A Multidimensional Quality Assessment of Synthetic Movement Data for Post-stroke Rehabilitation

Aline P.A Ferreira[1], Lucas O. Alves[1], Fabio A.F Molinares[1], Marcelo P. Guimarães[2], Vinícius F.S. Mota[1], Leonardo C.D. Rocha[3], and Diego R.C. Dias[1(✉)]

[1] Universidade Federal do Espírito Santo, Vitória, Brazil
{aline.p.ferreira,lucas.o.alves}@edu.ufes.br,
{fabio.molinares,vinicius.f.mota,diego.dias}@ufes.br
[2] Universidade Federal de São Paulo, São Paulo, Brazil
marcelo.paiva@unifesp.br
[3] Universidade Federal de São João del-Rei, São João del-Rei, Brazil

Abstract. It is difficult to create good artificial intelligence models for motor rehabilitation when we do not have sufficient clinical data. To solve this, we tested two approaches to create synthetic data for the squat exercise: the use of controlled Gaussian noise and the use of GAN networks. We analyzed whether this data was statistically accurate, whether it made clinical sense, and whether it was useful in practice. The results showed a clear difference. Gaussian noise was much more efficient, preserving the rhythm of the movement and the anatomical logic, which helped improve the performance of models such as Random Forest and LSTM. GAN networks, on the other hand, although they reproduced the joint angles well, failed to maintain the continuity of movement over time, proving to be less consistent for practical use. These findings suggest that, for complex and highly coordinated multi-joint movements, controlled stochastic methods may currently outperform standard generative deep learning models in maintaining structural integrity, providing empirical evidence for future research in data-driven rehabilitation.

Keywords: Synthetic Data · Neuromotor Rehabilitation · Generative Adversarial Networks · Time Series Analysis · Data Augmentation

1 Introduction

The availability of high-quality data is fundamental to the development and validation of accurate and reliable statistical models in clinical rehabilitation. However, clinical data collection faces significant challenges that compromise research in this field. Among these challenges are the high costs of acquiring motion capture equipment, the maintenance of specialized infrastructure, and temporal

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026, LNCS 16768, pp. 233–250, 2027.
https://doi.org/10.1007/978-3-032-30488-9_15

constraints in the collection process, which often require weeks or months to obtain an adequate sample. Furthermore, patient recruitment for longitudinal studies further limits the quantity and quality of data available for research.

To capture movement data in neuromotor rehabilitation, inertial sensors are positioned on different body joints or optical devices are employed, representing a specific type of time series [1,4,5,7,8]. Each movement is recorded as a sequence of frames, where each row represents a temporal instance containing complete information on joint angles and coordinates from various sensors. However, given the multiple temporal structures involved, these data require validation methodologies that not only consider marginal statistical properties but also preserve temporal dynamics and the clinical plausibility of the generated data. Given these challenges, synthetic data generation becomes a promising approach to overcome the limitations of available real-world data.

To create this data, we used artificial intelligence models that have gained prominence, such as generative networks (GANs) [6], known for creating very realistic synthetic data. However, we needed to understand which technique works best when the challenge is dealing with time series of clinical movements [2]. After all, before using this data in practice, it is essential to ensure that it is faithful to real patterns, since this has a direct impact on health [3]. Therefore, we decided to compare two different approaches: one based on Gaussian noise and the other using GANs themselves.

Our analysis was divided into four main criteria to ensure data quality [9]. First, we checked statistical accuracy, that is, whether the generated numbers accurately reflect the real data. Then, we verified the movement rhythm, ensuring that the temporal sequence is natural. We also evaluated clinical logic to ensure that the created movements are physically possible and respect human anatomy [10]. Finally, we tested practical utility, confirming whether this new data actually helps improve the performance of artificial intelligence models when used in training.

Thus, we emphasize that no single metric can fully capture the quality of synthetic data in clinical applications, underscoring the need for a multidimensional approach to evaluate the generated data comprehensively. Therefore, we consider that the results provide empirical evidence on which synthesis methodology produces higher-quality data for clinical movement time series, opening important avenues for future research and applications in neuromotor rehabilitation.

2 Methodology

In this section, we describe the techniques used to generate, evaluate, and compare synthetic time-series motion data, focusing on two distinct approaches: controlled Gaussian noise and GANs. The experimental development was carried out in a Google Colab environment using Python 3.12. The workflow is available

on GitHub[1], where the generation, preprocessing, multidimensional evaluation, and performance comparison steps are documented.

2.1 Dataset and Preprocessing

The data used in this study consist of time series of body movements captured by inertial sensors placed on multiple joints of the human body during the execution of standardized neuromotor rehabilitation tasks. For the analyses, we used the deep squat, a fundamental movement in post-stroke rehabilitation that requires complex coordination among the hip, knee, and ankle. We organize the data into a CSV file containing 27,063 temporal records (frames) and 65 columns, representing 100 unique movement series (identified by the record field). Each complete movement series consists of approximately 270 frames. In parallel, we used 4 additional movements as a negative class for training classifier models, including `Straight Line Lunge, Step over Obstacle, Lateral Lunge, and Rise from Sit`. We loaded these movements from separate CSV files, totaling 390 movement series in the negative class.

Given the data context, the first preparatory step involved the automated identification of signal variables based on column labels. The identification strategy classified the variables into three categories:

- Joint angles (8 variables): Identified by the presence of the substring angle in lowercase within the column name, capturing joint rotations such as `shoulderLangle, shoulderRangle, elbowLangle, elbowRangle, hipLangle, hipRangle, kneeLangle, kneeRangle`.
- Three-dimensional coordinates (36 variables): Identified by the suffix X, Y, or Z in the column name, resulting in 12 sensors × 3 coordinate axes (`lshoulderX, rshoulderX, lelbowX, ..., rankleZ`).
- Transverse variables (2 variables): Identified by the presence of "Transv" in the name, capturing transverse rotations - `shoulderLTransv, shoulderRTransv`.

The process ensures that a total of 46 signal variables are submitted to subsequent tests to verify that they were correctly identified.

2.2 Implementation of Generation Techniques

Gaussian noise-based synthesis is performed by selecting an existing real-world motion and adding Gaussian noise correlated with its characteristics. For each signal variable (coordinates, angles, cross-sectional variables), we defined a specific standard deviation: 0.02 for coordinates, 0.05 for angles, and 0.03 for cross-sectional variables. We sampled noise from a normal distribution with a mean of 0 and a standard deviation proportional to the real-world data's standard deviation. We limited the resulting values to the 1st and 99th percentiles of the

[1] https://github.com/aline-pires/TCC.git.

real-world data to ensure they remained within physically feasible ranges. This method preserved the temporal structure and dynamics of the original motion while introducing realistic variability.

For the multivariate time series GAN, we chose to implement it using `PyTorch`. Real-world sequences were resampled to a fixed length of 300 time steps, hip-centered, height-normalized (when available), and global z-score-standardized. The generator received as input a 96-dimensional Gaussian noise vector, concatenated with an explicit temporal embedding, composed of 1D convolutional residual blocks with progressive dilations and 192 intermediate channels. The discriminator was implemented with 1D convolutions and spectral normalization, operating simultaneously on position, velocity, and acceleration.

We conducted the training in two stages. In the first stage, we pre-trained on all available movements for 300 epochs. Subsequently, we applied a specific refinement to the deep squat movement for 1000 epochs. In both stages, we set a batch size of 32 and distinct learning rates for the generator and the discriminator: higher for the generator and slightly lower for the discriminator, to stabilize the adversarial process. For the generator's objective function, we combined adversarial loss, feature matching, and temporal regularization terms, with additional constraints introduced progressively after a warm-up period. As for the final synthetic sequences, these were generated directly by the trained generator and used exclusively in the data augmentation experiments.

2.3 Multidimensional Evaluation Protocol

First, we structured the quality assessment around verifying statistical accuracy, which evaluates the degree to which synthetic data distributions reproduce the statistical characteristics of real data. We applied two metrics: the Kolmogorov-Smirnov (KS) test and the Wasserstein distance.

We applied the KS test exclusively to the 36 three-dimensional coordinate variables, as these are continuous variables suitable for this test. The operational procedure involved: For each variable with coordinates $i \in 1, 2, ..., 36$, we extracted the vectors of real and synthetic values, then applied the KS test to obtain a two-tailed p-value. We defined approval by the criterion $p\text{value} > 0.05$, indicating the absence of statistical evidence of a significant difference between the distributions, generating a table of approvals and rejections based on this criterion.

We applied the Wasserstein distance to the 10 angles and transversal variables (8 angles + 2 transversals), providing a continuous distance metric that captures the multidimensional structure of the distributions.

The protocol we implemented included calculating the Wasserstein distance using the `wasserstein_distance()` function, already implemented in the `SciPy` library, with the standard Euclidean metric, and calculating the amplitude of the real data; it also involved normalizing the distance by the amplitude, allowing comparison between variables of different scales. Finally, we applied the acceptance criterion $W_{\text{norm}} < 0.15$. We chose this criterion of 0.15 to represent a tolerance of up to 15% of the observed range of motion in the real data. Approvals

in this criterion indicate not only the preservation of marginal distributions, but mainly the preservation of structured and biologically significant inter-articular correlations.

We assessed temporal dynamic fidelity using the autocorrelation function (ACF) for each signal variable and each unique motion series (identified by record); thus, we initially grouped the dataset by the record variable. For each record and variable, we extracted the time series and subsequently calculated the autocorrelation function for lags $k = 0, 1, 2, ..., 24$, capturing dependencies for up to 24 time steps. The procedure included a protection against very short series: if $T < k + 1$, the record was ignored. We also employed the function `statsmodels.tsa.stattools.acf()` for this purpose.

After calculating the autocorrelation for all vectors, we established the acceptance criterion as the mean absolute error $\text{MAE}_{\text{ACF}} < 0.10$, reflecting that autocorrelation structures are considered high-fidelity when they diverge by less than 10% on average. Values above this limit indicate that the generation technique substantially altered the temporal memory of the motion.

In addition to evaluating temporal dynamics through the autocorrelation function, we used multivariate DTW to quantify the similarity between complete motion trajectories. The procedure, implemented identically for both synthesis methods (Gaussian noise and GAN), used global z-score normalization based on real data statistics (mean and standard deviation calculated across the 46 signal variables). For each comparison, we extracted the multidimensional temporal matrices $(T_i, 46)$, where T_i represents the number of frames in the i series, applying DTW via `fastdtw()` with Euclidean distance in 46-dimensional space and normalizing the result by the optimal alignment path length.

To establish a common baseline, we calculated a DTW distribution between pairs of distinct real movements (any real movement versus any other real movement), generating a reference for the natural variability observed in real executions of the deep squat. The unified acceptance criterion was $\text{DTW}_{\text{norm}} \leq 2 \times \text{median}(\text{baseline}_{\text{real-real global}})$, allowing synthetic trajectories to exhibit up to twice the variation typically observed between distinct real executions of the same movement. This factor of 2 reflects a conservative tolerance for natural variations while rejecting systematic deviations from expected motor dynamics.

The pairing strategy was such that for each of the 100 generated synthetic series, the DTW was calculated against all 100 real series, systematically recording the lowest DTW distance, a nearest neighbor approach. This approach evaluated how closely each synthetic data trajectory matched the cloud of plausible real trajectories. Thus, both methods generated distributions of 100 minimum DTW values (one per synthetic series), allowing for direct comparison under the same *baseline* and acceptance criterion.

Finally, the evaluation of temporal dynamics concluded with verification of physical plausibility through derivative analysis. For each time series of coordinates and angles, velocity and acceleration were calculated. To be realistic, synthetic data cannot exhibit excessive accelerations (instantaneous changes in velocity) or high-frequency jitter; therefore, the 99th percentile of absolute accel-

eration was computed from real data, and the number of synthetic samples violating this limit was counted. Thus, the established criterion was less than 2% of synthetic samples exceeding the 99th percentile of the real data.

Regarding the third evaluation aspect, clinical plausibility, the method we developed involved defining limits based on the range of motion observed in the real data (1st and 99th percentiles) for each variable. We verified whether all synthetic values remained within these ranges, with relative differences considered acceptable; values completely outside the observable range indicate anatomically impossible movements. The criterion was that 95% of the synthetic samples were within the limits observed in the real data. We tabulated these results, indicating the variable, the upper and lower limits, the minimum and maximum synthetic values, the percentage of conformity, the approval status, and the applied criterion.

Finally, we evaluated the practical utility of synthetic data through data augmentation experiments designed to simulate scenarios of progressively scarce real data. Our experimental protocol was based on the central hypothesis that high-quality synthetic data can improve the performance of predictive models, such as Random Forest and LSTM, trained with limited real samples. To this end, we defined five experimental scenarios corresponding to different levels of availability of real-time series of deep squat movements, considering sets with 10, 20, 30, 50, and 70 positive series. For each scenario, we adopted a comparative protocol under two distinct conditions: **a)** training using exclusively real data; and **b)** training with real data + an equivalent amount of synthetic data (1:1 data augmentation). In all analyses, we kept the test set fixed, comprising 30 positive real series and 117 negative real series, for a total of 147 samples, ensuring a consistent basis for comparing the incremental impact of including synthetic data.

Regarding the predictive models, two complementary approaches were evaluated. The first was based on Random Forest, where the temporal mean, standard deviation, minimum value, maximum value, amplitude, and mean velocity were first extracted for each movement time series and for each of the 46 signal variables, resulting in a total of 276 features per movement, allowing for the capture of both statistical characteristics and compatibility with these machine learning models. Thus, using these features, 100 trees were configured with the Gini impurity criterion and automatic class weighting (`class_weight='balanced'`) to handle class imbalance. To ensure robustness, the experiments were repeated with five distinct seeds, ranging from 100 to 104. In each repetition, random samples of positive real series were selected according to the scarcity scenario, along with a fixed set of negative examples. In synthetic-data scenarios, positive synthetic series were added in quantities equivalent to the real ones. The model was then trained on this augmented set and evaluated on the fixed test set, recording performance metrics.

Furthermore, we implemented an LSTM model to directly explore the temporal dependencies of the series, without the need for prior statistical aggregation. The architecture we employed consisted of a masking layer, followed by an LSTM

layer with 32 memory units, a dropout layer for regularization, and a dense layer with sigmoid activation for binary classification. We trained the model with the Adam optimizer, the binary cross-entropy loss function, and 25 training epochs. We also standardized the time series to a fixed maximum length of 300 time steps via post-padding, and used a fixed threshold of 0.5 for obtaining predictions on the test set, with performance metrics calculated analogously to the Random Forest experiment.

In both approaches, we evaluated the model performance using accuracy, precision, recall, and F1-Score. For each experimental configuration, we aggregated the results via the mean and standard deviation of the metrics across the five seeds, allowing for the quantification of both average performance and the variability associated with the sampling process. We considered confirmed the practical utility of the synthetic data when consistent gains in F1-Score and other metrics were observed in models trained with real data augmented by synthetic data, especially in the most extreme scarcity scenarios (10 and 20 real series). In summary, Table 1 consolidates the predefined acceptance criteria for each evaluation dimension:

Table 1. Consolidated Acceptance Criteria for Synthetic Data Evaluation.

Dimension	Metric	Acceptance Criterion
Statistical Fidelity	Kolmogorov-Smirnov	p-value > 0.05 per coordinate
	Wasserstein Distance	$W \leq 0.15 \times$ amplitude
Temporal Dynamics	Multivariate DTW	$\text{DTW}_{\text{norm}} \leq 2 \times \text{median}(\text{base}_{\text{real-real}})$
	ACF/PACF	RMSE < 0.10 vs. real data
	Velocity/Acceleration	$< 2\%$ of samples outside real P99
Clinical Plausibility	Anatomical Limits	95% within real 1–99 percentiles
Practical Utility	Predictive Classification	Analysis of metric gains

3 Results

The initial analysis automatically identified 46 signal variables in the deep squat data, distributed into three categories: 8 joint angle variables, 36 three-dimensional coordinate variables (X, Y, Z) derived from 12 inertial sensors positioned on the shoulders, elbows, wrists, hips, knees, and ankles, and 2 transverse variables (rotation). This multidimensional structure captures the movement in its entirety, reflecting the simultaneous linear and rotational displacements of multiple joints.

3.1 Statistical and Temporal Quality

The statistical fidelity analysis revealed contrasting patterns between the two generation methods. These were evaluated using the KS test on the 36 coordinates and the Wasserstein distance on the 10 angles/transverse variables.

The KS test, which we applied exclusively to three-dimensional coordinates (X, Y, Z), demonstrated good preservation of Gaussian noise, achieving 78% approval (28/36 coordinates), with low KS statistics (0.0096–0.0111) and p-values consistently above 0.05, as observed in Fig. 1, indicating that the marginal distributions of the coordinates were preserved with high statistical fidelity. Rejections (8/36) were concentrated in the more peripheral parts of the body, probably due to the greater sensitivity of the KS test in low-variance regions. However, for the GAN, there was a strong contrast, with 0% approval (0/36), very high KS statistics, and p-values close to 0 in all coordinates. Therefore, we believe this result indicates that the marginal distributions of the coordinates generated by the GAN diverge significantly from the true ones, completely rejecting the hypothesis of statistical similarity.

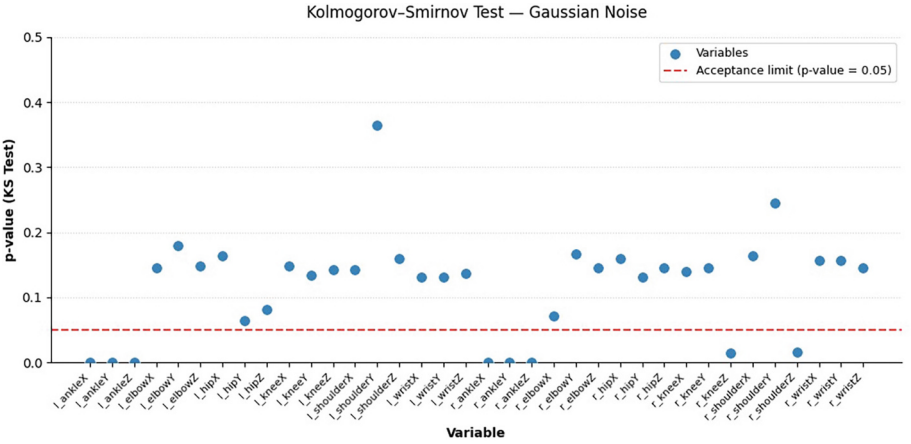
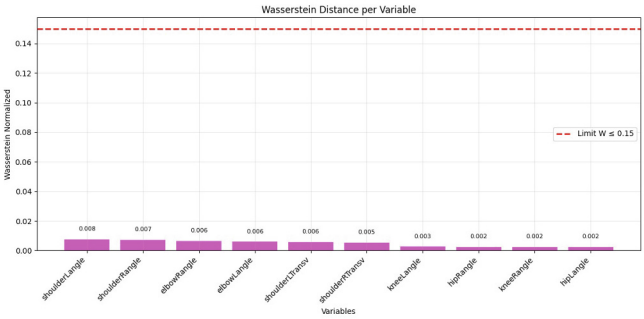
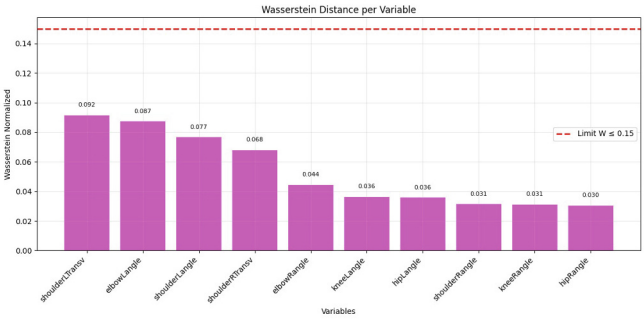


Fig. 1. Kolmogorov-Smirnov Test P-values for Noise Data.

Both methods, however, demonstrated excellent preservation of angular distributions. Gaussian noise obtained 100% approval (10/10) with extremely low normalized Wasserstein distances, while the GAN also reached 100%, all below the 0.15×span threshold. This means that, in this context, the methods demonstrated excellent preservation of angular distributions. It is worth noting that the GAN, despite the limitations observed in spatial coordinates, was capable of generating angles with distributions statistically compatible with the real data; however, the results are not yet as uniform and low as the synthetic data generated by Gaussian noise, as seen in Figs. 2a and 2b.



(a) Noise Data



(b) GAN Data

Fig. 2. Wasserstein Distance Comparison: Noise vs. GAN.

Thus, these results show that, despite limitations in coordinates, both methods preserve essential angular structures, establishing a basis for subsequent dynamic analyses based on angles.

3.2 Temporal Adequacy and Sequential Structure of Synthetic Data

The analysis of the Gaussian noise-based model's capacity to preserve temporal persistence via the autocorrelation function revealed that, globally, this method demonstrated excellent preservation, as seen in Fig. 3. It is observed that the temporal decay pattern of the autocorrelation is consistently preserved, with a large part of the lags remaining within the variability envelope observed in the real data. This behavior indicates that the Gaussian noise-based model maintains, globally, the temporal dependency structure of the original series. This visual evidence reflects the total approvals of 87.0% (40/46 variables) and extremely low autocorrelation error (MAE), especially in Z coordinates, ranging from 0.0004 to 0.0006, and main angles, ranging from 0.002 to 0.028. This performance evidences that temporal correlation was maintained during generation.

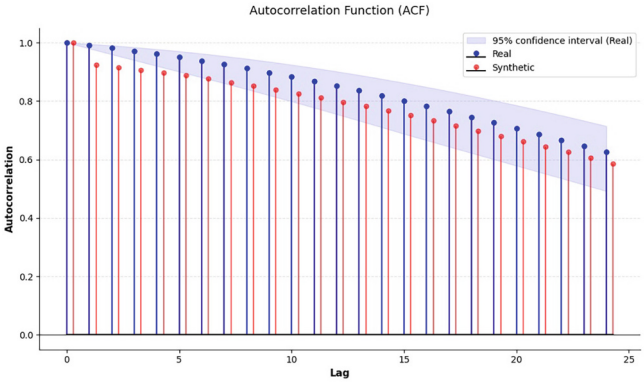


Fig. 3. Global Mean Autocorrelation Function for the Synthetic Set Generated by Gaussian noise vs. Real.

Figures 4 and 5 present two representative variables of the analyzed set: one with minimum autocorrelation error and another with the maximum observed error. In the best case, the synthetic curve totally overlaps the real curve, whereas in the worst case, an abrupt drop in synthetic ACF is observed in the first lags, contrasting with the persistence observed in the real data. This pattern indicates that for certain joints, such as ankles, which do not exhibit much movement during the deep squat, Gaussian noise introduces perturbations that compromise temporal continuity.

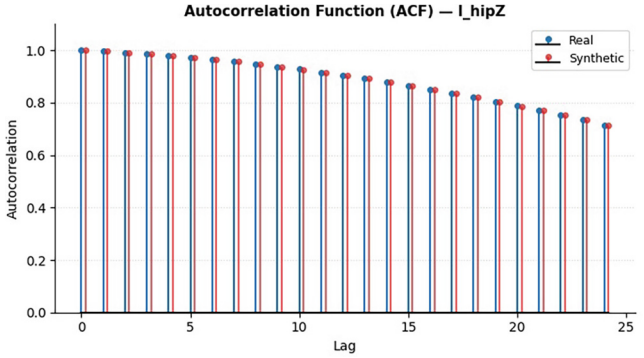


Fig. 4. Representative Variable (Noise): Lowest MAE.

In contrast, Fig. 6 evidences that the global mean ACF of the GAN-generated data presents greater divergence along the lags compared to the real data. Although the general shape of the curve is preserved, a steeper decay in autocorrelation is observed with increasing lag, indicating a partial loss of temporal persistence. This progressive deviation reflects the model's greater difficulty in

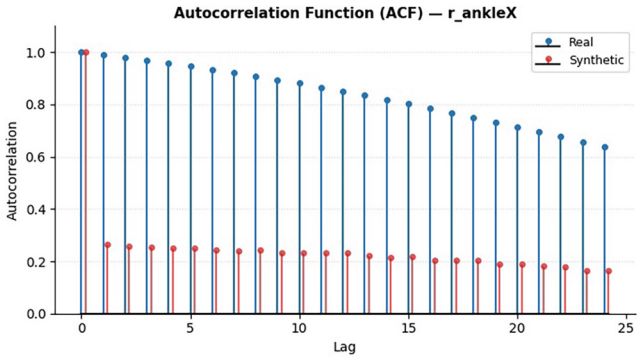


Fig. 5. Representative Variable (Noise): Highest MAE.

maintaining medium- and long-term temporal dependencies. This reflected an approval rate of 67.4% (31/46 variables), with substantially larger ACF errors in the Y and Z coordinates, ranging from 0.125 to 0.410, suggesting greater perturbation in temporal continuity. Failures were more concentrated in the more peripheral body parts, such as the elbows, wrists, and ankles.

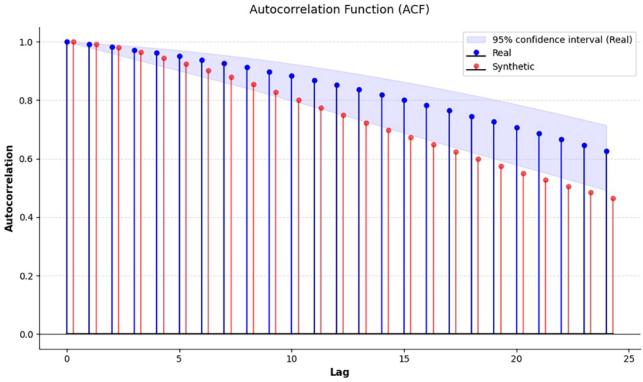
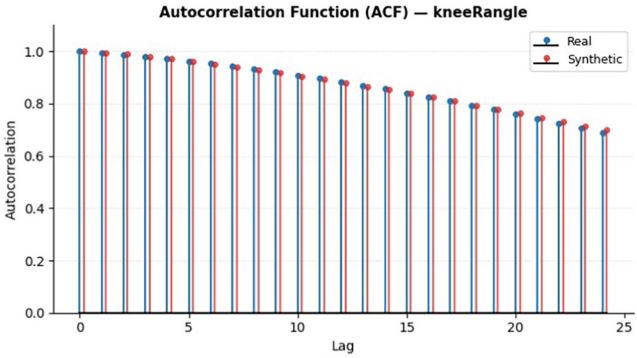


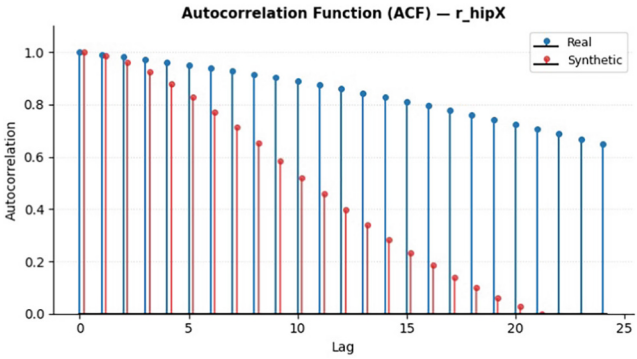
Fig. 6. Global mean autocorrelation function for the synthetic set generated by GAN vs. real.

Figures 7a and 7b present two representative variables of the analyzed set: one with minimum autocorrelation error and another with the maximum observed error. In the best case, corresponding to the variable *kneeRangle*, the synthetic curve almost entirely overlaps the real one, demonstrating that the model can accurately capture the temporal dependency structure in certain angular dynamics. However, in the worst-case scenario, we observed that for the variable *r_hipX*, the synthetic ACF decays abruptly, diverging sharply from the stability and persistence of the real data over the lags. This pattern indicates that the GAN intro-

duces artificial volatility in certain coordinates, failing to preserve the memory of the motion, which justifies the high MAE values detailed in Table 2.



(a) Representative Variable GAN: Lowest MAE



(b) Representative Variable GAN: Highest MAE

Fig. 7. Analysis of GAN Temporal Continuity: Best vs. Worst Performing Variables.

A direct comparison of the methods using the multivariate DTW test shows a difference in performance. The first method obtained approval in 100% of cases (100 out of 100 pairs), with a median distance of 0.1940, so low that it matches the real-real level (median: 6.60; limit: 13.20). This confirms a precise preservation of sequential dynamics. Conversely, the GAN presented opposite performance, with 0% approval and a median DTW of 75.88 – 5.75 times above the acceptable limit. Such a result shows that the GAN failed to capture the temporal structure, to the point that all 100 synthetic series were more dissimilar to the real ones than any pair of real series, completely losing the original multidimensional trajectory.

After analysis, we can see that, visually, this difference is clear in Fig. 8a, where the DTW distances of each synthetic series are plotted in relation to the

Table 2. Summary Table of Temporal Continuity Analysis.

Method	Approved	Rate (%)	Mean MAE	Max MAE
Gaussian Noise	40/46	87.0	0.0427	0.5879
GAN	31/46	67.4	0.1134	0.4101

100 real series. Gaussian noise forms a dense and cohesive pattern close to zero, while GAN creates a vastly dispersed cloud, very far from the origin. Furthermore, Gaussian noise not only meets the statistical criterion, but dominates it by a significant margin, as evidenced in Fig. 8b. Therefore, GAN not only fails in this context, but fails systematically and categorically.

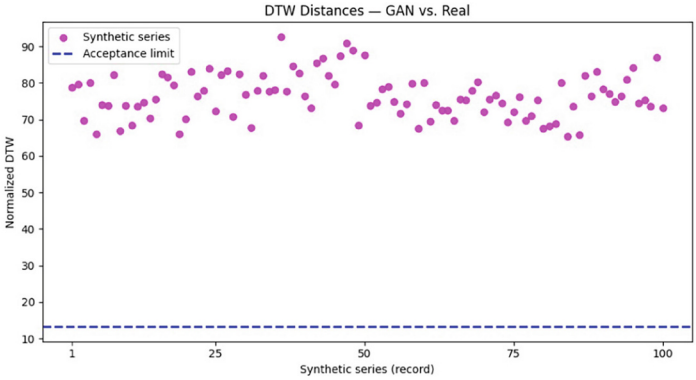
Finally, in the third evaluation of the temporal dimension, we assessed kinematic smoothness, where Gaussian noise performed critically, with 0% approval in the acceleration envelope analysis. Despite perfect DTW and excellent autocorrelation, synthetic accelerations exceeded the established limit (the 99th percentile of the real data) for all 46 variables, with an average violation of 55.66%. However, GAN performed better in the envelope, with 13.04% approval (6/46 variables) and an average violation percentage of 42.53%, indicating that some main trajectories maintained the smoothness of the real motion. This result reveals an interesting phenomenon: Gaussian noise faithfully preserved the ordered temporal structure of the series (DTW 100%, ACF 87%), but produced exaggerated acceleration amplitudes, as if amplifying the velocity peaks, and failed to generate smooth motion.

In summary, the three tests reveal that Gaussian noise maintains temporal structure (DTW 100%, ACF 87%) but fails in kinematics (acceleration 0%), while the GAN performs poorly for the majority but is not catastrophic (DTW 0%, ACF 67%, acceleration 13%). Thus, noise offers ordered sequences with exaggerated peaks, and the GAN generates high-frequency jitter that corrupts dynamics but preserves smoothness in some variables.

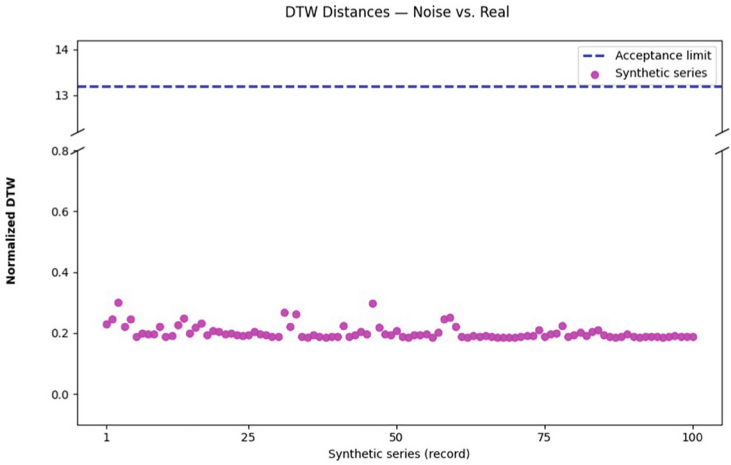
3.3 Clinical Adequacy

The results for verifying anatomical limits for noise data achieved 100% compliance (8 of 8 joint angles). All angle variables remained within the 1st–99th percentiles observed in the real data; in fact, the minimum and maximum of each variable in the synthetic data are exactly the same as those in the real data. However, this perfect agreement is expected, given that during noise generation, a limitation to the 1st–99th percentiles was applied. However, its verification validates that there were no implementation errors and that the synthetic data does not contain anatomically impossible movements.

The GAN data obtained 37.5% compliance (3 of 8 joint angles). The angle variables that remained within the 1st–99th percentiles observed in real data were those of the shoulder and the right side of the hip. However, it is interesting to note that no variable presented catastrophic violations, as all 8 had more than 91% of samples within the limits, as can be seen in Table 3.



(a) Normalized DTW Distances GAN vs. Real



(b) Normalized DTW Distances Noise vs. Real

Fig. 8. Analysis of Normalized DTW Distances Comparing Real Data Against GAN and Noise.

3.4 Performance in Predictive Models

Overall, the results reveal contrasting behaviors between the two generation methods evaluated. While Gaussian noise presented consistent and robust gains, especially in the scarcest scenarios, the GAN demonstrated heterogeneous effects, with isolated benefits and, in some cases, performance degradation.

As shown in Table 4, the addition of synthetic data generated by Gaussian noise proved particularly effective in extreme scarcity scenarios. With only 10 positive real series, the F1-score increased from 0.88 to 0.98, driven by an expressive jump in recall (from 0.79 to 0.97), without loss of precision. Similar results were observed in the scenario with 20 real series, where the inclusion of syn-

Table 3. Analysis of GAN Joint Angles.

Variable	Lower Limit	Upper Limit	Syn. Min	Syn. Max	% Inside
shoulderLangle	114	139	112.57	134.70	99.92
shoulderRangle	110	138	106.16	138.58	99.72
elbowLangle	127	169	124.90	182.69	93.44
elbowRangle	130	165	125.33	180.24	94.89
hipLangle	49	159	22.68	158.17	93.69
hipRangle	48	157	29.43	162.65	95.00
kneeLangle	58	169	36.52	175.39	92.11
kneeRangle	56	167	34.99	179.73	91.42

thetic data led the model to achieve maximum recall and accuracy of 99.6%. As the quantity of real data increased, gains became neutral or showed slight degradation, indicating that the main benefit of Gaussian noise occurs precisely in compensating for severe scarcity, and when sufficient real data are available, synthetic data causes more confusion to the model.

Regarding the results of using synthetic data generated by the GAN presented in Table 4, the Random Forest classifier reveals a less stable pattern. In the most critical scenario (10 real series), including synthetic data resulted in lower F1 and recall scores than training only on real data. From 20 real series onwards, the results became essentially equivalent between the two scenarios, with isolated gains observed only in the intermediate scenario of 30 series, where the model increased the F1-score to 0.99. These findings suggest that although the GAN may contribute in specific contexts, its effect is not systematically beneficial when evaluated in terms of generalization on real data.

We observed similar results in the LSTM analysis, although with some particularities, as seen in Table 5. For Gaussian noise, the addition of synthetic data led to consistent gains in F1-score and recall in all scenarios, again highlighting the regimes of greater scarcity. In the scenario with only 10 real series, the first metric increased from 0.88 to 0.96, accompanied by a substantial improvement in accuracy. This pattern reinforces the idea that, in this context, Gaussian noise acted as an effective regularization mechanism, helping the model learn more robust decision boundaries when real data are limited in deep neural network-based models.

In the case of the GAN, the results with LSTM reinforce the heterogeneity observed in Random Forest, as shown in Table 5. In the most extreme scenario, the inclusion of synthetic data led to a performance drop, with reductions in F1-score and recall. In intermediate scenarios, effects oscillated between modest gains and rare losses, while in scenarios with greater availability of real data, metrics rose slightly. These behaviors suggest that although the GAN can improve metrics in some scenarios, on average, it does not yield consistent practical gains for supervised tasks in this type of model, unlike noise data.

Table 4. Random Forest Performance Comparison: Gaussian Noise vs. GAN Augmentation.

		Gaussian Noise				GAN			
Scenario	Model	F1	Prec	Rec	Acc	F1	Prec	Rec	Acc
10	Real Only	0.876	1.000	0.786	0.956	0.893	1.000	0.813	0.962
10	Real + Syn	0.982	1.000	0.966	0.993	0.854	1.000	0.753	0.950
20	Real Only	0.982	1.000	0.966	0.993	0.972	1.000	0.947	0.989
20	Real + Syn	0.990	0.981	1.000	0.995	0.972	1.000	0.947	0.989
30	Real Only	0.976	0.980	0.973	0.990	0.973	0.987	0.960	0.989
30	Real + Syn	0.981	0.964	1.000	0.992	0.997	0.994	1.000	0.999
50	Real Only	0.984	0.968	1.000	0.993	0.993	0.987	1.000	0.997
50	Real + Syn	0.977	0.956	1.000	0.990	0.993	0.987	1.000	0.997
70	Real Only	0.981	0.963	1.000	0.992	0.984	0.969	1.000	0.993
70	Real + Syn	0.968	0.939	1.000	0.986	0.990	0.981	1.000	0.996

Table 5. LSTM Performance Comparison: Gaussian Noise vs. GAN Augmentation.

		Gaussian Noise				GAN			
Scenario	Model	F1	Prec	Rec	Acc	F1	Prec	Rec	Acc
10	Real Only	0.879	0.964	0.820	0.956	0.900	0.972	0.840	0.962
10	Real + Syn	0.959	0.961	0.960	0.984	0.775	0.948	0.693	0.927
20	Real Only	0.937	0.903	0.980	0.970	0.936	0.917	0.960	0.973
20	Real + Syn	0.977	0.957	1.000	0.990	0.964	1.000	0.933	0.986
30	Real Only	0.911	0.874	0.966	0.958	0.963	0.967	0.960	0.985
30	Real + Syn	0.971	0.943	1.000	0.988	0.953	0.960	0.947	0.981
50	Real Only	0.971	0.944	1.000	0.988	0.977	0.963	0.993	0.990
50	Real + Syn	0.984	0.968	1.000	0.993	0.990	0.987	0.993	0.996
70	Real Only	0.987	0.975	1.000	0.994	0.977	0.956	1.000	0.990
70	Real + Syn	0.990	0.981	1.000	0.996	0.984	0.969	1.000	0.993

4 Conclusion

In this work, we evaluated and compared the quality of synthetic body movement data generated by two distinct approaches: controlled Gaussian noise and GANs, in the context of post-stroke neuromotor rehabilitation. The multidimensional analysis that we conducted encompassed statistical fidelity, temporal dynamics, clinical plausibility, and practical utility, revealing a contrasting scenario between the methods, with important implications for applications in clinical data scarcity scenarios.

The results demonstrated that controlled Gaussian noise achieved superior performance in statistical fidelity of spatial coordinates and in preservation of

temporal structure, with 100% approval in the multivariate DTW test and high autocorrelation preservation (87%). Furthermore, it proved to be clinically plausible, fully respecting the anatomical limits defined from the real data. However, it revealed limitations in kinematic smoothness, with excessive accelerations in all analyzed variables, indicating that while it preserves the temporal sequence, it may introduce many movement peaks.

On the other hand, the GAN generated joint angles with distributions that were statistically compatible with those of real data. However, it failed to replicate the temporal structure (0% DTW agreement) and showed significant differences in the distributions of spatial coordinates. Despite having better preserved kinematic smoothness in some variables, its practical utility proved to be heterogeneous and less consistent, though not catastrophic, with isolated gains and even performance degradation in the extreme scarcity scenario.

From the perspective of utility in classification tasks, Gaussian noise has proven an effective tool for data augmentation in scenarios with limited real-world data, especially when combined with models such as Random Forests and LSTMs. Its gains were more significant and uniform, particularly in contexts of greater scarcity, acting as an effective regularizer and improving the model's generalization. On the other hand, GAN, although it has shown potential in specific contexts, has not delivered consistent benefits, suggesting that its application requires further adjustments and more rigorous validation.

Thus, in this study, we demonstrated that, for the specific motion of the deep squat, controlled Gaussian noise is a more suitable and reliable approach for generating synthetic data in data-scarce scenarios, especially when preserving temporal structure and anatomical plausibility are priorities. GAN, in turn, while promising in theory, still needs substantial refinements for direct application in this domain, particularly in capturing long-term temporal dependencies.

Acknowledgment. The authors acknowledge partial funding from the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Funding Code 001; the National Council for Scientific and Technological Development - Brazil (CNPq) - Funding Codes 442344/2023-6, 444602/2024-0, and 407466/2025-8; the Espírito Santo Research and Innovation Support Foundation (FAPES) – Call No. 02/2026 (Participation in Technical-Scientific Events) – Grant No. 428/2026 and project #2024-9G1WP BPC; and the Minas Gerais Research Funding Foundation (FAPEMIG)

References

1. Brandão, A.F., et al.: Biomechanics sensor node for virtual reality: a wearable device applied to gait recovery for neurofunctional rehabilitation. In: International Conference on Computational Science and Its Applications, pp. 757–770. Springer (2020)
2. Brophy, E., Wang, Z., She, Q., Ward, T.: Generative adversarial networks in time series: a systematic literature review. ACM Comput. Surv. **55**(10), 1–31 (2023)
3. Buess, L., Keicher, M., Navab, N., Maier, A., Tayebi Arasteh, S.: From large language models to multimodal ai: a scoping review on the potential of generative ai in medicine. Biomed. Eng. Lett. **15**(5), 845–863 (2025)

4. Camargo, L.F., Miguel, G.R., Feitosa, J.D.C., Rodrigues, L.G.S., Dias, D.R.C., Brega, J.R.F.: Combining computer vision and artificial intelligence to support the rehabilitation of post-stroke patients. In: International Conference on Computational Science and Its Applications, pp. 415–432. Springer (2025)
5. Dias, D.R., Alvarenga, I.C., Guimarães, M.P., Trevelin, L.C., Castellano, G., Brandão, A.F.: eStreet: virtual reality and wearable devices applied to rehabilitation. In: International Conference on Computational Science and Its Applications, pp. 775–789. Springer (2018)
6. Iglesias, G., Talavera, E., González-Prieto, Á., Mozo, A., Gómez-Canaval, S.: Data augmentation techniques in time series domain: a survey and taxonomy. Neural Comput. Appl. **35**(14), 10123–10145 (2023)
7. Ribeiro, E., et al.: A body tracking framework for neuromotor rehabilitation with multimedia application support. In: International Conference on Computational Science and Its Applications, pp. 345–360. Springer (2025)
8. Rodrigues, L.G.S., et al.: Supervised classification of motor-rehabilitation body movements with RGB cameras and pose tracking data. J. Interact. Syst. (2022)
9. Yoon, J., Jarrett, D., Van der Schaar, M.: Time-series generative adversarial networks. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
10. Zhu, W., et al.: Human motion generation: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **46**(4), 2430–2449 (2023)