keypoint Extraction

After employing the MoveNet model for dataset preparation, we use it in the first step of our process to generate personalized rehabilitation exercises. MoveNet, an advanced pose detection model, accurately identifies keypoints from a person's image, such as the shoulders, elbows, hips, and knees, even under challenging conditions. We utilize 13 keypoints to represent the body, which then serve as inputs for our TCGAN. This generates new keypoints that depict movements tailored to the rehabilitation exercises.

3.3. TCGAN Architecture

Our TCGAN (Figure 3) leverages two neural networks and an adversarial training process. In a typical Conditional GAN [25], a generator G and a discriminator D are trained simultaneously in a min-max game defined as follows:

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

The generator G (composed of G_0+G_1) produces synthetic data from a latent vector z_0 and a conditional label y, generating exercise scenes customized to the patient's position and size. The discriminator D assesses the likelihood that the generated data x is real given the condition y. The first term, $E_{x\sim pdata}$[log-D(x|y)], maximizes the probability that D correctly identifies real data. The second term, $E_{z\sim pz}$ [log (1−D (G(z|y)|y)], minimizes probability that D mistakes generated data for real. G and D's minimax game aims to generate realistic data.

3.3.1. Temporal Generator

The temporal generator G_0 in this study has a crucial role in generating bodily keypoints that represent sequences of movements. G_0 produces a set of latent variables z_1^t (for t=1,..., T) from an input latent vector z_0 where T is the number of frames of the exercise sequence. The inputs to G_0 consist of three components:

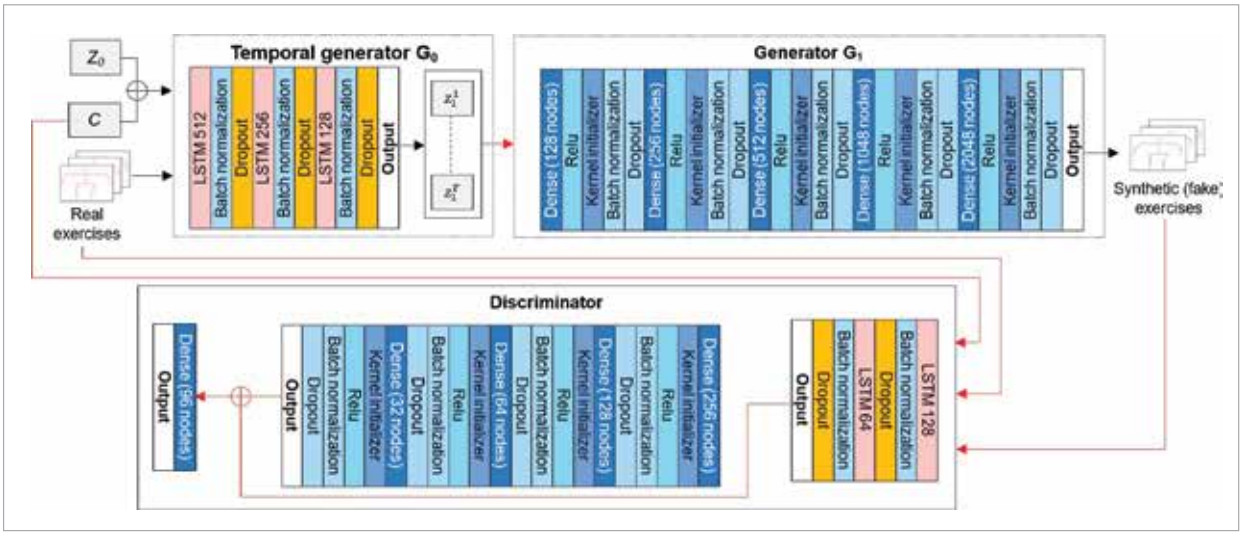
- Latent vector z_0 with shape 100,
- Sequence of frames shaped (T, 13 × 2), where each frame consists of 13 keypoints with (x, y) coordinates,
- Condition vector with a shape of seven (condition_dim = 7, exercise number, coordinates of head and feet to determine the size and position of the patient on the video).

LSTM (Long Short Term Memory) layers are used to store temporal dependencies in skeletal data and are one of the most effective tools because they learn long-range dependencies in sequential data. This can serve as a basis for accurate modelling of human movements over time. Given a skeletal keypoints sequence $X = (x_1, x_2, \ldots, x_t)$, where $x_t \in R^{13}$ implies 13 keypoints at time step t, this means the input sequence extends over T frames, each consisting of 13 skeletal keypoints for the human body joints at every specific 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 (rate 0.3) to ensure regularization and training stability. The output of the LSTM layers is a 2D array with a shape of (T, 128), representing hidden states for each time step. These latent variables are further fed into the image generator G_1. This yields a complete sequence of generated skeletal keypoints over the T frames. This is important because it ensures the temporal coherency of the generated keypoints for an accurate representation of the movements of interest in the given rehabilitation exercise. The final result of G is a smooth keypoint-based frame sequence of shape (T, 13 × 2) to help patients perform correct rehabilitation movements.

3.3.2. Generator

The generator G_1 is responsible for generating realistic sequences of motions represented by skeletal

Figure 3
Illustration of the proposed TCGAN.



keypoints for use in physical rehabilitation tasks. This generator's architecture is designed to produce context-specific and coherently changing sequences of motion that resemble to realistic human movements. The generator creates tailored rehabilitation exercises through mapping hidden variables to structured layers. It receives inputs from the temporal generator G_0, in the form of a series of latent variables ($z_i^t(t=1,\ldots, T)$). Furthermore, these variables have the shape of (T, 128). The generator consists of a series of fully connected layers. Dense layers learn complex representations of the input data through weighted linear combinations, followed by a non-linear activation function (Rectified Linear Unit (ReLU)). We used ReLU because of its simplicity, its ability to minimize the vanishing gradient effect, and its efficiency in deep networks. The G_1 architecture contains several dense layers with sizes of 128, 256, 512, 1024 and 2048 units. Each layer contributes to the numerous transformations of features from previous layers to produce the required output representation of skeletal keypoints. The generator uses these dense layers to capture the relationships between patterns, ensuring that the final output corresponds to the pattern that moves in a realistic manner. To optimize G_1's performance, we added batch normalisation and dropout (rate 0.3) between dense layers to improve training stability and enhance robustness, ensuring realistic and coherent motion sequences for rehabilitation.

The generator also includes Kernel Initializers to predefine the weight initialization in the dense layers. We used the "he" initializer, which is particularly suited for layers that use ReLU activations. The output from the last dense layer is reshaped into a series of coordinates representing the skeletal keypoints.

In summary, the TCGAN has been designed to generate realistic sequences of skeletal movements. An LSTM architecture ensures a temporal connection between frames, while dense layers capture relationships within the conditioned data. We also utilize other techniques, such as Batch Normalization and Kernel Initialization, to enhance training stability and efficiency.

3.3.3. Discriminator

The discriminator of the TCGAN incorporates both spatial and temporal pathways to assess the authenticity of generated skeletal sequences conditioned on specific movement attributes. Both fake and real data are combined with the condition vector and fed into the discriminator. The spatial discriminator consists of dense layers. The first layer has 256 neurons, followed by 128, 64, and finally 32 neurons. Every dense layer is equipped with the ReLU activation function which introduces non-linearity. Thus, the model can learn complex relationships between keypoints in each frame. To improve the training efficiency of these layers, we apply a kernel initializer to optimize the initial weight distribution. Additionally, dropout (rate 0.3) and batch normalisation layers are included to enhance robustness.

The temporal discriminator includes two LSTM layers. The first layer has 128 units and the second has 64 units. These process the temporal aspects of the skeletal sequences. To stabilize the training by normalizing of the layer outputs, every LSTM layer is followed by batch normalization. Additionally, a dropout and batch normalization techniques are added, which helps prevent overfitting by randomly desactivating some neurons during training. The output that gets fused from both spatial and temporal pathways is passed through fully connected layer (96 neurons = 64 + 32) with a sigmoid activation function to provide a probability score indicating whether the sequence is real or generated. Thanks to this combined architecture, the TCGAN can simultaneously evaluate the spatial accuracy and temporal consistency of the movements.

3.3.4. Improved Loss Function

We enhance our TCGAN by introducing a new loss function with several terms to ensure realistic and temporally consistent skeletal sequences. To enhance the discrimination between real and fake samples, we first add a Kullback-Leibler (KL) divergence term to the loss function of the discriminator. This brings the real and the generated data distributions closer together while enhancing separability between them.

$$L_D = E_{x\sim pdata}[\log D(x,y)] + E_{z\sim pz}[\log(1-D(G(z,y),y))] + \beta \cdot KL(p_{data} \| p_{fake}). \quad (2)$$

In Equation (2), β is a weighting factor that is responsible for controlling the strength of the KL diver-

gence. The symbols p_{data} and p_{fake} in the formula refer to the real data distribution (train dataset) and generated data distribution (GAN output) respectively. This term enables it easier for the discriminator to distinguish between true and false data, thereby accelerating the learning process.

We propose that the generator incorporates a penalty term inspired by the Wasserstein distance with gradient penalty. The generator aims to create a sequence that tricks the discriminator while simultaneously minimizing the style difference between these sequences and real ones. The L2 difference is calculated between the generator's output and the style target.

$$L_G = E_{z\sim pz}\,[-D(G(z,y),y)] + \gamma.\|G(z,y)-\hat{G}(z,y)\|^2. \quad (3)$$

L_G in Equation 3 represents the Generator Loss. The constant γ controls the influence of the style difference term. By incorporating this term, the sampling process generates samples that appear more realistic while preserving the stylistic properties of the real data.

The TCGAN's overall loss function includes terms corresponding to smoothness regularization and a gradient penalty, which together define the global loss function.

$$L_{total} = L_D + L_G + \lambda_{smooth} \cdot L_{smooth} + \lambda_{GP} \cdot L_{GP} + \alpha \cdot KL\,(p_{data} \parallel p_{fake}), \quad (4)$$

$$L_{smooth} = \Sigma^{T-1}\|x_{t+1} - x_t\|^2, \quad (5)$$

$$L_{GP} = \lambda\,.\,E_{\hat{x}\sim P\hat{x}}\,[(\|\nabla\hat{x}\,D(\hat{x}\,,y)\|_2 - 1)^2]. \quad (6)$$

The hyperparameters that control the smoothness and the gradient penalties are λ_{smooth} and λ_{GP}, respectively, while α is the weighting factor for K.L. divergence. This loss function enables the training of a generator that produces coherent skeletal sequence aligned with the real data distribution while stabilizing training through regularization and penalty terms. Equation (6) represents the Gradient Penalty used in Wasserstein GANs with regularization. It enforces a Lipschitz constraint on the discriminator D by ensuring that the norm of its gradient remains close to 1 for interpolated samples between real and generated data. This helps stabilize training and addresses the weight clipping issue found in standard WGANs. The term λ controls the strength of the penalty, and the expectation is taken over these interpolated samples.

3.4. Model Training

TCGAN Training Algorithm

1. Initialize models and hyperparameters:
-Load the real data.
-Initialize the generator (G) and discriminator (D) models.
-Define hyperparameters: learning rate, batch size, λ_{smooth}, λ_{GP}, epochs, etc.
2. Training Loop:
For each epoch:
Step 1: Train the Discriminator
-Sample a batch of real skeletal sequences from the real data.
-Generate a batch of fake sequences using the generator.
-Calculate loss function L_D
if (LD < thresholdD) **then**
Train Discriminator
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: $W'=U\Sigma'V^T$
    Update discriminator weights:
      $W \leftarrow W-\eta\nabla L_D$
  End if
Step 2: Train the Generator
-Sample random noise z and action label y.
-Generate fake skeletal sequences using the generator.
-Train the generator, minimizing the generator loss L_G.
Step 3: Compute Total Loss L_{Total}
Step 4: Update Models
  Update the weights of the D and G based on their respective losses.

The training of our TCGAN (as described in the algorithm above) begins by setting key hyperparameters such as the learning rate, batch size, and others. After each training epoch, the dataset is shuffled, and batches of real skeletal sequences along with their corresponding labels are sampled. During this phase, the generator creates fake skeletal sequences using a latent vector (sampled from a Gaussian distribution) concatenated with conditional labels. The discriminator is trained on real data to learn how to distinguish between fake and real samples. To prevent overfitting, the discriminator training process is disabled if its loss function falls below a threshold value. We apply singular value clipping for stability and perform weight updates based on the computed gradients. We employed curriculum learning during training by starting with simpler rehabilitation exercises (standing and balance), then progressively introducing more complex movements (assisted knee raises and assisted lateral leg swings) to improve model stability and performance. We also used stratified sampling to ensure each batch contains a balanced representation of each exercise, which prevents model bias toward more frequent or simpler movements and improves generalization across all exercise types.

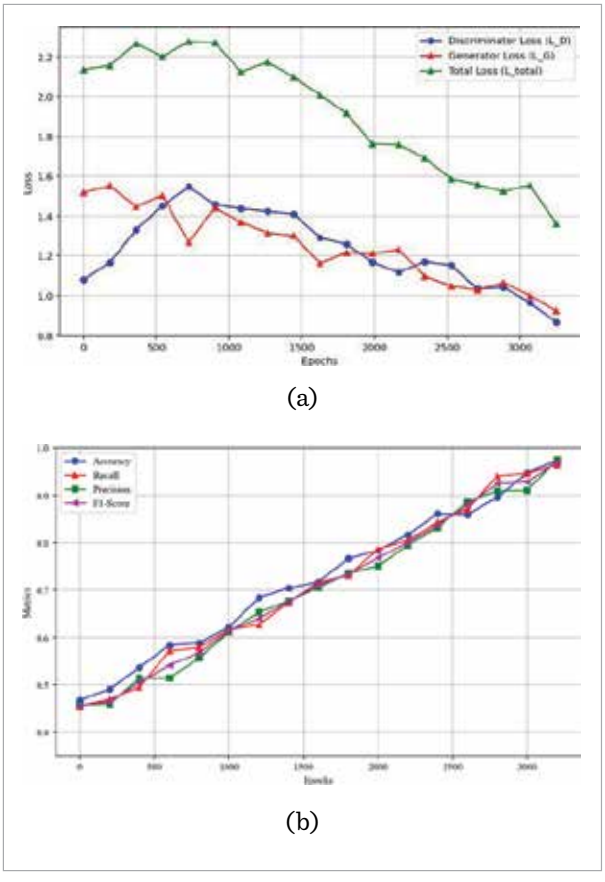
Table 1 presents the hyperparameter settings of the TCGAN. These values were selected based on several parameters as discussed in the next sections.

Table 1
Hyperparameters for the TCGAN Model.

Hyperparameter	Setup value
Dropout Rate (for G and D)	0.3
Learning rate η (for G and D)	0.0001
Batch size	64
Epochs	3150
Threshold	0.2
Maximum singular value threshold τ	0.1
λ smooth	0.1

The graph in Figure 4(a) displays curves of the generator, discriminator, and total loss functions obtained from training the TCGAN algorithm using the best hyperparameter values shown in Table 1. The generator exhibits excellent performance during training (Figure 4(b)). Initially, the accuracy, precision, recall, and F1 score metrics increase steadily at a good rate. For example, accuracy rises from 25% at the 1st epoch to 99% at the 3150th epoch. After the 3150th epoch, the model's accuracy remains stable (>99%). Moreover, the other metrics demonstrate similar stability. This indicates the model's strong learning capacity and its ability to generate realistic data. The TCGAN has been successfully trained, as evidenced by the loss curves, which show that both the discriminator loss (LD) and generator loss (LG) stabilize. The gradual decrease in Ltotal reflects effective optimization and regularization. The generator progressively improves at creating realistic sequences over time, as demonstrated by the declining of LG value, while the discriminator remains sufficiently challenged. This balance highlights the stability of the system.

Figure 4
Generator/Discriminator loss functions (a) and evolution of metrics (b) during training.



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. In our TCGAN, mode collapse is avoided thanks to its conditional nature. Data generation is controlled by specific conditions, primarily the exercise class, which prevents the model from producing repetitive or limited outputs. By leveraging these conditional inputs, the TCGAN ensures coherence and balance in the generated data.

4.2. Quantitative Model Evaluation

Quantitative GAN generator evaluation refers to the calculation of specific numerical scores used to summarize the quality of generated sequences.

4.2.1. The Fréchet Inception Distance (FID)

The Fréchet Inception Distance (FID) [6] is a measure of similarity between two sets of images. It is used for assessing the quality of the data generated by the proposed TCGAN. FID measures the similarity between the feature distributions of real and generated data. It is given by the following equation, which has been adapted for numerical data (keypoints):

$$FID = \|\mu 1 - \mu 2\|^2 + Tr\,(C1 + C2 - 2 \times \sqrt{C1 \times C2}), \quad (7)$$

Table 2
Obtained FID for each exercise across different epochs

Epoch	100	1000	2000	3000	3150
Standing and balance	4.78	2.07	1.65	0.85	0.84
Assisted lateral leg swings	5.01	2.21	1.69	0.91	0.90
Assisted knee raises	4.69	2.11	1.68	0.89	0.89
Total (entire dataset)	4.90	2.15	1.65	0.88	0.87

where μ1 and μ2 are the means of the features of real and generated data, respectively, and C1 and C2 denote the covariance matrices of the feature vectors of the real and generated data, respectively. The trace Tr is a function from linear algebra [28]. Interpreting FID values is crucial for assessing the quality of data generated by models like the TCGAN. A FID score close to zero indicates that the generated data is nearly indistinguishable from the real data. Careful selection and tuning of hyperparameters have significantly enhanced the performance of our TCGAN. As shown in Table 2, FID scores exhibit a clear decline across training epochs, indicating progressive improvement in the realism of the generated skeletal sequences. By epoch 3150, the total FID score reaches 0.87, demonstrating strong alignment between synthetic and real data distributions, which is a key indicator of high-quality motion synthesis. The most significant reduction in FID occurs between epochs 100 and 1000, highlighting the effectiveness of the curriculum learning strategy and improved loss function in accelerating model adaptation. The minimal change observed after epoch 3000 indicates that the model has reached near-optimal performance. Among individual exercises, standing and balance achieves the lowest FID score (0.84), due to the frontal camera view, which enables more accurate detection of keypoints compared to the side views used for the other two exercises. The small gap between individual exercise scores and the total dataset score confirms well-distributed learning, supported by stratified sampling and conditional inputs that prevent bias toward any specific movement type.

4.2.2. Classification Performance

One of the most effective ways to validate the performance of the TCGAN is by using the FID score to assess the quality of the generated data. However, this evaluation alone does not fully ensure the efficiency of the TCGAN. To comprehensively validate our model's capabilities, we employed three machine learning classification algorithms: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF). These algorithms enabled us to evaluate the quality of the synthetic data generated by the TCGAN and its similarity to real data. For training these classifiers, we used the same real dataset that was utilized during the training of our TCGAN.

After training these algorithms, we conducted tests using various sizes of synthetic data generated by our TCGAN. These tests were evaluated using several standard machine learning metrics: accuracy, precision, recall, and F1 score. Testing across different data sizes ensures the reliability and robustness of

our TCGAN, making it suitable for real-world applications where data size and variability are critical. The results of these tests are presented in Table 3. Across all configurations, the models achieve exceptionally high accuracy, precision, recall, and F1 scores, often exceeding 99%. These findings strongly indicate that the skeletal motion sequences generated by our TCGAN are statistically coherent, closely resembling real human movement patterns. This consistency across diverse classifiers is particularly significant. RF, SVM, and KNN differ fundamentally in how they model data, ranging from distance-based reasoning (KNN) to margin optimization (SVM) and ensemble decision-making (RF). The fact that all models perform equally well implies that the synthetic data lacks artifacts or biases that could mislead one type of classifier while favouring another. In other words, the generated movements exhibit generalizable structure, rather than overfitting to a specific model or dataset split. To ensure that these high scores reflect genuine generalization rather than memorization, we applied rigorous validation techniques, including cross-validation and testing on independent sub-datasets. These measures help confirm that the observed performance results from the quality of the synthetic data rather than from overfitting or noise exploitation. In summary, the consistently high classification metrics across models and dataset sizes provide strong empirical evidence that the TCGAN generates realistic, structurally sound motions.

Table 4 presents classification performance across four individuals performing three rehabilitation exercises: standing and balance, assisted lateral leg swings, and assisted knee raises. A time-based classifier was implemented to evaluate entire exercise se-

Table 3
Test results on datasets with different sizes.

Dataset size	1000			100000			1000000		
Classifier	RF	SVM	KNN	RF	SVM	KNN	RF	SVM	KNN
Accuracy	99.93%	99.69%	99.97%	99.84%	99.42%	99.22%	99.94%	99.64%	99.58%
Precision	99.37%	99.46%	100%	99.59%	99.84%	100%	99.29%	99.70%	99.66%
Recall	99.32%	99.12%	100%	99.99%	99.31%	100%	99.24%	99.94%	99.78%
F1-score	99.34%	99.29%	100%	99.79%	99.57%	100%	99.26%	99.82%	99.72%

Table 4
Performance metrics for each exercise and person.

Exercise	Person N°	Repetitions	Accuracy	Precision	Recall	F1- score	Mean Variance (px^2)
Standing and balance	1	30	99.00%	98.98%	98.64%	98.81%	7.2 ± 1.1
	2	30	98.56%	98.34%	98.67%	98.50%	8.4 ± 1.2
	3	30	98.42%	99.01%	98.36%	98.68%	9.7 ± 1.8
	4	30	98.88%	98.68%	99.00%	98.84%	6.8 ± 0.9
Assisted lateral leg swings	1	30	97.14%	96.69%	97.33%	97.01%	18.6 ± 2.9
	2	30	98.45%	96.72%	98.33%	97.52%	15.8 ± 2.4
	3	30	96.98%	96.68%	97.00%	96.84%	21.3 ± 3.5
	4	30	98.35%	96.73%	98.34%	97.53%	14.2 ± 2.1
Assisted knee raises	1	30	98.06%	99.00%	95.51%	97.23%	16.5 ± 2.3
	2	30	97.59%	97.66%	97.33%	97.50%	19.8 ± 3.1
	3	30	96.94%	96.67%	96.99%	96.83%	24.7 ± 4.2
	4	30	97.11%	96.33%	97.64%	96.98%	20.4 ± 3.4

quences generated by our TCGAN. This assessment differs from the per-image evaluations reported in Table 3, where classification was performed on individual frames. The time-based classifier was trained on real data using a neural network-based architecture and is structurally similar to the discriminator, which also functions as a binary classifier. The last column of Table 4 reports the Mean Variance (in pixels squared), a metric that quantifies motion uncertainty by measuring positional variability across multiple model outputs generated under identical input conditions. Specifically, we applied Monte Carlo Dropout during inference, keeping dropout active (rate = 0.3) and generating 50 synthetic movement sequences per input using the same latent vector z and condition y. The variance across these outputs provides an estimate of model confidence in keypoint placement. Each value in the final column thus reflects the average positional variance across all keypoints over the 50 samples, with standard deviation indicating local variability (e.g., 20.4 ± 3.4 px^2 means a mean variance of 20.4 px^2 with 3.4 px^2 fluctuation).

Despite all movements being executed at similar speeds, the best results are consistently observed during standing and balance tasks, where average accuracy reaches 98.72% and pixel variance remains low (an average of 8.0 ± 1.3 px^2), indicating strong spatial stability and temporal coherence. This superior performance results not only from the temporal modelling capabilities of the LSTM-based generator (G_0) but also from the improved visibility of keypoints when the person faces the camera directly. In this frontal view, MoveNet achieves more accurate and stable detection of skeletal keypoints, and it is easier for the patient to follow the movements since all the keypoints are clearly visible. In contrast, knee raises and leg swings exercises are performed from the side view, resulting in minor drops in recall and F1-score (e.g., Person 1 achieves 95.51% recall for knee rises). These decreases stem from challenges such as partial occlusion, limb overlap, or reduced visibility of certain joints. The increase in pixel variance observed in these side-view exercises (up to 24.7 ± 4.2 px^2) correlates with these fluctuations in performance, indicating that reduced visibility introduces greater uncertainty into keypoint localization, even though movement speed remains constant. This method of uncertainty estimation reveals a strong correlation (R^2 = 0.89) between Mean Variance and F1 scores, confirming that higher variance corresponds to lower classification confidence, which making Mean Variance a reliable indicator of motion quality. However, the classifier still maintains high overall scores, demonstrating that the generated sequences preserve sufficient realism and structural integrity. This resilience directly reflects the impact of several model components. The KL divergence-enhanced loss improves separability between real and generated data distributions, as seen in the consistently high precision values (e.g., 99.00% for Person 1 during knee raises). The gradient penalty contributes to training stability, which reflected in robust performance even under reduced visibility conditions such as side-view exercises. The smoothness regularization term suppresses jitter and abrupt transitions, especially effective in frontal-view sequences like standing and balance, where pixel variance remains low (8.0 ± 1.3 px^2) and classification scores are highest. Furthermore, the system's ability to maintain performance across unseen users supports the effectiveness of curriculum learning and stratified sampling, which ensure balanced exposure to exercise types and reduce overfitting. Importantly, the conditional input vector, including head and feet coordinates, enables motion adaptation to each patient's biomechanics, which is a key factor in maintaining consistency despite inter-individual variation.

Based on a psychotherapist's remarks, 28 px^2 was set as the threshold for acceptable motion quality. We noticed that movements exceeding this threshold were considered unreliable.

Table 5

Average of compensatory movements observed in Pre- and Post- our approach intervention.

Exercises	Before using our approach	After using our approach
Standing and balance	21	13
Assisted lateral leg swings	19	13
Assisted knee raises	23	12
Total	63	38

Table 5 presents a comparative evaluation of compensatory movement frequency before and after integrating our TCGAN-based real-time visual

guidance system. Across 240 trials (120 pre- and 120 post-intervention, with 40 trials per exercise), involving four participants performing three rehabilitation exercises, we observe a clear and clinically meaningful reduction in compensatory behaviours. The total number of compensatory movements decreased from 63 (pre- intervention) to 38 (post-intervention), representing a reduction of approximately 40%. This decline was consistent across all exercise types: standing and balance (−38%), assisted lateral leg swings (−32%), and assisted knee raises (−48%). These results strongly indicate that real- time visual guidance based on patient-specific skeletal references significantly improves movement accuracy and reduces reliance on maladaptive strategies. Importantly, the residual compensatory movements observed post- intervention were attributed by the physiotherapist primarily to inherent physical limitations, such as reduced range of motion or muscle weakness, rather than errors in following the generated guidance. This distinction reinforces the clinical relevance of our findings: while our system cannot fully overcome physiological impairments, it effectively minimizes compensations that arise from poor form awareness or lack of immediate corrective feedback, which are modifiable factors in home- based rehabilitation. This outcome highlights the value of real-time visual scaffolding in promoting correct motor patterns early in the rehabilitation process, potentially reducing the risk of long-term maladaptive plasticity and improving functional recovery. The data thus support our system's role not only as a motion generation tool but also as a behavioural intervention that actively shapes movement quality through continuous visual reinforcement.

4.3. Qualitative Model Evaluation

To assess the quality of the sequences generated by our TCGAN, we used data from the experiment involving the four individuals described in the previous section. As we mentioned it, tests were conducted in diverse environments with varying lighting conditions to evaluate the model's robustness. The participants, differing in size and body type, wore different outfits for each session. The first observation is that, across all exercises, the participants' landmarks were successfully detected (Step 1 of the process), enabling an accurate generation of exercises (Step 2 of the process).

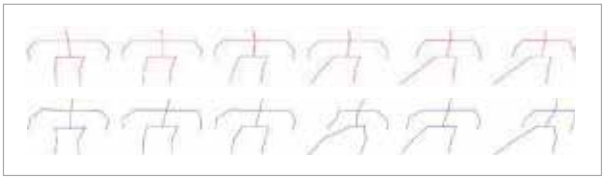
Table 6
Evaluation of training exercises: speed of creation and physiotherapist satisfaction.

Exercise	Number of exercises	Time to create 120 exercises	Positive opinion of the physiotherapist
Standing and balance	120	1.82 second	For 96.69% of exercises
Assisted lateral leg swings	120	1.75 second	For 96.67% of exercises
Assisted knee raises	120	1.98 second	For 95.26% of exercises
Average	120	1.85 second	For 96.21% of exercises

A key outcome of this evaluation is presented in Table 6, which quantifies both the efficiency and clinical acceptability of the generated exercises. The average time required to generate 120 exercise sequences was just 1.85 seconds, indicating strong computational efficiency and suitability for real-time deployment. More importantly, physiotherapist assessments revealed high satisfaction rates across all exercise types, with an average approval rate of 96.21%. Notably, standing and balance achieved the highest score at 96.69%, implying that the model performs particularly well in generating stable, posture-focused movements. These high approval rates indicate that the generated sequences are not only structurally coherent but also clinically meaningful, closely mirroring correct human movement patterns as demonstrated by trained individuals. This aligns with visual comparisons shown in Figure 5.

Another qualitative evaluation based on visual comparison is illustrated in Figure 5, which includes two image sequences depicting a standing and balance exercise. The first sequence, generated by our TCGAN, visually demonstrates the model's ability to

Figure 5
Standing and balance exercise generated by TCGAN (first row) and from dataset (second row).



produce realistic movements that closely follow the intended trajectory and posture of the exercise. The second sequence, taken from a dataset of real humans performing the same motion, serves as a reference. A comparison of the two sequences shows that the TCGAN-generated sequence effectively captures the essence of the real movement, indicating that the model can generate coherent and anatomically accurate exercises.

Taken together (Table 6 and Figure 5), these results confirm that our TCGAN effectively generates realistic, patient-adapted rehabilitation exercises that are both computationally efficient and visually accurate, making them suitable for real-time guidance in unsupervised home-based rehabilitation scenarios.

5. Ablation Study

In this section, we present a series of ablation experiments designed to evaluate the impact of key components and design choices in our model. Through systematic modifications, including changes to the sampling strategy, architectural components, and regularization techniques, we assess their contributions to performance, robustness, and generalization. The results from these experiments provide valuable insights into the relative importance of each factor and guide the refinement of our approach. Tests were conducted using input data collected from the study outlined in Table 4.

Table 7 evaluates how variations in camera resolution and frame rate affect motion realism. At 480p, performance remains strong (FID: 0.88; Accuracy: 97.17%), indicating that MoveNet retains sufficient spatial sensitivity for rehabilitation tasks where gross limb positioning is critical. However, at 240p, accuracy drops by over 3%, and FID rises sharply to 1.15, revealing that insufficient pixel density impairs keypoints detection, introducing noisy skeletal input into the TCGAN. This likely disrupts temporal coherence in generated sequences, as evidenced by lower F1-scores, highlighting the system's dependence on accurate initial pose estimation. Reducing the frame rate from 30 FPS to 10 FPS has negligible impact across all metrics, confirming that the LSTM- based generator requires only one accurate skeletal snapshot per movement repetition. This aligns with our conditional architecture, which synthesizes full motion sequences based on positional cues such as head and foot coordinates. The stability under low frame rates also supports deployment on resource-constrained devices. The interaction between low resolution and low frame rate (240p/10 FPS) yields the worst results, underscoring that while temporal redundancy can be reduced, spatial fidelity cannot be fully compensated. Our normalization and KL-divergence-enhanced loss help mitigate some noise, but they cannot fully recover lost structural information.

Table 8 presents a comparative evaluation of the TCGAN (our current model) and Pose2PoseGAN in terms of generated movement quality. TCGAN outperforms Pose2PoseGAN across all metrics, achieving superior FID (0.87 vs. 1.17) and higher classification performance (Accuracy: 97.96%, F1- score: 97.69%). This performance gain stems from TCGAN's ability to model long-term temporal dependencies, allowing it to generate accurate poses with smooth and continuous motion, which is crucial for rehabilitation applications. In contrast, Pose2PoseGAN is effective at generating transitions between individual poses but struggles to maintain motion continuity over extended sequences. This limitation results in slightly lower overall performance. None-

Table 7

Robustness evaluation of TCGAN across varying camera resolutions and frame rates.

Metrics	1080p 30FPS (Current)	480p 30FPS	240p 30FPS	1080p 10FPS	480p 10FPS	240p 10FPS
FID	0.87	0.88	1.15	0.88	0.90	1.15
Accuracy	97.96%	97.17%	94.85%	97.68%	96.83%	94.50%
Precision	97.62%	96.00%	91.96%	96.50%	95.50%	91.41%
Recall	97.76%	95.52%	92.42%	96.50%	95.02%	91.88%
F1-score	97.69%	95.76%	92.19%	96.50%	95.26%	91.65%

theless, it remains suitable for tasks involving repetitive or short- range movements, where precise pose-to-pose transitions are more important than global temporal coherence.

Table 8
Comparative evaluation of GAN architectures for movement generation.

Metrics	TCGAN (Current model)	Pose to pose
Total FID	**0.87**	1.17
Accuracy	**97.96%**	94.17%
Precision	**97.62%**	91.37%
Recall	**97.76%**	90.91%
F1-score	**97.69%**	91.14%

Table 9 compares the impact of stratified and uniform sampling strategies on TCGAN's performance. While both approaches yield high- quality results, stratified sampling consistently leads to marginally better outcomes across all metrics. The total FID improves slightly (0.87 vs.0.88), and classification performance sees measurable gains in precision (97.62% vs. 96.98%), recall (97.76% vs. 96.02%), and F1-score (97.69% vs. 96.50%). These improvements are attributed to stratified sampling's ability to preserve class balance and representation of rare movement types during training. This ensures that the model is exposed to diverse yet proportionally distributed motion patterns, enhancing generalization and reducing overfitting to dominant movement categories. In contrast, uniform sampling introduces slight class imbalance, which may cause underrepresentation of less frequent motions and result in reduced recall. Overall, these findings highlight the importance of data distribution strategies in generative motion models and validate our use of stratified sampling to support robust and consistent performance across varied rehabilitation movements.

Table 9
Impact of sampling strategies on TCGAN performance.

Metrics	Stratified sampling (Current model)	Uniform sampling
Total FID	**0.87**	0.88
Accuracy	**97.96%**	97.67%
Precision	**97.62%**	96.98%
Recall	**97.76%**	96.02%
F1-score	**97.69%**	96.50%

Table 10 evaluates the impact of different loss function designs on TCGAN performance and training stability. Our composite loss, which integrates adversarial loss, KL divergence, smoothness regularization, and gradient penalty, achieves the best FID score (0.87) and highest classification metrics, indicating superior realism and temporal coherence in generated motion sequences. In contrast, using only a biomechanical loss or Wasserstein-only objective leads to reduced accuracy (96.33% and 94.33%, respectively) and increased FID (0.92 and 1.16), showing that purely domain-based or distributional losses fail to capture both realism and dynamic movement patterns. The MSE-based loss performs worst (FID: 1.30), producing rigid, unnatural sequences and suffering from mode collapse, as reflected in low precision (88.83%) and high instability (σ/μ = 0.29). Training stability was assessed using the coefficient of variation (σ/μ) of the generator loss over the final 10% of training epochs, where lower values indicate smoother and more consistent convergence. These results highlight the need for a hybrid loss that balances adversarial realism, biomechanical plausibility, and temporal smoothness. Training stability also degrades significantly without our improved loss, with the gradient penalty and KL term playing key roles in aligning real and fake data distributions while enforcing Lipschitz continuity. Smoothness regularization further ensures natural transitions between frames, essential for generating rehabilitation-appropriate guidance.

Table 11 compares TCGAN performance across different architectural designs for the generator and discriminator. Our current model, featuring an LSTM-based temporal generator and dual-path (spatial-temporal) discriminator, achieves the lowest FID (0.87) and highest classification scores, demonstrating superior realism and temporal coherence in generated motion sequences. Replacing the generator with a Transformer (Transformer-G) results in slightly higher FID (0.89) and reduced precision (95.96%), demonstrating that while Transformers capture long-range dependencies, they may introduce instability in sequential skeletal generation

Table 10
Effect of loss function variants on TCGAN performance and training stability.

Metrics	Current composite	Biomechanical loss	Wasserstein only	MSE
Total FID	**0.87**	0.92	1.16	1.30
Accuracy	**97.96%**	96.33%	94.33%	92.17%
Precision	**97.62%**	95.43%	91.88%	88.83%
Recall	**97.76%**	93.53%	90.95%	87.50%
F1-score	**97.69%**	94.47%	91.41%	88.16%
Training stability (σ/μ)	**0.02**	0.09	0.19	0.29

Table 11
Impact of generator and discriminator architecture variants on TCGAN performance.

Metrics	Current model	Transformer-G	Temporal-Only-D	GRU-G+CNN-D	TCN-G
Total FID	**0.87**	0.89	1.05	1.18	0.89
Accuracy	**97.96%**	97.00%	95.17%	93.83%	97.50%
Precision	**97.62%**	95.96%	93.85%	92.23%	96.48%
Recall	**97.76%**	95.00%	91.50%	89.00%	96.00%
F1-score	**97.69%**	95.48%	92.66%	90.59%	96.24%

without careful positional encoding or masking. The Temporal-Only-D variant, which removes the spatial pathway from the discriminator, shows a significant FID increase (1.05), indicating that spatial accuracy is essential for realistic pose alignment. Using GRU-based generation with CNN-based discrimination (GRU-G+CNN-D) further degrades all metrics (FID: 1.18), highlighting limitations in capturing temporal dynamics with shallow recurrent units and local CNN features. In contrast, a Temporal Convolutional Network (TCN-G) maintains strong performance (FID: 0.89), showing that convolutional architectures can also model temporal dependencies effectively if properly designed. These results confirm the effectiveness of LSTM layers in capturing temporal dependencies within skeletal sequences, while the dual-path discriminator enhances both spatial accuracy and temporal coherence. This architectural choice is critical for real-time rehabilitation applications, where realistic and anatomically plausible motion generation directly influences patient engagement, movement learning, and execution precision.

Table 12 evaluates how reducing the number of keypoints affects both motion realism (measured by FID) and perceived clinical value. Interestingly, FID scores remain relatively stable across all configurations (ranging from 0.87 to 0.89), indicating that even with fewer keypoints, the TCGAN maintains high fidelity in generating motion sequences. However, physiotherapist evaluations reveal a significant drop in perceived usefulness when using fewer than 9 keypoints. While models using 13 or 11 keypoints were rated as "Good", the 9- keypoint model was rated "Average", and the 7-keypoint variant was deemed "Bad". This shows that while motion realism may be preserved metrically, clinical relevance diminishes when too few keypoints are used, limiting the system's ability to capture essential joint movements for rehabilitation guidance. This highlights a critical trade-off: although minimal skeletal data may suffice for basic motion generation, accurate representation of major joints, especially in the limbs, is necessary for meaningful real- time guidance. These findings support our decision to use 13 keypoints, ensuring both technical performance and clinical applicability in guiding stroke patients through corrective exercises.

Table 13 evaluates how dynamic dropout (DD) and adversarial pruning (AP) affect TCGAN perfor-

Table 12
Effect of keypoint number on clinical usefulness.

Metrics	13 keypoints	11 keypoints	9 keypoints	7 keypoints
Total FID	**0.87**	0.88	0.87	0.89
Physiotherapist opinion on usefulness	**Good**	Good	Average	Bad

Table 13
Impact of dynamic dropout (DD) and adversarial pruning (AP) on TCGAN performance.

Metrics	FID	Accuracy	Precision	Recall	F1-score
Current model	0.87	97.96%	97.62%	97.76%	97.69%
DD (G_0)	0.91	96.50%	95.00%	94.53%	94.76%
DD (G_1)	0.89	97.19%	95.57%	96.04%	95.80%
DD (temporal D)	1.14	94.83%	92.50%	92.04%	92.27%
DD (spatial D)	**0.83**	**98.83%**	**98.50%**	**98.01%**	**98.25%**
AP (G_0, 40%)	1.32	91.50%	87.50%	87.06%	87.28%
AP (G_1, 40%)	1.03	95.50%	93.50%	93.03%	93.27%
AP (spatial D, 40%)	0.88	97.46%	96.62%	96.98%	96.80%
AP (temporal D, 40%)	1.39	89.17%	84.00%	83.58%	83.79%

mance. Interestingly, applying DD to different components yields mixed results. When applied to the spatial discriminator, DD improves all metrics (FID: 0.83; Accuracy: 98.83%), confirming that controlled neuron suppression enhances generalization by reducing overfitting to specific pose configurations. In contrast, DD applied to the temporal discriminator significantly degrades FID (1.14) and classification scores, indicating that suppressing temporal pathways disrupts motion coherence. Adversarial pruning generally harms performance, especially when applied to the temporal discriminator (FID: 1.39; F1: 83.79%) or G_0 (FID: 1.32; F1: 87.28%), confirming that removing critical neurons impairs the model's ability to generate smooth, realistic sequences. However, AP on the spatial discriminator causes only minor degradation (FID: 0.88), demonstrating it retains enough spatial modelling capacity for rehabilitation-appropriate guidance.

Table 14 demonstrates the critical role of LSTM layers in generating temporally coherent skeletal sequences. Removing temporal modelling entirely (i.e., using dense layers only) leads to a sharp FID increase (1.33), significant drops in all classification metrics, and a low motion coherence score of 4.1, indicating clearly disjointed or unnatural movement sequences. Replacing the multi-layer LSTM with a single 512-unit layer improves performance (FID: 0.95; motion coherence: 8.3), but still underperforms compared to our full model. Reducing the architecture to a 2-layer LSTM (256 and 128 units) yields results closer to the current model (FID: 0.90; F1: 95.92%; motion coherence: 9.2), showing that depth and hidden state capacity are key to capturing long-term dependencies in motion data. The high physiotherapist-rated motion coherence (9.6) of our full model confirms that multi-layer LSTM enhances realism and smoothness, essential for guiding patients through complex rehabilitation exercises. These findings validate our architectural choice, ensuring both quantitative performance and clinically meaningful output.

Table 15 shows that both batch normalization (BN) and gradient penalty (GP) play critical roles in stabilizing TCGAN training and improving motion generation quality. When both components are used together (current model), the system achieves the lowest FID (0.87), highest classification scores, and best training stability ($\sigma/\mu = 0.02$), confirming their combined effectiveness. Training stability was evaluated using the coefficient of variation (σ/μ) of the

Table 14
Contribution of LSTM layers to motion coherence in TCGAN.

Metrics	Current model	No temporal layers (Dense only)	1-Layer LSTM (Single 512-unit LSTM)	Reduced LSTM (2-layer LSTM 256 and128)
FID	**0.87**	1.33	0.95	0.90
Accuracy	**97.96%**	90.73%	95.99%	96.95%
Precision	**97.62%**	87.96%	95.41%	95.69%
Recall	**97.76%**	84.00%	92.57%	96.15%
F1-score	**97.69%**	85.93%	93.97%	95.92%
Motion coherence (1-10) *	**9.6**	4.1	8.3	9.2

*Motion coherence: physiotherapist opinion from 0 to 10

generator loss during the last 10% of training epochs, with lower values indicating more stable and consistent convergence. Removing both BN and GP leads to severe degradation: FID rises to 1.35, accuracy drops to 88.94%, and training instability soars ($\sigma/\mu = 0.35$). This indicates that without these mechanisms, the model struggles to align real and generated data distributions, resulting in poor skeletal sequences and unstable learning dynamics.

Using GP alone improves stability over no BN + no GP ($\sigma/\mu = 0.20$), but performance remains suboptimal (FID: 1.08; Accuracy: 94.64%), indicating that while GP helps enforce Lipschitz continuity, it cannot fully compensate for the lack of BN in normalizing layer inputs. In contrast, BN alone significantly boosts performance (FID: 0.95; Accuracy: 96.49%) and moderately improves stability ($\sigma/\mu = 0.07$), highlighting its importance in maintaining consistent feature distributions during training. These findings support our architectural choice to combine BN and GP, ensuring both stable training and high-quality, realistic motion generation essential for reliable rehabilitation guidance.

Table 15
Impact of batch normalization (BN) and gradient penalty (GP) on training stability and performance.

Metrics	Current model	GP only	BN only	No BN + No GP
FID	**0.87**	1.08	0.95	1.35
Accuracy	**97.96%**	94.64%	96.49%	88.94%
Precision	**97.62%**	92.89%	95.45%	85.71%
Recall	**97.76%**	91.04%	94.03%	80.60%
F1-score	**97.69%**	91.96%	94.74%	83.08%
Training stability (σ/μ)	**0.02**	0.20	0.07	0.35

Table 16
Impact of loss function components on TCGAN performance.

Metrics	Current model (Full Loss)	Without KL divergence	Without Wasserstein distance
FID	**0.87**	0.90	0.91
Accuracy	**97.96%**	96.87%	97.30%
Precision	**97.62%**	95.19%	95.45%
Recall	**97.76%**	95.65%	97.67%
F1-score	**97.69%**	95.42%	96.55%

Table 16 evaluates how individual components of the TCGAN loss function affect motion generation quality and realism. Our full loss, which combines adversarial loss, KL divergence, smoothness regularization, and gradient penalty, achieves the best FID score (0.87) and highest classification metrics, confirming its effectiveness in generating realistic and temporally coherent skeletal sequences. Removing the KL divergence term increases FID to 0.90 and significantly reduces precision (95.19%) and F1-score (95.42%), indicating that this component plays a key role in aligning real and generated data distributions. The drop in separability likely leads to less distinct or noisy motion patterns, affecting both realism and consistency across frames.

Disabling the Wasserstein distance component also degrades performance (FID: 0.91), though recall remains nearly unchanged (97.67%). This imply that while Wasserstein distance contributes to visual fidelity and sample diversity, the KL term is more critical for structural accuracy and sharpness in pose estimation. These findings validate our composite loss formulation, where each term addresses a specific challenge: KL divergence improves distribution alignment, Wasserstein distance enhances realism, and smoothness/gradient penalty ensure training stability and temporal coherence, which are essential properties for real-time rehabilitation guidance.

Table 17
Comparison of curriculum learning and standard learning techniques in TCGAN.

Metrics	With curriculum learning	Without curriculum learning
FID	**0.87**	0.89
Accuracy	**97.96%**	97.32%
Precision	**97.62%**	96.41%
Recall	**97.76%**	96.85%
F1-score	**97.69%**	96.63%
Epochs	**3150**	3700

Table 17 compares the performance of TCGAN trained with and without curriculum learning. The model trained using curriculum learning achieves slightly better quantitative results across all metrics while requiring fewer training epochs (3150 vs. 3700). This implies that curriculum learning enhances both convergence efficiency and the quality of motion generation. The observed improvement in classification metrics indicates that introducing exercises progressively, from simpler to more complex movements, enables the generator to learn smoother and more realistic skeletal transitions. This staged learning approach mitigates early overfitting to complex motion patterns and contributes to more stable training dynamics. Additionally, curriculum learning improves generalization by ensuring balanced exposure to various exercise types through stratified sampling, as detailed in Section 3.4. The reduction in required epochs further implies faster convergence without compromising realism or accuracy, thereby increasing computational efficiency during training.

6. Comparative Study

The comparative analysis in Table 18 highlights the key differences between our proposed approach and existing solutions for rehabilitation exercises. As shown in the second column, prior methods fall into two main categories: vision-based and wearable- based. The majority of these approaches rely on classification techniques, providing feedback only after exercises are completed. While this method allows for performance assessment, it fails to prevent patients from repeatedly executing incorrect movements, which can slow down recovery or even cause further complications. Unlike these approaches, our proposed method offers real-time skeletal representation, providing patients with continuous guidance as they perform exercises. This feature ensures that incorrect postures and compensatory movements are immediately corrected, reducing the risk of long-term impairments. Rather than simply classifying movements as correct or incorrect after execution, our system dynamically adjusts to the patient's real-time posture, ensuring that each movement aligns with the prescribed rehabilitation exercise. Furthermore, our method introduces a personalized and adaptive rehabilitation framework, in contrast to the static classification-based models used in prior works such as [5, 14, 17]. By integrating LSTM and ANN, our approach dynamically tailors exercises to each patient's unique characteristics enhancing rehabilitation efficiency and effectiveness. Unlike conventional methods, which do not adapt to individual differences, our model evolves with the patient's recovery trajectory, ensuring optimal exercise execution. Another key advantage of our approach is the precision of skeletal representation. While some vision-based models (e.g., [26] and [20]) employ CNNs and LSTMs for movement classification, they lack interactive correction mechanisms. In contrast, our system actively guides the patient through the correct motion, minimizing the risk of compensatory movements. This level of precision, achieved through real-time motion tracking, significantly enhances rehabilitation outcomes. Additionally, our fully vision-based approach eliminates the need for wearable sensors, addressing one of the major drawbacks of wearable-based methods. While wearables such as those in [21, 12, 3] provide precise motion tracking, they often introduce high

Table 18
Comparison of the proposed approach with the state-of-the-art.

Reference	Data acquisition	Used technique	Method
[26]	Vision	Classification	Temporal Convolutional Network
[5]	Vision	Classification	GCN
[21]	Wearable	Classification	CNN
[14]	Vision/Wearable	Classification	ResNet3D-50, MLPMixer, Transformer encoders
[12]	Wearable	Classification	Fully Connected Neuronal Network, LSTM
[23]	Wearable	Classification	CNN
[3]	Wearable	Classification	SVM
[29]	Wearable	Classification	1D CNN
[16]	Vision	Classification	DT, LR, SVM, LSTM, Artificial Neural Network
[17]	Vision	Classification	DT, LR, SVM, LSTM, Feedforward Neural Network
[20]	Vision	Classification	CNN, LSTM
Proposed approach	**Vision**	**Creation of tailored exercise to be mimicked by the patient**	**LSTM, Artificial Neural Network**

costs, discomfort, and usability challenges, making them less practical for long-term rehabilitation. Unlike hybrid systems such as [14], which combine vision and wearable data but still rely on post-exercise classification, our approach ensures seamless, hardware-free deployment, making it more scalable and accessible for home-based and tele- rehabilitation settings. By combining real-time correction, personalized adaptation, and a hardware-free setup, our approach significantly improves rehabilitation efficiency. Patients receive instant feedback, allowing them to correct mistakes immediately rather than waiting for post-exercise evaluations. This not only enhances patient autonomy but also accelerates recovery by ensuring that each movement is performed correctly from the outset.

7. Conclusion

In this study, a TCGAN-based motion generation system for post-stroke physical rehabilitation exercises has been proposed. 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 the rehabilitation session. By dynamically generating motion sequences adapted to each patient's characteristics (such as height, posture, and positioning) our system enables precise and effective rehabilitation, reducing the likelihood of incorrect or compensatory movements that may hinder recovery.

The experimental results confirm the high accuracy and realism of the generated skeletal sequences, with a FID score of 0.87, demonstrating a strong similarity between synthetic and real motion data. Furthermore, our comparative analysis with state-of-the- art rehabilitation techniques highlights the superiority of our approach in terms of real- time correction and personalized movement adaptation, leading to more effective rehabilitation outcomes. Additionally, evaluations using machine learning classifiers further validate the quality of the generated exercises, showing that they closely resemble real human movements. The reduction of compensatory movements observed in 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 broader range of rehabilitation exercises, enhancing adaptability to diverse patient profiles, and integrating additional patient data, such as EMG signals or biomechanical feedback, to improve mo-

tion accuracy. Clinical validation through extensive trials with stroke patients will be crucial to assess 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. We will also focus on modelling patient- specific factors, such as fatigue and motor learning, with adaptive mechanisms to adjust guidance based on performance and progress.

References

1. Boukhennoufa, I., Zhai, X., Utti, V., Jackson, J., McDonald-Maier, K. D. Wearable Sensors and Machine Learning in Post-Stroke Rehabilitation Assessment: A Systematic Review. Biomedical Signal Processing and Control, 2022, 71, 103197. https://doi.org/10.1016/j.bspc.2021.103197
2. Burton, J. K., Ferguson, E. E. C., Barugh, A. J., Walesby, K. E., MacLullich, A. M. J., Shenkin, S. D., Quinn, T. J. Predicting Discharge to Institutional Long-Term Care After Stroke: A Systematic Review and Meta-Analysis. Journal of the American Geriatrics Society, 2018, 66(1), 161-169. https://doi.org/10.1111/jgs.15101
3. Cai, S., Li, G., Su, E., Wei, X., Huang, S., Ma, K., Zheng, H., Xie, L. Real-Time Detection of Compensatory Patterns in Patients with Stroke to Reduce Compensation During Robotic Rehabilitation Therapy. IEEE Journal of Biomedical and Health Informatics, 2020, 24(9), 2630-2638. https://doi.org/10.1109/JBHI.2019.2963365
4. Chen, J., Or, C. K., Chen, T. Effectiveness of Using Virtual Reality-Supported Exercise Therapy for Upper Extremity Motor Rehabilitation in Patients with Stroke: Systematic Review and Meta-Analysis of Randomized Controlled Trials. Journal of Medical Internet Research, 2022, 24(6), e24111. https://doi.org/10.2196/24111
5. Deb, S., Islam, M. F., Rahman, S., Rahman, S. Graph Convolutional Networks for Assessment of Physical Rehabilitation Exercises. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2022, 30, 410-419. https://doi.org/10.1109/TNSRE.2022.3150392
6. Dowson, D. C., Landau, B. V. The Fréchet Distance Between Multivariate Normal Distributions. Journal of Multivariate Analysis, 1982, 12(3), 450-455. https://doi.org/10.1016/0047-259X(82)90077-X
7. Edwards, J. D., Dominguez-Vargas, A. U., Rosso, C., Branscheidt, M., Sheehy, L., Quandt, F., et al. A Translational Roadmap for Transcranial Magnetic and Direct Current Stimulation in Stroke Rehabilitation: Consensus-Based Core Recommendations from the Third Stroke Recovery and Rehabilitation Roundtable. International Journal of Stroke, 2024, 19(2), 145-157. https://doi.org/10.1177/17474930231203982
8. Everard, G., Luc, A., Doumas, I., Ajana, K., Stoquart, G., Edwards, M. G., Lejeune, T. Self-Rehabilitation for Post-Stroke Motor Function and Activity - A Systematic Review and Meta-Analysis. Neurorehabilitation and Neural Repair, 2021, 35(12), 1043-1058. https://doi.org/10.1177/15459683211048773
9. Gómez-Portes, C., Carneros-Prado, D., Albusac, J., Castro-Schez, J. J., Glez-Morcillo, C., Vallejo, D. PhyRe Up! A System Based on Mixed Reality and Gamification to Provide Home Rehabilitation for Stroke Patients. IEEE Access, 2021, 9, 139122-139137. https://doi.org/10.1109/ACCESS.2021.3118842
10. Hadjipanayi, C., Banakou, D., Michael-Grigoriou, D. Virtual Reality Exergames for Enhancing Engagement in Stroke Rehabilitation: A Narrative Review. Heliyon, 2024, 10(18), e37581. https://doi.org/10.1016/j.heliyon.2024.e37581
11. Jung, H.-T., Park, J., Jeong, J., Ryu, T., Kim, Y., Lee, S. I. A Wearable Monitoring System for At-Home Stroke Rehabilitation Exercises: A Preliminary Study. In 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 2018, 13-16. https://doi.org/10.1109/BHI.2018.8333358
12. Kaku, A., Parnandi, A., Venkatesan, A., Pandit, N., Schambra, H., Fernandez-Granda, C. Towards Data-Driven Stroke Rehabilitation via Wearable Sensors and Deep Learning. In Machine Learning for Healthcare Conference, 2020, 143-171.
13. Kashi, S., Polak, R. F., Lerner, B., Rokach, L., Levy-Tzedek, S. A Machine-Learning Model for Automatic Detection of Movement Compensations in Stroke Patients. IEEE Transactions on Emerging Topics in Computing, 2020, 9(3), 1234-1247. https://doi.org/10.1109/TETC.2020.2988945
14. Kim, D., Park, J. E., Kim, M. J., Byun, S. H., Jung, C. I., Jeong, H. M., Woo, S. R., Lee, K. H., Lee, M. H., Jung, J.-W., et al. Automatic Assessment of Upper Extremity Function and Mobile Application for Self-Administered Stroke Rehabilitation. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2024, 32, 652-661. https://doi.org/10.1109/TNSRE.2024.3358497

15. Lee, M. H., Siewiorek, D. P., Smailagic, A., Bernardino, A., Bermúdez i Badia, S. Learning to Assess the Quality of Stroke Rehabilitation Exercises. In Proceedings of the 24th International Conference on Intelligent User Interfaces, California, USA, 2019, 218-228. https://doi.org/10.1145/3301275.3302273

16. Lee, M. H., Siewiorek, D. P., Smailagic, A., Bernardino, A., Bermúdez i Badia, S. An Exploratory Study on Techniques for Quantitative Assessment of Stroke Rehabilitation Exercises. In Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, Genoa, Italy, 2020, 303-307. https://doi.org/10.1145/3340631.3394872

17. Lee, M. H., Siewiorek, D. P., Smailagic, A., Bernardino, A., Bermúdez i Badia, S. Design, Development, and Evaluation of an Interactive Personalized Social Robot to Monitor and Coach Post-Stroke Rehabilitation Exercises. User Modeling and User-Adapted Interaction, 2023, 33(2), 545-569. https://doi.org/10.1007/s11257-022-09348-5

18. Leong, S. C., Tang, Y. M., Toh, F. M., Fong, K. N. Examining the Effectiveness of Virtual, Augmented, and Mixed Reality (VAMR) Therapy for Upper Limb Recovery and Activities of Daily Living in Stroke Patients: A Systematic Review and Meta-Analysis. Journal of Neuroengineering and Rehabilitation, 2022, 19(1), 93. https://doi.org/10.1186/s12984-022-01071-x

19. Li, C., Cheng, L., Yang, H., Zou, Y., Huang, F. An Automatic Rehabilitation Assessment System for Hand Function Based on Leap Motion and Ensemble Learning. Cybernetics and Systems, 2020, 52(1), 3-25. https://doi.org/10.1080/01969722.2020.1827798

20. Liao, Y., Vakanski, A., Xian, M. A Deep Learning Framework for Assessing Physical Rehabilitation Exercises. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2020, 28(2), 468-477. https://doi.org/10.1109/TNSRE.2020.2966249

21. Lin, P. J., Zhai, X., Li, W., Li, T., Cheng, D., Li, C., Pan, Y., Ji, L. A Transferable Deep Learning Prognosis Model for Predicting Stroke Patients' Recovery in Different Rehabilitation Trainings. IEEE Journal of Biomedical and Health Informatics, 2022, 26(12), 6003-6011. https://doi.org/10.1109/JBHI.2022.3205436

22. Luo, Z., Lim, A. E. P., Durairaj, P., Tan, K. K., Verawaty, V. Development of a Compensation-Aware Virtual Rehabilitation System for Upper Extremity Rehabilitation in Community-Dwelling Older Adults with Stroke. Journal of NeuroEngineering and Rehabilitation, 2023, 20(1), 56. https://doi.org/10.1186/s12984-023-01183-y

23. Panwar, M., Biswas, D., Bajaj, H., Jöbges, M., Turk, R., Maharatna, K., Acharyya, A. Rehab-Net: Deep Learning Framework for Arm Movement Classification Using Wearable Sensors for Stroke Rehabilitation. IEEE Transactions on Biomedical Engineering, 2019, 66(11), 3026-3037. https://doi.org/10.1109/TBME.2019.2899927

24. Pereira, C. M., Greenwood, N., Jones, F. From Recovery to Regaining Control of Life - The Perspectives of People with Stroke, Their Carers and Health Professionals. Disability and Rehabilitation, 2020, 43(20), 2897-2908. https://doi.org/10.1080/09638288.2020.1722263

25. Saito, M., Matsumoto, E. Temporal Generative Adversarial Nets with Singular Value Clipping. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017, 2830-2839. https://doi.org/10.1109/ICCV.2017.308

26. Sardari, S., Sharifzadeh, S., Daneshkhah, A., Loke, S. W., Palade, V., Duncan, M. J., Nakisa, B. LightPRA: A Lightweight Temporal Convolutional Network for Automatic Physical Rehabilitation Exercise Assessment. Computers in Biology and Medicine, 2024, 173, 108382. https://doi.org/10.1016/j.compbiomed.2024.108382

27. Shahmoradi, L., Almasi, S., Ahmadi, H., Bashiri, A., Azadi, T., Mirbagherie, A., Ansari, N. N., Honarpishe, R. Virtual Reality Games for Rehabilitation of Upper Extremities in Stroke Patients. Journal of Bodywork and Movement Therapies, 2021, 26, 113-122. https://doi.org/10.1016/j.jbmt.2020.10.006

28. Solanki, A., Nayyar, A., Naved, M. Generative Adversarial Networks for Image-to-Image Translation. Elsevier, 2021. https://doi.org/10.1016/C2020-0-00284-7

29. Wang, J., Li, C., Zhang, B., Zhang, Y., Shi, L., Wang, X., Zhou, L., Xiong, D. Automatic Rehabilitation Exercise Task Assessment of Stroke Patients Based on Wearable Sensors with a Lightweight Multichannel 1D-CNN Model. Scientific Reports, 2024, 14(1), 19204. https://doi.org/10.1038/s41598-024-68204-1

30. Wang, X., Fu, Y., Ye, B., Babineau, J., Ding, Y., Mihailidis, A. Technology-Based Compensation Assessment and Detection of Upper Extremity Activities of Stroke Survivors: Systematic Review. Journal of Medical Internet Research, 2022, 24(6), e34307. https://doi.org/10.2196/34307

31. Wu, X., Zhang, Q., Qiao, J., Chen, N., Wu, X. Calligraphy-Based Rehabilitation Exercise for Improving the Upper Limb Function of Stroke Patients: Protocol for an Evaluator-Blinded Randomised Controlled Trial. BMJ Open, 2021, 12(5), e052046. https://doi.org/10.1136/bmjopen-2021-052046