UDC: 004.932:004.942:004.895

Z-TRANSCGAN: A Z-SCORE NORMALIZED TRANSFORMER-BASED CONDITIONAL GAN FOR GAIT DATA AUGMENTATION







Yuanyuan Sun[1], Fei Liu[2], Yue Yang[2], Huarong Shao[2], Shodikulova Gulandom Zikriyayevna[3],
Babamuradova Zarrina Bakhtiyarovna[3], Bing Ji[1]
Bing Ji - Corresponding author. E-mail addresses b.ji@sdu.edu.cn (Bing Ji) School of Control Science and Engineering, Shan-dong University, Jinan, 250061, China
1 - School of Control Science and Engineering, Shandong University, Jinan, 250061, China;
2 - Engineering Research Center for Sugar and Sugar Complex, National-Local Joint Engineering Laboratory of Polysaccharide Drugs, Key Laboratory of Carbohydrate and Glycoconjugate Drugs, Shandong Academy of Pharmaceutical Science, Jinan, Shandong 250101, China;
3 - Samarkand State Medical University, Republic of Uzbekistan, Samarkand

Z-TRANSCGAN: ЮРИШ МАЪЛУМОТЛАРИНИ ТЎЛДИРИШ УЧУН Z-БАҲОЛАШНИ НОРМАЛЛАШТИРУВЧИ ЎЗГАРТИРГИЧГА АСОСЛАНГАН ШАРТЛИ ГЕНЕРАТИВ-ТЎПЛОВЧИ ТАРМОҚ

Юанюан Сун[1], Феи Лиу[2], Юэ Янг[2], Ҳуаронг Шао[2], Шодикулова Гуландом Зикрияевна[3],
Бабамурадова Заррина Бахтияровна[3], Бинг Жи[1]
Бинг Жи - мухбир муаллиф. Электрон почта манзиллари b.ji@sdu.edu.cn (Бинг Жи) Назорат фанлари ва муҳандислиги мактаби, Шан-донг университети, Жинан, 250061, Хитой
1 - Назорат фанлари ва муҳандислиги мактаби, Шан-донг университети, Жинан, 250061, Хитой;
2 - Шакар ва шакар комплекси муҳандислик илмий-тадқиқот маркази, Полисахарид дори воситалари миллий-маҳаллий қўшма муҳандислик лабораторияси, Шандун фармацевтика фанлари академияси углевод ва гликоконюгат дори воситалари асосий лабораторияси, Жинан, Шандун 250101, Хитой;
3 - Самарқанд давлат тиббиёт университети, Ўзбекистон Республикаси, Самарқанд ш.

Z-TRANSCGAN: УСЛОВНАЯ ГЕНЕРАТИВНО-СОБИРАТЕЛЬНАЯ СЕТЬ НА ОСНОВЕ ПРЕОБРАЗОВАТЕЛЯ С НОРМАЛИЗАЦИЕЙ Z-ОЦЕНКИ ДЛЯ ДОПОЛНЕНИЯ ДАННЫХ О ПОХОДКЕ

Юаньюан Сун[1], Фэй Лю[2], Юэ Ян[2], Хуаронг Шао[2], Шодикулова Гуландом Зикрияевна[3],
Бабамурадова Заррина Бахтияровна[3], Бин Джи[1]
Бин Джи - корреспондент-автор. E-mail addresses b.ji@sdu.edu.cn (Bing Ji) Факультет науки и техники управления, Университет Шаньдун, Цзинань, 250061, Китай
1 - Факультет науки и техники управления, Университет Шаньдун, Цзинань, 250061, Китай;
2 - Инженерно-исследовательский центр сахара и сахарного комплекса, Национально-местная совместная инженерная лаборатория полисахаридных препаратов, Ключевая лаборатория углеводных и гликоконъюгированных препаратов, Шаньдунская академия фармацевтических наук, Цзинань, Шаньдун 250101, Китай;
3 - Самаркандский государственный медицинский университет, Республика Узбекистан, г. Самарканд

e-mail: info@sammu.uz

Резюме. *Юриш таҳлили касалликларга ташхис қўйиш, шахсни текшириш ва реабилитацияни баҳолаш учун зарур. Бироқ, юришни таҳлил қилишда чуқур ўрганиш ёндашувларининг самарадорлиги қиммат, кўп меҳнат талаб қиладиган ва қатъий махфийлик қоидаларига бўйсунадиган юриш вақт қаторларининг кенг кўламли юқори си-*

*фатли маълумотларини тўплашнинг мураккаблиги билан чекланади. HOA маълумотлар тўпламидаги экспериментал натижалар шуни кўрсатадики, Z-TransCGAN **бу** TTS-CGANдан устун бўлиб, ўртача таснифлаш аниқлигини (ACC) 1,42% га ва эгри чизиқ остидаги майдонни (AUC) 0,85% га оширишга эришади. Ушбу натижалар Z-TransCGAN нинг юриш таҳлили учун маълумотларни тўлдириш стратегияси сифатида самарадорлигини тасдиқлайди, бу эса синтетик маълумотлар ишлаб чиқаришни ҳам, пастки оқим бўйича таснифлаш самарадорлигини ҳам яхшилайди.*

Калит сўзлар: *Чуқур ўрганиш; Шартли генератив рақобатли тармоқлар (CGAN); Юриш таҳлили; Маълумотларни тўлдириш; Трансформер.*

Abstract. *Gait analysis is essential for disease diagnosis, identity verification, and rehabilitation assessment. However, the effectiveness of deep learning approaches in gait analysis is hindered by the difficulty of collecting large-scale, high-quality gait time-series data, which is costly, labor-intensive, and subject to strict privacy regulations. Experimental findings on the HOA dataset indicate that Z-TransCGAN surpasses TTS-CGAN, achieving a 1.42% increase in average classification accuracy (ACC) and a 0.85% increase in the area under the curve (AUC). These results validate the efficacy of Z-TransCGAN as a data augmentation strategy for gait analysis, improving both synthetic data generation and downstream classification performance.*

Keywords: *Deep learning; Conditional Generative Adversarial Networks (CGANs); Gait analysis; Data augmentation; Transformer.*

I. Introduction. Deep learning has been increasingly employed in gait analysis tasks, including disease diagnosis, identity verification, and rehabilitation assessment. However, the effectiveness of these models largely depends on access to large-scale, high-quality training datasets [1]. Unlike image or text data, which are readily available online, gait time-series data must be acquired through specialized sensors, rendering the data collection process expensive, labor-intensive, and constrained by strict privacy regulations [2, 3]. These issues often result in limited sample sizes and significant class imbalance across different gait categories. As a result, models trained on such datasets are prone to overfitting and typically exhibit poor generalization to unseen subjects or underrepresented classes.

To address data scarcity, researchers have traditionally relied on handcrafted augmentation techniques, time domain based methods, e.g. window slicing, temporal shifting, and time scaling [4, 5], simulate walking variations by modifying segment durations or shifting time indices. In the frequency domain, transformations like Fourier and wavelet decomposition [6] manipulate spectral features to introduce rhythm variability. Noise injection [7], often implemented via Gaussian or random noise, is also commonly used to improve model robustness. While these approaches are simple and computationally efficient, they typically suffer from limited generative diversity and may disrupt the complex spatiotemporal dependencies intrinsic to gait sequences. Moreover, these methods often rely on strong prior assumptions and lack adaptability to the nonlinear and dynamic nature of real-world gait signals, which can negatively impact downstream classification performance.

Due to the advantages of Generative Adversarial Networks (GANs) developed in 20148, deep generative models have gained widespread popularity for data augmentation and obtained a seris of exciting findings in different areas [9-11]. By learning to generate realistic samples from limited datasets, GANs offer a promising alternative to manual augmentation by expanding datasets in a data-driven manner.

More recently, GANs have been increasingly utilized in the area of time-series data analysis. A comprehensive survey 12highlights their advantages, including the ability to augment small datasets, generate novel samples, recover corrupted sequences, reduce noise, and even produce privacy-preserving synthetic datasets. Time-series GAN models like C-RNN-GAN [13], RCGAN [14], TimeGAN [15], and SigCWGAN [16] typically adopt recurrent neural network (RNN) architectures due to their temporal modeling capabilities. However, RNN-based GANs often struggle with long-range dependencies and suffer from vanishing gradients, limiting their effectiveness in generating longer or more complex sequences.

To overcome these difficulties, Transformer-based architectures [17], which leverage self-attention mechanisms to model long-range dependencies, have been introduced into generative tasks. Transformer modules have improved performance in various GAN frameworks for both vision and text domains [18, 19], and their theoretical advantages, particularly their ability to model long sequences without recurrent operations, make them attractive for time-series generation.

The challenge in generating gait data lies in maintaining the spatiotemporal dynamics, such as the coordination of joint movements throughout the gait cycle, which requires precise conditional control during the generation process. Conditional GANs (CGANs), first introduced by Mirza et al. [20], extend GANs by integrating class labels or other auxiliary variables into both the generator and discriminator, enabling class-specific synthetic data generation. For instance, TTS-CGAN effectively generates multi-class biological signals by adding a classification head to the discriminator and using labels as conditioning input [20]. This method can effectively preserves the discriminative features related to gait and its association with disease by validating the classification ability of the synthetic data. However, as noted in the original literature, TTS-CGAN has primarily been applied to relatively stationary, low-dimensional, and short-duration sequences, where the generation process is comparatively easier and less noisy.

In contrast, gait signals are typically non-stationary, high-dimensional, and long-duration, which significantly increases generation difficulty. Specifically, these signals are often contaminated with heterogeneous noise, where the magnitude and distribution of noise vary across dimensions. This leads to a key limitation in multivariate gait data generation: the imbalance of feature scales across dimensions exacerbates the model's tendency to overfit

noisy, high-variance features while neglecting meaningful low-variance ones. During training, dimensions with larger amplitudes disproportionately dominate the optimization process, distorting the generator's focus and causing the synthesized sequences to exhibit uneven temporal fluctuations. Ultimately, this degrades both the realism and the class separability of the generated data, making it difficult to produce high-fidelity, label-consistent synthetic gait signals. To address this limitation, Z-TransCGAN is introduced—a Z-score Normalized Transformer-based Conditional GAN tailored for generating multivariate, label-specific gait time-series data. Prior to training, Z-score normalization is employed on the input data, transforming each feature to possess a mean of zero and a standard deviation equal to one. This preprocessing step effectively eliminates discrepancies in feature scales, mitigates the impact of heterogeneous noise, and ensures uniform contribution from each feature during the generation process. Consequently, the model achieves enhanced signal stationarity, leading to improved generation quality.

The core conditioning strategy involves concatenating label embeddings with the generator input and incorporating a classification module into the discriminator. This structure enables the generation of class-specific synthetic sequences while allowing the discriminator to simultaneously distinguish between real and synthetic sequences and classify their categories. To validate the preservation of disease-relevant gait characteristics, the utility of the generated data is assessed through downstream classification performance.

The principal contributions of this study can be summarized as follows:

Z-TransCGAN is proposed as a Transformer-based conditional GAN tailored for generating multi-class labeled gait time-series data.

Z-score normalization is applied to pre-process training data, mitigating inter-dimensional noise interference and enhancing the quality of generated non-stationary signals.

Classification experiments are conducted on the publicly available HOA dataset, combining real and synthetic data to demonstrate the effectiveness of the augmentation-based pipeline in improving classification accuracy.

The structure of this paper is organized as follows: Section II offers a detailed description of the proposed method. Section III presents comprehensive experiments to evaluate the method's performance. Finally, Section IV concludes the study and discusses potential directions for future research.

II. Proposed method.

A. Motivations. In real-world gait analysis scenarios, collecting sufficient high-quality gait data through wearable sensors is often unrealistic due to various practical limitations. Specifically, (1) gait signals are influenced by factors such as sensor placement, individual variability, and environmental disturbances, which result in inconsistent signal quality; (2) Gait time-series data require specialized sensors for collection, which incurs high costs, significant labor, and is subject to stringent privacy regulations. As a result, the limited availability of data significantly hampers the performance and generalization ability of diagnostic models.

Due to their powerful generative capabilities, GAN-based approaches have been extensively employed to alleviate the issue of limited gait data. However, existing GAN-based approaches, including TTS-CGAN, rarely address the intrinsic challenge of heterogeneous noise interference across multivariate gait signals. Without proper normalization, the diverse scales and variances among different dimensions lead to an imbalanced training process, where the generator tends to overfit high-variance dimensions while neglecting subtle but critical temporal dependencies. This imbalance not only amplifies noise artifacts but also disrupts the continuity and realism of generated sequences, resulting in suboptimal data augmentation quality.

To tackle the above issues, a Z-score Normalized TTS-CGAN framework, termed Z-TransCGAN, is proposed. Specifically, z-score normalization is applied prior to training to standardize all input features, ensuring a mean of zero and a variance of one, thereby mitigating the impact of heterogeneous noise and eliminating the influence of scale disparities between dimensions. By ensuring that all features contribute equally during training, Z-TransCGAN enhances the model’s capacity to capture the underlying spatiotemporal patterns and promotes a smoother generation process, leading to more coherent, realistic, and stable synthetic gait signals that better reflect true gait dynamics. The effectiveness of this approach is validated through classification experiments combining real and synthetic data, demonstrating significant improvements in downstream task performance with the augmented dataset.

B. Proposed Z-TransCGAN-based gait classification Method

1) Overall Framework: The proposed framework comprises four main stages, as shown in Figure 1. (1) Preprocessing of Gait Time-Series Data: The original multivariate gait time-series data are initially standardized using z-score normalization to eliminate scale differences across features. After normalization, the data are transformed into a C×H×W image-like format, where C denotes the number of channels and H is set to 1. Additionally, positional encodings and class label embeddings are incorporated into the data to facilitate downstream modeling. (2) Design and Training of Z-TransCGAN: The Z-TransCGAN is constructed to synthesize multi-class gait data. The model includes a generator and a discriminator [22]. The generator is used to produce data corresponding to specific class labels, while the discriminator aims to distinguish between real and synthetic samples and classify the input into the correct category. Through adversarial training, both components are iteratively optimized to enhance the realism and diversity of the generated sequences. (3) Data Assessment and Selection: A similarity-based evaluation mechanism is employed to monitor the quality of the generated sequences. Based on this assessment, the model checkpoint that produces the highest-quality synthetic samples is selected for subsequent augmentation experiments. (4) Data Augmentation and Classification: High-quality synthetic samples are integrated with the original training data to form an augmented dataset. This enriched dataset is then utilized to train the classification model, aiming to enhance its performance on the testing set by improving generalization.

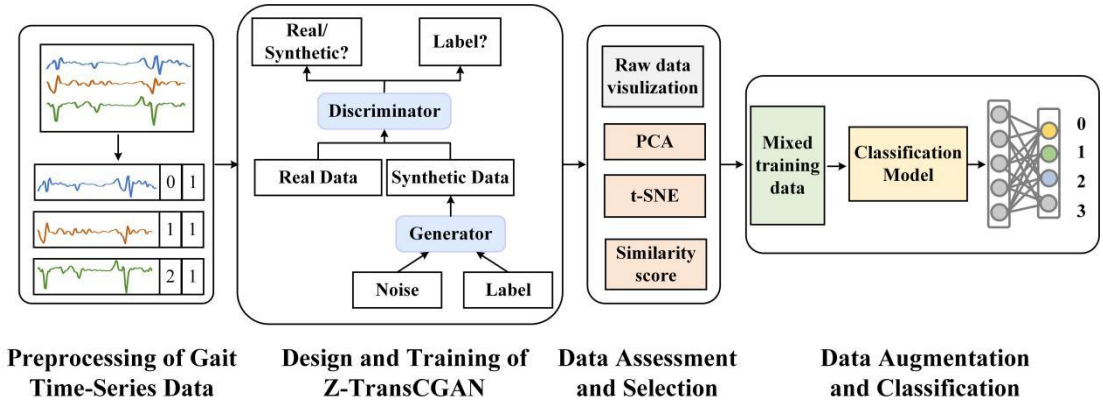

Figure 1: Gait classification algorithm based on Z-TransCGAN enhancement

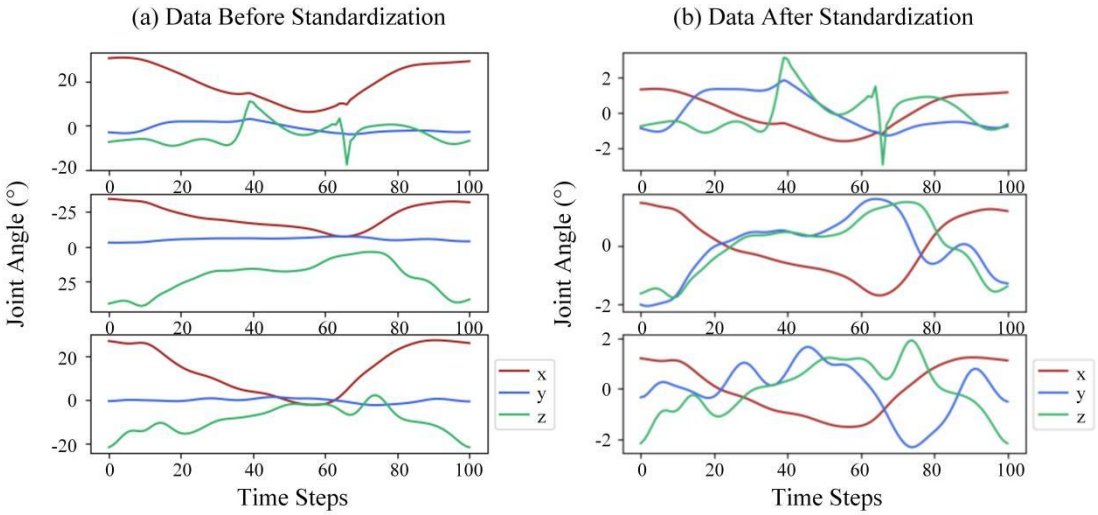

Figure 2: Comparison of data before and after standardization

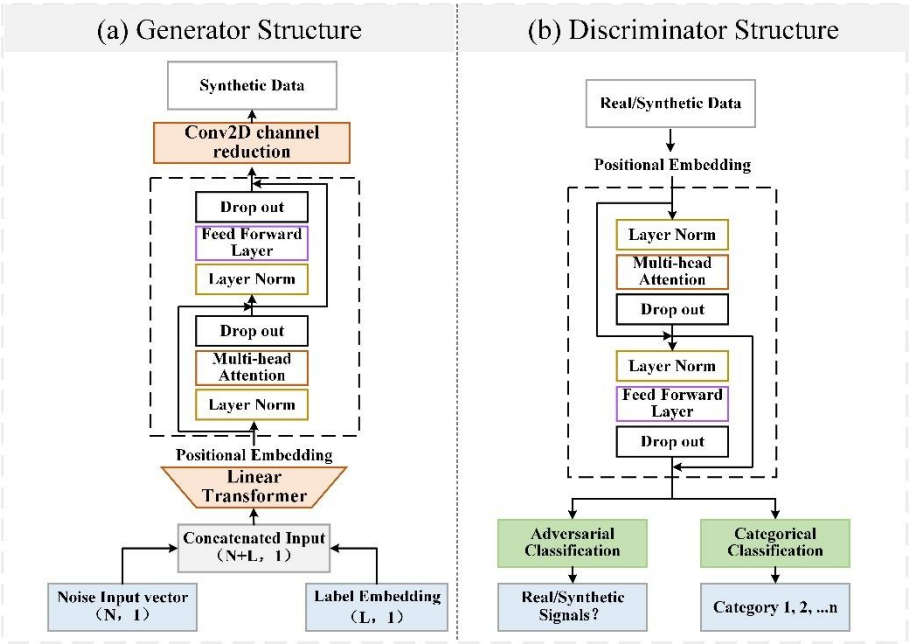

Figure 3: Overall architecture of Z-TransCGAN model based on Transformer

2) Preprocessing of Gait Time-Series Data: In our experiments, the original multivariate gait time-series data, initially with dimensions of (BatchSize, C, W), first undergoes z-score normalization to eliminate feature scale discrepancies and enhance training stability. As shown in Figure 2, the fundamental concept of z-score normalization is to standardize the original data by utilizing its mean and standard deviation, thereby producing a dataset with a mean of zero and a standard deviation of one. Specifically, for a given dataset, the mean (μ) and standard deviation (σ) are calculated as:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (1)$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \quad \textbf{(2)}$$

Based on these, each data point x_i can be standardized using:

$$z_i = \frac{x_i - \mu}{\sigma}, \quad i = 1, 2, ..., n \quad (3)$$

This normalization process not only removes scale differences across features but also improves the stability and convergence efficiency of algorithms.

After normalization, the data is treated as an image with a height of 1 and reshaped into the format C × H × W, where C denotes the number of signal channels (analogous to image color channels like RGB), H is the fixed height (1 for time-series data), and W corresponds to the sequence length (the number of time steps). Thus, each input sequence is represented as a tensor of shape (BatchSize, C, 1, W).

To facilitate model training, the sequence is divided into W/N segments based on a selected patch size N. A learnable positional encoding is then appended to the end of each patch to preserve temporal order information. As a result, the input to the discriminator adopts the shape (BatchSize, C, 1, (W/N) + 1).

3) Design and Training of Z-TransCGAN: The Z-TransCGAN network proposed in this study adopts Transformer encoder architectures for both the generator and the discriminator [21]. Each Transformer encoder is composed of two main components: a multi-head self-attention mechanism and a feedforward multilayer perceptron (MLP) with GELU activation. Layer normalization is applied prior to each component, while dropout layers are inserted after each component to mitigate overfitting. Residual connections are incorporated between components to preserve information flow across the network.

As shown in Figure 3(a), the generator G is designed to synthesize signals based on both a random latent vector z and a target class label c. The label c is randomly assigned, enabling the generator to learn to produce signals corresponding to different categories. Specifically, the generator takes a one-dimensional vector of length N, consisting of uniformly distributed random values in the range (0, 1). The embedded label vector is then concatenated to the end of this input vector. For instance, if the generator receives a 1D vector of shape (100, 1) and a label embedding of shape (10, 1), the final input becomes a (110, 1) vector. This combined input is subsequently transformed into a sequence with the same length and embedding dimension as the real signal.

The sequence is subsequently segmented into multiple patches, with learnable positional encodings are added to each patch to prepare the data for input into the Transformer encoder blocks. To ensure that the output synthetic sequence matches the shape of the real signal, a Conv2D layer with a kernel size of (1, 1) is applied after the Transformer encoder. This convolutional layer preserves the width and height of the sequence while adjusting the number of channels. Specifically, the generator output with shape (BatchSize, Hidden_Dim, 1, Time_Length) is mapped to (BatchSize, Real_Dim, 1, Time_Length), where Real_Dim corresponds to the number of channels in the real data. Through this process, a multivariate time-series signal with the same shape as the real data is generated from the random noise vector.

As shown in Figure 3(b), the discriminator is responsible for determining whether an input signal is real or synthetic, as well as classifying the signal into its respective category. The architecture of the discriminator is inspired by the Vision Transformer (ViT) model [23]. In ViT, an image is evenly divided into patches of equal width and height, each of which is flattened into a vector and projected into an embedding space through a linear layer, resulting in an embedding vector. Since Transformers inherently lack a sense of positional order, learnable positional encodings are added to preserve the spatial relationships between patches.

As described in the 'Preprocessing of Gait Time-Series Data' section, each input time-series is treated as an image with a fixed height of 1, where the temporal steps align with the image's width, and the signal channels correspond to color channels. To enable the use of positional encodings, the width dimension is evenly partitioned into patches while keeping the height unchanged. The resulting embedded vectors are then passed through a standard Transformer encoder, comprising multi-head self-attention and feedforward layers, to generate high-dimensional feature representations that capture long-range dependencies within the pseudo-image sequence. Finally, a fully connected layer performs classification based on these feature representations.

4) Loss Function Design: The objective functions for optimizing the generator G and the discriminator D are defined as follows:

$$L_D = -L_{adv} + \lambda L_{cls}^r \quad \textbf{(4)}$$

$$L_G = -L_{adv} + \lambda L_{cls}^f \quad \textbf{(5)}$$

Here, L_{adv} denotes the adversarial loss, which evaluates the discriminator's ability to distinguish real signals from synthetic ones. L_{cls}^r and L_{cls}^f refer to the classification losses for real and synthetic data, respectively, assessing the discriminator's ability to correctly assign input signals to their respective class labels. The hyperparameter λ controls the relative importance between the classification and adversarial losses.

To encourage the generator to produce synthetic signals that are indistinguishable from real ones, the following adversarial loss is adopted:

$$L_{adv} = E_x[\log D_{adv}(x)] + E_{z,c}[\log(1 - D_{adv}(G(z,c)))] \quad \textbf{(6)}$$

Here, $G(z,c)$ denotes the synthetic signal produced by the generator, conditioned on the random noise vector z and the target class label c, while D_{adv} aims to distinguish real signals from those generated by G. During GAN training, the generator seeks to minimize L_{adv}, whereas the discriminator aims to maximize it. Hence, a negative sign is applied to L_{adv} in Equation (5), allowing

the discriminator to maximize the adversarial objective while the generator minimizes it.

Given a latent noise vector z and a target class label c, the algorithm aims to generate a synthetic output signal that can be correctly classified into the specified class c. To this end, a classification head is integrated into the discriminator, and classification losses are included during the optimization of both the discriminator and the generator. The classification loss comprises two components:

Real classification loss is used to optimize D and is defined as:

$$L_{cls}^{r} = E_{x,c^{'}}[-\log D_{cls}(c^{'} \mid x)] \quad \textbf{(7)}$$

Here, L_{cls}^{r} denotes the classification loss for real signals, where x is a real input signal, and c is its ground-truth label. Minimizing this loss enables D to correctly classify real inputs into their respective original categories.

Fake classification loss is used to optimize G and is defined as:

$$L_{cls}^{f} = E_{z,c}[-\log D_{cls}(c \mid G(z,c))] \quad \textbf{(8)}$$

Here, L_{cls}^{f} represents the classification loss for synthetic signals, where c represents the target class label used during generation. Minimizing this loss encourages G to produce synthetic signals that can are accurately classified into the designated class.

5) Design of Classification Model: This study utilizes the GLIR-GaitNet framework proposed in 24, which comprises two primary components: the GL-JCFE module and the PIR module. The GL-JCFE module comprises three submodules: the local 2D residual module, which captures local features within the three degrees of freedom of the same joint; the global dynamic graph learning module, which extracts global features across joints; and the MCE module, which enhances the complementarity between these two types of features. The PIR module addresses feature imbalance during multi-feature fusion by incorporating SIM loss, thereby improving the interaction between global and local features. Finally, classification results are derived by inputting the fused multi-feature representation into a fully connected layer for triple classification.

C. Gait classification Based on the Z-TransCGAN

In summary, the training procedure of the gait classification algorithm based on the conditional adversarial enhancement consists of four main stages: raw data preprocessing, Z-TransCGAN model training, augmentation weight selection, and mixed-data classification. The detailed training process is shown in Table 1.

III. Experiments and results analysis

A. Dataset

The HOA gait dataset, provided by Dijon University Hospital (France), is publicly available for multi-severity classification research. It can be accessed at https://waikato.github.io/weka-wiki/downloading_weka/, with trial registration on ClinicalTrials.gov (NCT01907503), dated 17 July 2013. This dataset includes gait data from 182 participants, consisting of both healthy individuals and those diagnosed with hip osteoarthritis (HOA).

Table 1: Training process of gait classification algorithm based on Z-TransCGAN enhancement

Input: Input data $D = \{x_i, y_i\}_{i=1,2,3,4}$, total training epochs for the adversarial network E_{GAN}, latent vector z, class labels c, hyperparameter λ, and total training epochs for classification E.
Normalize the input data, reshape it to a C×H×W format, and embed the corresponding class label and positional encodings.
Initialize counter e_{GAN} =0;
While $e_{GAN} \le E_{GAN}$, do:
Generate synthetic data x_{syn} conditioned on random latent vector z and class label c;
Compute the discriminator loss L_D using Equation (4);
Update the discriminator parameters;
Compute the generator loss L_G using Equation (5);
Update the generator parameters;
$e_{GAN} = e_{GAN} + 1$;
Initialize counter e=0;
While $e \le E$, do:
Augment the training dataset with generated data based on selected augmentation weights to form a mixed dataset $\{x_i, x_i^{GAN}, y_i\}_{i=1,2,3,4}$;
Update the parameters of the gait classification model;
$e = e + 1$;

Table 2: HOA Dataset description, including the category name and the size of samples

Class	Category	Numbers
0	Level 0	80
1	Level 2	18
2	Level 3	47
3	Level 4	37

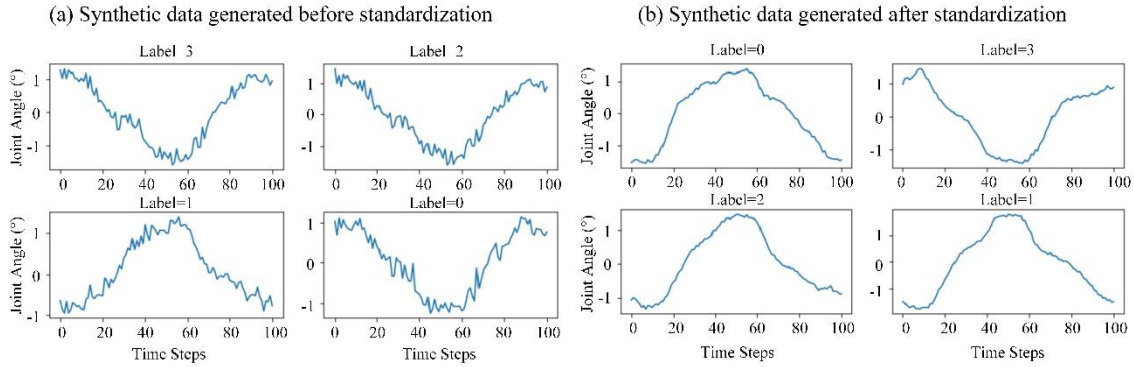

Figure 4: Comparison of synthetic data on a single X-axis before and after standardization

For consistency, ten gait samples were collected per participant. Data were recorded using eight optical cameras (100 Hz) and two force plates (1000 Hz) while subjects walked along a 6-meter track [25]. The raw motion data were segmented into gait cycles and resampled to 101 time points. Angular velocity signals from six bilateral joints—ankles, knees, and hips—were used to create 18-dimensional feature vectors for each frame, totaling 101 frames per sample, as detailed in Table 2. Disease severity was categorized into four levels: 0 (asymptomatic), 2, 3, and 4, reflecting increasing symptom severity.

B. Experimental Setup

1) Implementation Details: The proposed framework is implemented using the PyTorch library and trained on a system equipped with an NVIDIA RTX 3090 GPU. The model is trained for 150 epochs with a batch size of 8. Optimization is performed using the Stochastic Gradient Descent (SGD) algorithm, with a momentum of 0.9. The initial learning rate is set to 0.001 and is reduced to 0.0001 after 70 epochs.

2) Evaluation Metrics: To provide a comprehensive evaluation of the model's performance, two metrics are utilized: Accuracy (ACC) and Area Under the Curve (AUC). Accuracy is calculated as:

$$Accuracy = \frac{\sum_{i=1}^{4} \frac{TP + TN}{TP + TN + FP + FN}}{4} \quad \textbf{(9)}$$

Here, TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively.

The AUC, which calculates the area beneath the Receiver Operating Characteristic (ROC) curve, is used to investigate the capacity of distinguishing between classes. A higher value of AUC represents the better model performance.

3) Cross-validation: Given the limited size of the dataset, five-fold cross-validation is employed to ensure reliable evaluation. The dataset is divided into five stratified folds, maintaining the class distribution within each fold. In each iteration, four folds are used for training and the remaining one for validation, with each fold serving as the validation set exactly once. Additionally, to account for the fact that each subject contributes multiple gait samples, Stratified Group K-Fold cross-validation is applied 26. This method ensures that data from a single subject are not simultaneously included in both the training and validation sets, effectively preventing data leakage and offering a more realistic evaluation of the model's generalization performance in subject-independent scenarios.

C. Evaluation of Generated Samples

To assess the quality of synthetic time-series data produced by the adversarial model, this study employs four complementary methods: Raw data visualization, Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (t-SNE), and similarity score computation.

4) Raw data visualization: Visualizing augmented time-series data provides an intuitive means to assess the fidelity of synthetic sequences generated through data augmentation. Theoretically, for synthetic data to be considered reliable, it should exhibit statistical consistency with real data across critical kinematic dimensions—such as amplitude fluctuations, base frequency cycles, and time-frequency evolution characteristics. This ensures that the synthetic data maintains morphological similarity to real signals, thereby confirming its practical usability for gait analysis and offering a straightforward basis for evaluating the effectiveness of augmentation strategies.

As depicted in Figure 4, synthetic data generated before and after standardization is compared with real data. The corresponding real data is shown as the red curve in Figure 2(b). Both pre- and post-standardization synthetic data closely align with real data in terms of amplitude, periodicity, and overall trend, confirming their visual similarity. However, the pre-standardization synthetic data exhibit noticeable noise and irregularities, resulting in less smooth curves and greater susceptibility to heteroscedastic noise interference between multidimensional sequences. In contrast, the post-standardization synthetic data demonstrate improved smoothness and reduced noise, highlighting the importance of standardization in improving the quality and reliability of synthetic data for gait analysis applications.

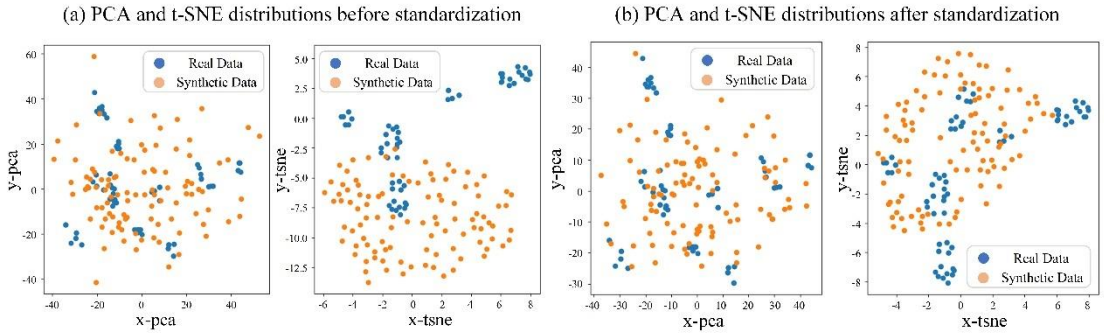

Figure 5: The distribution of real and synthetic data after dimensionality reduction using PCA and t-SNE methods before and after standardization

Table 3: Similarity score between real data and synthetic data before and after standardization

Standardization	Similarity Score			
	Label=0	Label=1	Label=2	Label=3
Without	0.979	0.762	0.867	0.902
With	0.993	0.895	0.976	0.924

Visualizations with PCA and t-SNE: To assess the distribution of synthetic data in the feature space and its similarity to real data, PCA and t-SNE are used for dimensionality reduction. PCA identifies the principal components accounting for the largest variance in the data and projects high-dimensional sequences into a lower-dimensional space. Examining the PCA scatterplots allows assessment of whether the synthetic data extends the original distribution and fills gaps between clusters. In contrast, t-SNE excels at visualizing local structures, helping us determine if the augmented data forms clusters similar to those of the real data or introduces new groupings.

As shown in Figure 5, the distributions of synthetic and real data before and after standardization were compared using PCA and t-SNE. The PCA plots reveal significant overlap between both standardized and non-standardized synthetic data and real data, indicating that the synthetic data captures the overall distribution and clustering structure. This overlap suggests statistical consistency with the real data and effective simulation of its global structure. Moreover, separation beyond the overlap indicates that the synthetic data introduces additional variability, enhancing dataset diversity and reducing overfitting risk. In the t-SNE plots, standardized synthetic data exhibits closer alignment with real data, forming well-defined clusters. This indicates that standardization improves the local structural fidelity of the generated data, better capturing the intricate relationships in the real data.

These observations confirm that standardization enhances both the global and local consistency of synthetic data, supporting its effectiveness as a high-quality augmentation method for gait analysis.

Similarity score computation: To quantitatively evaluate the similarity between synthetic and real gait sequences, the similarity score is defined using the average cosine similarity of feature vectors. For each sequence, seven statistical features—median, mean, standard deviation, variance, root mean square, maximum, and minimum are extracted—resulting in a 7-dimensional feature vector:

$$f =< feature_1, feature_2, ..., feature_m > \quad \textbf{(10)}$$

The cosine similarity between each pair of real and synthetic feature vectors is computed as follows:

$$\cos\_ sim_{ab} = \frac{f_R \cdot f_G}{\|f_R\| \|f_G\|} = \frac{\sum_{i=1}^{m} f_{Ri} f_{Gi}}{\sqrt{\sum_{i=1}^{m} f_{Ri}^2} \sqrt{\sum_{i=1}^{m} f_{Gi}^2}} \quad \textbf{(11)}$$

where f_R and f_G denote the feature vectors of real and synthetic signals, respectively, each of length m. The average cosine similarity is then calculated across all pairs of feature vectors belonging to the same class, providing a measure of overall similarity:

$$\text{avg}\_\cos\_ sim = \frac{1}{n} \sum_{i=1}^{n} \cos\_ sim_i \quad \textbf{(12)}$$

In our experiments, 30 synthetic samples per class were trained, and the cosine similarity scores were computed before and after standardization. The results, presented in Table 3, indicate that for each hip osteoarthritis (HOA) severity level, the synthetic samples generated by the standardized model exhibit higher average cosine similarity to real samples compared to those generated without standardization.

In summary, these similarity analyses of augmented time-series data suggest that incorporating multi-dimensional data standardization into the synthesis process enhances the fidelity of synthetic gait data, reducing noise and improving its alignment with real data.

Analysis of Enhanced Classification Algorithm Results. The impact of data augmentation on classification performance was evaluated by training models using a combination of real and synthetic data from the HOA dataset. The dataset is categorized into four classes based on severity levels: 80 samples from the control group (class 0), 18 samples with mild severity (class 1), 47 samples with moderate severity (class 2), and 37 samples with severe symptoms (class 3). To address the underrepresentation of moderate severity samples (class 2), a higher augmentation factor was applied specifically to this class. The experimental design involved constructing a new training set by merging the original samples with augmented data, and then comparing the results with models trained solely on the original dataset.

Table 4: The impact of different enhancement schemes on the accuracy of each fold

Enhancement schemes	Cross validation of ACC (%) for each fold					Average ACC (%)	Average AUC (%)
	Fold1	Fold 2	Fold 3	Fold 4	Fold 5		
No Augment	83.24	87.84	90.86	86.00	85.56	86.70	79.58
0.5 times	85.95	87.71	86.39	85.23	89.97	87.05	79.53
1 times	87.16	88.11	88.89	85.97	90.83	88.21	80.47
1.5 times	86.78	88.10	89.02	82.37	88.34	86.922	79.45

* On the basis of the original enhancement factor, the second type of data has been increased by an additional 0.5 times.

Table 5: Performance of various enhancement methods on the HOA dataset

Method	Cross validation of ACC (%) for each fold					Average ACC (%)	Average AUC (%)
	Fold1	Fold 2	Fold 3	Fold 4	Fold 5		
No Augment	83.24	87.84	90.86	86.00	85.56	86.70	79.58
MTS-CGAN	86.02	85.39	84.24	81.16	85.23	84.41	76.56
TTS-CGAN	86.70	87.91	85.86	86.72	86.75	86.79	79.62
Z-TransCGAN	87.16	88.11	88.89	85.97	90.83	88.21	80.47

To align with the cross-validation training scheme, augmentation models were trained independently for each fold, generating synthetic samples based on the corresponding training set. These augmented models were subsequently evaluated on their respective test sets to rigorously assess the effectiveness of the proposed data augmentation strategy. Table 4 presents the classification ACC across five folds under different augmentation strategies, along with the average ACC and AUC scores.

As presented in Table 4, augmenting the number of samples in classes 0, 1, and 3 by twofold and increasing class 2 by 1.5 times resulted in the highest average ACC of 88.21% and AUC of 80.47%. Compared to the baseline model without augmentation, this represents improvements of 1.51% in ACC and 0.89% in AUC. These results indicate that the proposed data augmentation approach effectively alleviates the challenges of limited and imbalanced data, thereby enhancing the overall classification performance.

E. Comparative Experiment Results and Analysis

To validate the effectiveness of the proposed gait data augmentation method, comparative experiments are conducted to assess the classification performance after generating augmented gait sequences using various time-series augmentation methods. To ensure a fair comparison, the same raw training and testing samples, along with identical classification models, are used across all algorithms. The detailed classification performance comparison results are presented in Tables 5.

1) No Augment: This baseline directly trains and tests the model on the original, non-augmented dataset, without applying any data augmentation techniques.

2) MTS-CGAN: Multivariate Time Series Conditional Generative Adversarial Network 27is a Transformer-based generative model that adjusts the generator's output using encoded contextual information. It enables a single model to learn the mixed distribution of data from multiple classes, allowing for realistic modeling of multivariate time series under various conditions.

3) TTS-CGAN: The Transformer Time-Series Conditional GAN 21, as introduced earlier, corresponds to the conditional adversarial augmentation network but omits the normalization preprocessing step. While this approach facilitates the generation of synthetic time-series data, it exhibits limitations in producing high-quality sequences for longer, non-stationary signals.

The results of the comparative experiments indicate that the proposed improved method consistently outperforms other approaches. Specifically, when MTS-CGAN was used for data augmentation, the accuracy of cross-validation decreased to varying extents in all folds. This indicates that the synthetic data produced by MTS-CGAN exhibited lower quality, and mixing this data with the original training set hindered the model from learning effective classification weights, ultimately leading to performance degradation. In contrast, both the TTS-CGAN and the improved Z-TransCGAN methods resulted in significant performance improvements. Notably, the incorporation of the standardization process further enhanced the quality of the synthetic data. Compared to TTS-CGAN, this improvement resulted in a 1.42% improvement in average ACC and a 0.85% gain in average AUC, effectively optimizing the general time-series augmentation algorithm for gait sequence synthesis.

Figure 6 displays the confusion matrix and ROC curves for the enhanced gait classification algorithm based on Z-TransCGAN, illustrating disease grading results. In Figure 6(a), the confusion matrix is presented, showcasing the model's performance across four categories. Each cell represents the normalized probability of a specific combination of predicted and true labels, offering valuable insights into the model's class-wise performance. Figure 6(b) illustrates the ROC curves, which demonstrate the trade-off between the true positive rate and the false positive rate at various thresholds. The AUC quantifies the model's discriminative ability, with values closer to 1 indicating better performance. The model performs best in class 0 (AUC = 0.9686), effectively diagnosing whether a subject has HOA. However, the diagnostic performance for class 1, representing mild patients, is relatively ambiguous (AUC = 0.6253), likely due to the less pronounced gait abnormalities in this group.

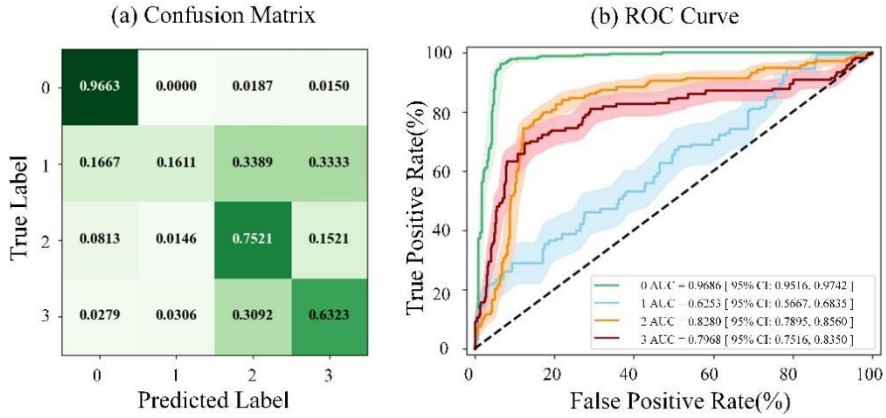

Figure 6: Confusion matrix and ROC curves for Z-TransCGAN-based gait classification

IV. Conclusion. This study presents Z-TransCGAN, a Transformer-based conditional GAN enhanced with Z-score normalization, specifically designed to generate high-quality multivariate gait time-series data for disease diagnosis and severity grading. By standardizing input features, Z-TransCGAN mitigates scale disparities and reduces heterogeneous noise interference, leading to improved model performance. Experimental validation on the HOA dataset demonstrates that Z-TransCGAN outperforms TTS-CGAN, achieving a 1.42% increase in average ACC and a 0.85% increase in average AUC. These findings underscore the effectiveness of Z-TransCGAN as a data augmentation method, enhancing both synthetic data generation and downstream classification tasks in gait analysis.

From a clinical perspective, the ability to generate synthetic gait data that accurately reflects disease-specific patterns is crucial for training robust diagnostic models, especially when real-world data is scarce. However, it is important to note that gait signals inherently contain individual-specific information, such as personal walking styles, which can inadvertently be learned by generative models. This unintended memorization poses privacy concerns and may affect the generalization ability of models trained on synthetic data.

To tackle this issue, future research should prioritize the development of methods that can disentangle disease-related features from individual-specific attributes within gait data. Approaches such as adversarial training, differential privacy, or attribute editing frameworks could be explored to anonymize synthetic gait sequences while preserving their diagnostic utility.

Literature:

[1] Tran L, Choi D. Data augmentation for inertial sensor-based gait deep neural network[J]. IEEE Access, 2020, 8: 12364-12378.

[2] Li J B, Farrell J W, Valles D. Data Augmentation for Classifying Multiple Sclerosis Severity through Inertial Measurement Unit-Based Gait Analysis[C]//2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2024: 6443-6450.

[3] Duong H T, Nguyen-Thi T A. A review: preprocessing techniques and data augmentation for sentiment analysis[J]. Computational Social Networks, 2021, 8(1): 1.

[4] Cao P, Li X, Mao K, et al. A novel data augmentation method to enhance deep neural networks for detection of atrial fibrillation[J]. Biomedical Signal Processing and Control, 2020, 56: 101675.

[5] Pan Q, Li X, Fang L. Data augmentation for deep learning-based ECG analysis[M]//Feature engineering and computational intelligence in ECG monitoring. Singapore: Springer Singapore, 2020: 91-111.

[6] Liu B, Zhang Z, Cui R. Efficient time series augmentation methods[C]//2020 13th international congress on image and signal processing, BioMedical engineering and informatics (CISP-BMEI). IEEE, 2020: 1004-1009.

[7] Iglesias G, Talavera E, González-Prieto Á, et al. Data augmentation techniques in time series domain: a survey and taxonomy[J]. Neural Computing and Applications, 2023, 35(14): 10123-10145.

[8] Creswell A, White T, Dumoulin V, et al. Generative adversarial networks: An overview[J]. IEEE signal processing magazine, 2018, 35(1): 53-65.

[9] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4681-4690.

[10] Bousmalis K, Silberman N, Dohan D, et al. Unsupervised pixel-level domain adaptation with generative adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 3722-3731.

[11] Zhang H, Xu T, Li H, et al. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 5907-5915.

[12] Brophy E, Wang Z, She Q, et al. Generative adversarial networks in time series: A survey and taxonomy[J]. arxiv preprint arxiv:2107.11098, 2021.

[13] Mogren O. C-RNN-GAN: Continuous recurrent neural networks with adversarial training[J]. arxiv preprint arxiv:1611.09904, 2016.

[14] Esteban C, Hyland S L, Rätsch G. Real-valued (medical) time series generation with recurrent conditional gans[J]. arxiv preprint arxiv:1706.02633, 2017.

[15] Yoon J, Jarrett D, Van der Schaar M. Time-series generative adversarial networks[J]. Advances in neural information processing systems, 2019, 32.

[16] Ni H, Szpruch L, Wiese M, et al. Conditional sig-wasserstein gans for time series generation. arxiv[J]. arxiv preprint arxiv:2006.05421, 2020.

[17] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
[18] Jiang Y, Chang S, Wang Z. Transgan: Two pure transformers can make one strong gan, and that can scale up[J]. Advances in Neural Information Processing Systems, 2021, 34: 14745-14758.
[19] Diao S, Shen X, Shum K, et al. TILGAN: transformer-based implicit latent GAN for diverse and coherent text generation[C]//Findings of the Association for Computational linguistics: ACL-IJCNLP 2021. 2021: 4844-4858.
[20] Mirza M, Osindero S. Conditional generative adversarial nets[J]. arxiv preprint arxiv:1411.1784, 2014.
[21] Li X, Ngu A H H, Metsis V. Tts-cgan: A transformer time-series conditional gan for biosignal data augmentation[J]. arxiv preprint arxiv:2206.13676, 2022.
[22] LI X, METSIS V, WANG H, et al. Tts-gan: A transformer-based time-series generative adversarial network[C]//International conference on artificial intelligence in medicine. Cham: Springer International Publishing, 2022: 133-143.
[23] LIANG Z, XU Y, HONG Y, et al. A Survey of Multimodel Large Language Models[C]//Proceedings of the 3rd International Conference on Computer, Artificial Intelligence and Control Engineering. 2024: 405-409.
[24] Shandong University Gait analysis disease diagnosis method and system based on feature interaction rebalancing: 2024,11665713 X [P]. February 18, 2025 (under application)
[25] Bertaux A, Gueugnon M, Moissenet F, et al. Gait analysis dataset of healthy volunteers and patients before and 6 months after total hip arthroplasty[J]. Scientific Data, 2022, 9(1): 399.
[26] Pan X, Yu Z, Yang Z. A multi-scale convolutional neural network combined with a portable near-infrared spectrometer for the rapid, non-destructive identification of wood species[J]. Forests, 2024, 15(3): 556.
[27] Madane A, Dilmi M, Forest F, et al. Transformer-based conditional generative adversarial network for multivariate time series generation[J]. arxiv preprint arxiv:2210.02089, 2022.

Z-TRANSCGAN: УСЛОВНАЯ ГЕНЕРАТИВНО-СОБИРАТЕЛЬНАЯ СЕТЬ НА ОСНОВЕ ПРЕОБРАЗОВАТЕЛЯ С НОРМАЛИЗАЦИЕЙ Z-ОЦЕНКИ ДЛЯ ДОПОЛНЕНИЯ ДАННЫХ О ПОХОДКЕ

Юаньюань Сун, Фэй Лю, Юэ Ян, Хуаронг Шао, Шодикулова Г.З., Бабамурадова З.Б., Бин Джи

Резюме. *Анализ походки необходим для диагностики заболеваний, проверки личности и оценки реабилитации. Однако эффективность подходов глубокого обучения в анализе походки сдерживается сложностью сбора крупномасштабных высококачественных данных временных рядов походки, что является дорогостоящим, трудоемким и подчиняется строгим правилам конфиденциальности. Экспериментальные результаты на наборе данных HOA показывают, что Z-TransCGAN превосходит TTS-CGAN, достигая 1,42% увеличения средней точности классификации (ACC) и 0,85% увеличения площади под кривой (AUC). Эти результаты подтверждают эффективность Z-TransCGAN как стратегии дополнения данных для анализа походки, улучшая как генерацию синтетических данных, так и производительность классификации ниже по потоку.*

Ключевые слова: *Глубокое обучение; Условные генеративные состязательные сети (CGAN); Анализ походки; Дополнение данных; Трансформер.*