



Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)



A Survey of Time Series Data Augmentation Techniques

Muntaha Abdel Hamza Hamid¹, Lamia Abed Noor Muhammed²

¹Department of Computer Science and IT, University of Al-Qadisiyah, Diwaniyah, Iraq, Email: master.student245@qu.edu.iq

Department of Computer Science and IT, University of Al-Qadisiyah, Diwaniyah, Iraq, Email: lamia.abed@qu.edu.iq

ARTICLE INFO

Article history:

Received: 03 /03/2026

Revised form: 02 /05/2026

Accepted : 03 /05/2026

Available online: 30 /09/2026

Keywords:

Time series Data, Augmentation

ABSTRACT

Machine learning modeling require for a large amount of training data. Where more training data example results more robust models, but it is not training data can be available in suitable quantity. Also, in some situations, the quality of data that have rich information is leak. So to deal with these problem, augmentation is one of important and common solutions. Augmentation is a method for creating new artificial examples from the available data examples and added to data set. According to its importance, different methods have been proposed to perform augmentation. There methods vary according to the data types. Time series is one of data types that require more for augmentation because this type of data is not available in the suitable quantity or quality. Different time series data augmentation methods have been suggested, from traditional methods that use statistics tool, until the methods based deep learning. This paper would be present a review of these methods.

MSC..

<https://doi.org/10.29304/jqcm.2026.18.32697>

1. Introduction

Data augmentation is an augmentation procedure that seeks to create novel artificial information based on the existing information so as to enhance the performance of a machine learning model by augmenting the size of the information or diversifying its patterns.

Augmentation has become an essential technique in machine learning, particularly in states where the availability of high-quality data is limited. In many real-world applications, collecting large-scale and well-balanced datasets is challenging due to factors such as cost, privacy constraints, or the rarity of certain events. As a result, models trained on insufficient data often suffer from overfitting and poor generalization when applied to unseen samples. In the majority of cases, augmentation is applied to expand the size of the samples utilized in training, i.e., the size of the dataset grows after the augmentation process, which aids the models to observe more examples and prevent the overfitting issue, which appears due to the insufficiency of samples [1], [2]. Nonetheless, this is not the only aim of the goal, because many of the modern approaches target augmentation by increasing the diversity of temporal patterns, by adding variation in the form, frequency spectrum, or the order of events, thus providing the model with

*Corresponding author: Muntaha Abdel Hamza Hamid

Email addresses: master.student245@qu.edu.iq

Communicated by 'sub etitor'

a greater generalization capacity [3], [4]. It is important to note that data augmentation is not only concerned with increasing the number of training samples, but also with enhancing the diversity of temporal patterns [4]. Traditional techniques such as jittering and scaling mainly apply simple transformations, which may enlarge the dataset but often fail to capture the complex dynamics of time series data [3]. To overcome these limitations, more advanced techniques developed deep learning approaches, particularly Generative Adversarial Networks (GANs), which are capable of learning data distributions and generating more realistic and diverse samples [5]. Conditional variants further improve this process by guiding generation using additional information such as class labels [6]. However, the quality of generated data remains a critical challenge. In this context, clustering techniques provide an effective solution by grouping similar patterns and enabling the selection of more representative and consistent samples [7]. This allows augmentation to be applied within more homogeneous subsets of the data, improving both the diversity and reliability of the generated samples. In addition, transformation-based methods that preserve temporal structure while introducing controlled variations further support this objective [8]. Overall, combining augmentation with clustering leads to better data quality and improved generalization, as it balances diversity with structural consistency in time series data [9].

In general, the benefits of augmentation can be summarized in three main areas: Increasing data size and improving training stability [10]. Expanding the time mode space and enhancing the generalizability of models [11]. Improving model performance and accuracy, especially in cases of scarce or unbalanced data [12].

The processing time series data and augmentation methods in the literature of research are categorized into three broad groups. The first one is the traditional methods that are based on the straightforward signal manipulations, i.e. cropping, rearranging, or noise addition. These are simple techniques which can be implemented and used without any learning models and which usually do not ruin the exact temporal structure of the series. The second type is comprised of the Machine Learning techniques which are based on models that can analyze a temporal pattern using algorithms like SVM or random forests. They have the advantage of augmented data but are restricted in the situation of the complex distribution changes. The third category is Deep Learning techniques, which are based on more sophisticated models, including generative networks and spectral transformation-based techniques, to create new more diverse and realistic data. Of these contemporary approaches, the STAug approach has been regarded as one of the most salient deep learning based approaches, where both time and frequency augmentation are used to preserve time coherence and enhance diversity [13]. The approach is based on the analysis of the series with the help of Empirical Mode Decomposition in order to create sub-components that are recombinable with random weights, and the linear mixing of the series is performed to obtain new data. The approach has demonstrated the capability of enhancing the performance of prediction models through the development of more varied samples without compromising the dynamics of the original signal as indicated in one study [14].

2. Method and Material

2.1. Time Series Data

Time series data can be described as an observation or a collection of observations or values that are measured in time series, where every new item is related to or directly or indirectly affected by the previous item [15]. This information is not just defined by the value, but also by the time-varying pattern by which the signal varies, e.g. long-term patterns, seasonality, variations, and short- or long-term cycles. Time series data have a wide variety of applications, such as industrial systems, consumption prediction, sensor systems, medical care, and financial markets [5], [16], [17].

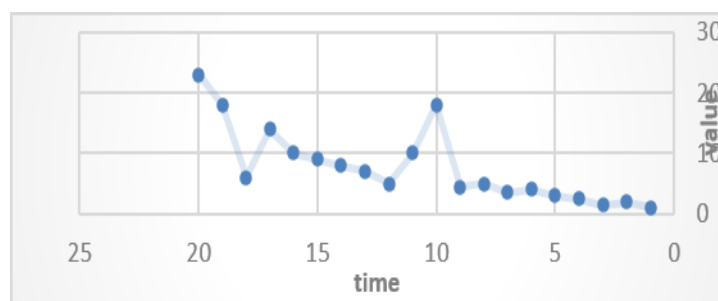


Figure (1) : Time Series Data.

The above figure is an illustration of time series data in which the points that are recorded indicate variations in temperature across consecutive times in the day. The chart shows that the values fluctuate at one point to the next, and this fact shows that time series are dynamic because one value is dependent on the temporal context. Such data is normally utilized to estimate trends, identify abnormal changes or understand thermal behavior during a given time. Other studies have highlighted the complex time nature of such data, which implies that time series involve both long and short term trends, seasonality and dynamic changes that are complex, and must be taken into consideration when developing models [14]. Many studies have also indicated that such data has characteristics of both time and frequency and their analysis is associated with the existence of trends, seasonality and complex dynamic changes that should be taken into consideration when developing models [14]. Time series studies have also demonstrated that time series capture the full structure of the system, and must be modeled in a way that comprehends the correlations of the signal and its constituents [12], [18].

2.2. Augmentation

Data augmentation is considered one of the most effective techniques for improving the quality and quantity of training data, especially in scenarios where data are limited or imbalanced. The main idea is to generate new samples from existing data while preserving their essential characteristics, which helps improve model robustness and reduce overfitting [13]. In time series applications, augmentation techniques aim to introduce controlled variations into the data without destroying the temporal structure. Traditional approaches apply simple transformations such as adding noise, scaling, time warping, or rearranging segments of the signal. These methods increase data diversity and allow models to learn more flexible representations of temporal patterns [10]. More advanced approaches rely on deep generative models, such as Generative Adversarial Networks (GANs) and diffusion-based models, to generate synthetic time series data that closely resemble real observations. These models are capable of learning complex data distributions and producing realistic variations while maintaining statistical properties of the original data [18]. Furthermore, augmentation can also be used to address class imbalance by generating additional samples for minority classes. This helps models better distinguish between different patterns and improves classification performance. Several studies have reported that the use of augmentation can lead to performance improvements ranging from 10% to 30%, depending on the dataset and application [19]. In addition, augmentation enhances the ability of models to handle noise and dynamic changes in real-world environments. By exposing models to more diverse states during training, it becomes possible to build systems that are more robust and adaptable to variations in time series data [20].

2.3. Time Series Data Augmentation

Time series data augmentation has become a key component in many applications such as forecasting, classification, and anomaly detection. The main objective of these techniques is to generate new samples that follow the same distribution as the original data while preserving their temporal characteristics [21]. Previous studies have shown that augmentation plays an important role in reducing overfitting and improving model generalization, particularly in applications that require high precision, such as medical diagnosis and industrial monitoring systems. For example, simple augmentation methods such as cutout, mixup, and window warping have been shown to improve the performance of deep learning models by up to 1%–45% on datasets from the UEA archive [22]. , survey studies have highlighted that both traditional techniques (e.g., cropping, stretching, and noise injection) and modern approaches (e.g., GANs and VAEs) are widely used to address data imbalance and enhance the detection of rare patterns in time series data [23]. These techniques contribute to building more stable and reliable models by providing diverse training samples. Moreover, augmentation can improve the representation of data in feature space. Visualization methods such as t-SNE and UMAP have shown that augmented data can lead to clearer separation between clusters, which helps improve classification performance and interpretability of the learned representations [23].

3. Literature Review

According to the methods that are used with the augmentation, the works in this field have three trends; statistical methods that are traditional and oldest one, machine learning methods, and the newest methods that are related with deep learning.

3.1. Traditional and Statistical Methods

Traditional data augmentation methods for time series rely on direct manipulation of the original signal without the need for complex learning models. These approaches are widely used due to their simplicity and low computational cost, making them suitable for scenarios with limited resources.

Annaki et al. (2024) [8], evaluated several traditional augmentation techniques on Human Activity Recognition datasets, including jittering, scaling, interpolation, smoothing, and time warping. The results showed that simple methods such as jittering and scaling achieved strong performance, reaching accuracy levels above 98% in some cases. However, more complex transformations inspired by image augmentation negatively affected performance, as they disrupted the temporal structure of the data

Oh et al. (2020) [13] proposed a time-series augmentation approach, in which interpolation is used to preserve the direction and trend of a series with low complexity, by creating virtual observation points with a higher density using the Cubic Spline function, and taking regular samples to create new synthetic series, to minimize the distortion of temporal structures in comparison to slicing or reordering. It was assessed on 85 datasets of the UCR archive with ResNet and augmentation in the training data. The findings indicated an increase in performance of 54 datasets, no effect on 9 datasets and adverse effect on 22 datasets with an average of 81.7 percent to 83.5 percent and a positive level of statistical significance ($P=0.0008$). Limitations were a reliance on performance on the interpolation function and density factor R choice, and possible performance deterioration when the characteristics were not changed based on the type of series.

3.2. Machine learning methods (hybrid / probabilistic)

Zhang et al. (2023) [14], proposed STAug, a method based on decomposing the time series into trend, seasonality, and noise components. Controlled transformations are applied to each component before recombining them into new sequences. The results showed that this structured approach improves model accuracy and robustness compared to simple transformations. However, the method is limited to data with clear seasonal patterns. The authors applied STAug to a variety of time series classification/forecasting data and showed that this structural augmentation is superior to simple random transformations, and the accuracy of the models is significantly enhanced, and their sensitivity to noise is mitigated. It is restricted to series which are representable by clear seasonality and trend and which may not be efficient when there are chaotic patterns in the data or rare events.

Akyash et al. (2021) [24], proposed a process of augmentation which is named DTW-Merge, where two sequences of a single type are aligned with the help of Dynamic Time Warping and consequently, the respective parts are merged into a new sequence that will have the general characteristics of the shape of a specific type and will not lose the time order. It was applied to augment UCR data like BeetleFly, and ResNet was trained on the original and augmented data. It was found that DTW-Merge was able to outperform a large variety of classical transformations (jittering, scaling, time warping, permutation, SPAWNER, WDBA, etc.) with an average accuracy of 83.07% and a positive t-test value indicating a statistically significant improvement. The principal shortcomings of the work are that the procedure is based on complete DTW alignment, which is computationally more expensive, and the merging is linear and does not consider multivariate relationships of a complex nature.

Ismail-Fawaz et al. (2023) [25], proposed a clustering-based augmentation approach using Dynamic Time Warping Barycenter Averaging (DBA) and its variants such as SoftDBA and ShapeDBA. In this method, K-means clustering is first applied to group similar time series, and then cluster centroids are generated using DBA techniques to represent the patterns of each group. The method improves classification and clustering performance by enhancing cluster representation. However, it assumes that samples within each cluster share similar shapes, which limits its performance when dealing with highly variable or high-dimensional multivariate time series.

Zhang et al. (2024) [26], presented M-SMOTE as an improved version of SMOTE designed to better handle time series characteristics. In this approach, minority samples are divided into safe, borderline, and noisy regions before generating synthetic data, allowing the model to focus on more reliable samples. The method uses enhanced similarity measures that consider temporal behavior rather than relying only on Euclidean distance. Experimental results showed noticeable improvements in classification accuracy and stability when compared to standard SMOTE. However, the method increases computational complexity and requires careful parameter tuning to achieve optimal performance.

Triggiano and Romito (2024) [27], Proposed a time series classification model that uses Gaussian process expansions and the expected signature as an extraction mechanism. During the first step, the K time series that are statistically consistent with the original time series are obtained by sampling a multivariate distribution using a fully supervised Gaussian Process model (with mean and covariance learned within the network). These samples would be subsequently added to the original signal in a long and better information bearing series. The signature and the expected signature of these series are then computed in the second stage and fed into a simple classifier with fixed dimensional representation. Experiments on both synthetic and real data demonstrated that this random Gaussian Process-based augmentation can produce superior signature model performance over not using augmentation, and nearly matches the performance of such strong models as 1NN-DTW and Hive-Cote 2. Conversely, the model is statistically stable upon the generation of multiple series K , thus taking longer to train, and the results are sensitive to the selection of normalization parameters during the computation of expected signature.

Compared to traditional methods, machine learning-based approaches provide more structured augmentation by considering similarity measures and clustering strategies. However, their performance is still limited when dealing with highly dynamic or high-dimensional time series data.

3.3. Deep learning methods

Hu et al. (2020) [28], proposed DATSING, a framework that integrates data augmentation with domain-adversarial learning. The approach generates augmented data across multiple domains and trains a model to learn domain-invariant representations using an adversarial discriminator. The method was evaluated on datasets from energy and transportation domains and showed improved generalization to unseen environments. However, it assumes the availability of domain labels and involves complex training procedures.

Hellerman et al. (2022) [29], proposed a financial time series augmentation approach based on image representation using Extended Intertemporal Return Plot (XIRP). In this method, time series values are mapped into images, where the diagonal represents original values and off-diagonal elements represent temporal returns. A WGAN-GP model is then used to generate synthetic images, which are converted back into time series data. The method achieved approximately 7% reduction in RMSE in 79% of the M4 datasets. However, the transformation to image space may lead to loss of fine temporal details, and performance is sensitive to the mixing ratio and model configuration.

Li et al. (2022–2023) [30], projected TTS-CGAN, a Conditional Generative Adversarial Network enhanced with Transformer architecture for generating future time series based on historical context. In this model, the generator produces synthetic sequences conditioned on past observations, while the discriminator distinguishes between real and generated samples using attention mechanisms to capture long-term temporal dependencies. The method was evaluated on industrial and IoT datasets, where it demonstrated improved predictive performance and enhanced anomaly detection capability compared to baseline models. The experimental results indicated that incorporating GAN-based augmentation can improve the quality of learned representations, although no fixed accuracy value was reported due to variation across datasets. However, the approach is sensitive to common GAN-related issues such as mode collapse and requires careful tuning of the loss function and training process to maintain generation stability and data quality.

El-Laham et al. (2023) [31], proposed StyleTime, a style-based generative model inspired by StyleGAN, designed to separate temporal content from style attributes such as amplitude and frequency. This allows controlled generation of new time series by recombining styles in the latent space. Experiments on synthetic and financial datasets showed that StyleTime generates high-quality data and improves classification performance when used for augmentation. However, the model is complex, computationally expensive, and the interpretation of style components remains challenging in practical applications.

Zhang et al. (2024) [32], proposed a hybrid framework for time series generation that combines data preprocessing, metric learning, and a multi-level variational autoencoder (ML-VAE), along with the MSTL decomposition framework. The method extracts temporal patterns such as trend and seasonality by transforming them into difference matrices, which are then integrated into the VAE to generate new data that preserves the original temporal structure. The results showed that the generated samples closely resemble real data, particularly when visualized in a three-dimensional space, indicating strong preservation of structural properties. Ablation studies confirmed that each component contributes significantly to the overall performance, as removing either the preprocessing or ML-VAE module leads to noticeable degradation. The model also outperformed other augmentation approaches such as CR-VAE, TimeGAN, and C-RNN-GAN in terms of similarity and predictive performance.

Ren et al. (2024) [33], proposed the Diff-MTS model, which is based on diffusion processes for generating multivariate time series while preserving both temporal and inter-variable relationships. The model progressively adds noise to the data and learns the reverse denoising process to reconstruct realistic future sequences. The generated samples are used as probabilistic augmentations that provide distributions rather than point estimates. Experimental results on traffic and energy datasets showed that Diff-MTS achieves better performance in terms of MAE and MSE compared to traditional Transformer and RNN models. However, the method is computationally intensive due to the multiple diffusion steps and presents challenges in controlling the trade-off between diversity and accuracy of generated samples.

Miljkovic et al. (2024) [34], focused on augmenting rare medical Electrogastragram (EGG) data using a GAN-based approach. The model employs 1D convolutional networks for both generator and discriminator and is trained on limited real samples to generate realistic synthetic signals. The augmented data significantly improved classification accuracy while preserving physiological characteristics of the signals. However, the limited size of the original dataset restricts the evaluation of generalization across diverse clinical conditions.

Oba et al. (2022) [35], proposed a dynamic data augmentation model based on a gating network that automatically selects the most suitable augmentation technique for each sample. The model integrates multiple expert networks and applies a feature consistency loss to ensure stable representation learning. The results demonstrated improved classification accuracy compared to traditional and automated augmentation methods such as RandAugment. However, the model complexity increases significantly due to the combination of multiple augmentation strategies.

Lee et al. (2023) [36] proposed the SeriesSleepNet model, which combines CNN and Bi-LSTM architectures for EEG-based sleep stage classification. The model incorporates a sliding window-based augmentation strategy to enhance temporal feature learning and address class imbalance. Experimental results showed accuracies of 0.87 and 0.84 on the Sleep-EDF and SHHS datasets, respectively. However, the model faces challenges in generalization across different devices and is sensitive to window configuration and loss weighting.

Saget et al. (2025) [37], proposed a resampling-based augmentation method designed for time series contrastive learning, particularly for remote sensing applications. The approach generates augmented samples by first up sampling the original time series and then extracting two disjoint subsequences that preserve the overall temporal coverage before resampling them back to the original length. These augmented sequences are used as positive pairs within a contrastive learning framework to improve representation learning. The method was evaluated on several agricultural datasets using Sentinel-2 imagery and demonstrated superior performance compared to traditional augmentation techniques such as jittering, resizing, and masking. The results showed that the proposed approach achieved state-of-the-art performance on the S2-Agri100 dataset, reaching approximately 72% classification accuracy when combined with VICReg and around 69% using SimCLR. However, the method depends on the design of subsequence sampling and requires large amounts of unlabeled data for effective training.

Overall, deep learning-based methods achieve the highest performance among augmentation techniques due to their ability to model complex temporal patterns. However, these methods often require high computational resources and careful tuning to ensure stable training.

Ref.	Author	Year	Method	Dataset	Accuracy	Limitation
[8]	Annaki et al.	2024	Jittering, Scaling, Smoothing, Time Warping, CutMix, Mixup	HAR	The best results were achieved with jittering and scaling, reaching >98% accuracy in some tasks.	Random methods lose temporal structure and may lead to decreased performance.
[13]	Oh. et al	2020	Cubic interpolation of the curved line to increase smoothness	85 datasets – UCR	Accuracy increased from 81.7% to 83.5% and improved across 54 datasets.	Performance depends on the choice of interpolation parameters and may cause distortion if the values are not adjusted.

[14]	Zhang et al.	2023	STAug – Breaking down the series (trend/seasonal/residual) + editing it	Forecasting + Classification	Achieves higher results than random transformations and a significant improvement in noise resistance	Less effective with chaotic or irregularly patterned strings
[24]	Akyash et al.	2021	DTW-Merge (String Merging via DTW Alignment)	UCR Datasets	Achieves an accuracy rate of 83.07%, surpassing traditional methods	Requires significant computational time due to DTW alignment
[25]	Ismail-Fawaz et al.	2023	ShapeDBA / SoftDBA (Paris Center centered on DTW)	Clustering + Classification	Improved representation of centers and increased accuracy in shape recognition	Assumes strong morphological similarity within each cluster
[26]	Zhang et al.	2024	M-SMOTE (Modified SMOTE for time series)	Multiple time series datasets	Improved performance compared to SMOTE	Requires fine-tuning of parameters to ensure stability + computational cost
[27]	Triggiano & Romito	2024	Gaussian Process Expansion + Expected Signature	Synthetic + Real datasets	Improved accuracy of signature models and closer performance to 1NN-DTW	Requires a large number of generated samples (K), which increases training time
[28]	Hu et al.	2020	DATSING (Augmentation + Domain Adversarial Learning)	Energy + Transport	Improving generalization across new domains	Requires domain identity knowledge
[29]	Hellermann et al.	2022	XIRP + WGAN-GP (Image-style TS augmentation)	M4 dataset	RMSE reduction by ~7% in 79% of data	Potential loss of fine details in strings
[30]	Li et al.	2022	TTS-CGAN (Conditional GAN + Transformer)	Bio signals	Improving the performance of traditional models and generating suitable data for anomaly detection	Mode collapse risk
[31]	El-Laham et al.	2022	StyleTime (StyleGAN for TS)	Finance + Synthetic	Generating controlled data (style maps)	High computational complexity
[32]	Zhang et al.	2024	ML-VAE + Metric Learning + MSTL (Hybrid DL)	Power/Energy datasets	Highest similarity to real data and outperforms TimeGAN	Requires fine-tuning of parameters to ensure stability

[33]	Ren et al.	2024	Diff-MTS (Diffusion for Multivariate Time Series)	Traffic, Energy	Improving MAE/MSE and achieving diversified future distributions	Large number of diffusion steps and high cost
[34]	Miljkovic et al.	2024	Medical EGG-GAN	EGG signals	Significant improvement in the accuracy of classifying gastric disorders	Limited original data hinder generalization
[35]	Oba et al.	2022	Dynamic DA with Gating Network	12 UCR datasets	Automatic selection of the best DA and improved accuracy compared to RandAug and α -Aug	Precise λ adjustment is required to prevent feature scattering
[36]	Lee, M. et al.	2023	SeriesSleepNet (CNN + BiLSTM + sliding-window augmentation)	Sleep-EDF, SHHS	Achieved accuracy of 0.87 and 0.84, surpassing traditional models.	Sensitivity to window size and difficulty generalizing across different devices.
[37]	Saget et al.	2025	Resampling Augmentation	Sentinel-2 / S2-Agri100	72% (VICReg), 69% (SimCLR)	Depends on sampling + need large amount of data

It is important to note that many of these methods do not report a single fixed accuracy value, as their performance is typically evaluated across multiple datasets and under different experimental settings. For this reason, the reported results are often expressed in terms of overall improvement or comparative performance rather than a specific accuracy value.

4. conclusion

A wide range of time series data augmentation techniques has been explored, and each method shows certain strengths as well as inherent limitations. Traditional approaches are simple and efficient, but they often fail to capture complex temporal patterns. Machine learning-based methods provide more structured augmentation, yet they remain limited when dealing with highly dynamic data. In contrast, recent deep learning-based techniques have demonstrated superior performance, mainly due to their ability to model complex data distributions and generate more realistic synthetic samples. Despite these advantages, deep learning methods are not without challenges, particularly in terms of computational cost and training stability. Therefore, relying on a single approach may not always lead to optimal results. A more promising direction is to combine the strengths of different augmentation strategies. In particular, integrating traditional or structured methods with deep learning techniques can help improve both data quality and diversity. Such hybrid approaches offer a balanced solution, leading to more reliable performance and better generalization across different time series applications.

References

- [1] N. Ashfaq, M. H. Khan, and M. A. Nisar, "Identification of Optimal Data Augmentation Techniques for Multimodal Time-Series Sensory Data: A Framework," *Information (Switzerland)*, vol. 15, no. 6, Jun. 2024, doi: 10.3390/info15060343.
- [2] D. Solís-Martín, J. Galán-Páez, and J. Borrego-Díaz, "D3A-TS: Denoising-Driven Data Augmentation in Time Series." [Online]. Available: <https://ssrn.com/abstract=4663081>
- [3] Z. Gao, H. Liu, and L. Li, "Data Augmentation for Time-Series Classification: An Extensive Empirical Study and Comprehensive Survey," Jun. 2025, doi: 10.1613/jair.1.17084.
- [4] I. Annaki, M. Rahmoune, and M. Bourhaleb, "Overview of Data Augmentation Techniques in Time Series Analysis," 2024. [Online]. Available: www.ijacsa.thesai.org

- [5] F. Huang and Y. Deng, "TCGAN: Convolutional Generative Adversarial Network for time series classification and clustering," *Neural Networks*, vol. 165, pp. 868–883, Aug. 2023, doi: 10.1016/j.neunet.2023.06.033.
- [6] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," Nov. 2014, [Online]. Available: <http://arxiv.org/abs/1411.1784>
- [7] D. Sisodia and L. Singh, "Clustering Techniques: A Brief Survey of Different Clustering Algorithms."
- [8] I. Annaki, M. Rahmoune, and M. Bourhaleb, "Overview of Data Augmentation Techniques in Time Series Analysis," 2024. [Online]. Available: www.ijacsa.thesai.org
- [9] R. Ilbert, T. V. Hoang, and Z. Zhang, "Data Augmentation for Multivariate Time Series Classification: An Experimental Study," Jun. 2024, [Online]. Available: <http://arxiv.org/abs/2406.06518>
- [10] K. M. Rashid and J. Louis, "Window-Warping: A Time Series Data Augmentation of IMU Data for Construction Equipment Activity Identification."
- [11] J. Paul *et al.*, "PSA-GAN: Progressive Self Attention GANs for Synthetic Time Series," Mar. 2022, [Online]. Available: <http://arxiv.org/abs/2108.00981>
- [12] L. Roque, C. Soares, V. Cerqueira, and L. Torgo, "L-GTA: Latent Generative Modeling for Time Series Augmentation," Jul. 2025, [Online]. Available: <http://arxiv.org/abs/2507.23615>
- [13] C. Oh, S. Han, and J. Jeong, "Time-series data augmentation based on interpolation," in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 64–71. doi: 10.1016/j.procs.2020.07.012.
- [14] X. Zhang, R. R. Chowdhury, J. Shang, R. Gupta, and D. Hong, "Towards Diverse and Coherent Augmentation for Time-Series Forecasting," Mar. 2023, [Online]. Available: <http://arxiv.org/abs/2303.14254>
- [15] R. Shumway and D. S. Stoo, "Springer Texts in Statistics Time Series Analysis and Its Applications With R Examples Fourth Edition."
- [16] E. Fons, P. Dawson, X. Zeng, J. Keane, and A. Iosifidis, "Evaluating data augmentation for financial time series classification," Oct. 2020, [Online]. Available: <http://arxiv.org/abs/2010.15111>
- [17] R. Moraffah *et al.*, "Causal Inference for Time series Analysis: Problems, Methods and Evaluation," Feb. 2021, [Online]. Available: <http://arxiv.org/abs/2102.05829>
- [18] X. Yuan and Y. Qiao, "Diffusion-TS: Interpretable Diffusion for General Time Series Generation," Oct. 2024, [Online]. Available: <http://arxiv.org/abs/2403.01742>
- [19] N. Ashfaq, M. H. Khan, and M. A. Nisar, "Identification of Optimal Data Augmentation Techniques for Multimodal Time-Series Sensory Data: A Framework," *Information (Switzerland)*, vol. 15, no. 6, Jun. 2024, doi: 10.3390/info15060343.
- [20] D. Solís-Martín, J. Galán-Páez, and J. Borrego-Díaz, "D3A-TS: Denoising-Driven Data Augmentation in Time Series." [Online]. Available: <https://ssrn.com/abstract=4663081>
- [21] N. A. Mat Leh, M. N. Ibrahim, M. K. Osman, A. Mohamad, and M. N. Mohd Hussain, "A Review on Time Series Data Augmentation Techniques for Deep Learning," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 62, no. 3, pp. 139–150, Dec. 2024, doi: 10.37934/araset.62.3.139150.
- [22] H. Yang and T. Desell, "Robust Augmentation for Multivariate Time Series Classification," Jan. 2022, [Online]. Available: <http://arxiv.org/abs/2201.11739>
- [23] G. Iglesias, E. Talavera, Á. González-Prieto, A. Mozo, and S. Gómez-Canaval, "Data Augmentation techniques in time series domain: A survey and taxonomy," Feb. 2024, doi: 10.1007/s00521-023-08459-3.
- [24] M. Akyash, H. Mohammadzade, and H. Behroozi, "DTW-Merge: A Novel Data Augmentation Technique for Time Series Classification," Aug. 2021, [Online]. Available: <http://arxiv.org/abs/2103.01119>
- [25] A. Ismail-Fawaz *et al.*, "ShapeDBA: Generating Effective Time Series Prototypes using ShapeDTW Barycenter Averaging," Sep. 2023, [Online]. Available: <http://arxiv.org/abs/2309.16353>
- [26] Hongchun Qu, "A Time Series Data Augmentation Method based on SMOTE," *Zhang Zheng*, May 2024.
- [27] F. Triggiano and M. Romito, "Gaussian Processes Based Data Augmentation and Expected Signature for Time Series Classification," *IEEE Access*, vol. 12, pp. 80884–80895, 2024, doi: 10.1109/ACCESS.2024.3408712.
- [28] H. Hu, M. J. Tang, and C. Bai, "DATSING: Data Augmented Time Series Forecasting with Adversarial Domain Adaptation," in *International Conference on Information and Knowledge Management, Proceedings*, Association for Computing Machinery, Oct. 2020, pp. 2061–2064. doi: 10.1145/3340531.3412155.

-
- [29] J. Hellermann, Q. Qian, and A. Shah, “Financial Time Series Data Augmentation with Generative Adversarial Networks and Extended Intertemporal Return Plots,” May 2022, [Online]. Available: <http://arxiv.org/abs/2205.08924>
- [30] X. Li, A. H. H. Ngu, and V. Metsis, “TTS-CGAN: A Transformer Time-Series Conditional GAN for Biosignal Data Augmentation,” Jun. 2022, [Online]. Available: <http://arxiv.org/abs/2206.13676>
- [31] Y. El-Laham and S. Vyetenko, “StyleTime: Style Transfer for Synthetic Time Series Generation,” in *Proceedings of the 3rd ACM International Conference on AI in Finance, ICAIF 2022*, Association for Computing Machinery, Inc, Nov. 2022, pp. 489–496. doi: 10.1145/3533271.3561772.
- [32] C. Zhang, H. Qin, Y. Zhang, C. Jiang, D. Zhang, and W. Deng, “Augmenting time series data: An interpretable approach with metric learning and variational autoencoders,” *International Journal of Electrical Power and Energy Systems*, vol. 161, Oct. 2024, doi: 10.1016/j.ijepes.2024.110190.
- [33] L. Ren, H. Wang, and Y. Laili, “Diff-MTS: Temporal-Augmented Conditional Diffusion-based AIGC for Industrial Time Series Towards the Large Model Era,” Jul. 2024, doi: 10.1109/TCYB.2024.3462500.
- [34] N. Miljković, N. Milenić, N. B. Popović, and J. Sodnik, “Data augmentation for generating synthetic electrogastrogram time series,” *Med. Biol. Eng. Comput.*, vol. 62, no. 9, pp. 2879–2891, Sep. 2024, doi: 10.1007/s11517-024-03112-0.
- [35] D. Oba, S. Matsuo, and B. K. Iwana, “Dynamic Data Augmentation with Gating Networks for Time Series Recognition,” May 2022, [Online]. Available: <http://arxiv.org/abs/2111.03253>
- [36] M. Lee, H. G. Kwak, H. J. Kim, D. O. Won, and S. W. Lee, “SeriesSleepNet: an EEG time series model with partial data augmentation for automatic sleep stage scoring,” *Front. Physiol.*, vol. 14, 2023, doi: 10.3389/fphys.2023.1188678.
- [37] A. Saget, B. Lafabregue, A. Cornuéjols, and P. Gançarski, “Resampling Augmentation for Time Series Contrastive Learning: Application to Remote Sensing,” Jun. 2025, [Online]. Available: <http://arxiv.org/abs/2506.18587>