Data Augmentation for Time-Series Classification: An Extensive Empirical Study and Comprehensive Survey

ZIJUN GAO, The School of Engineering, University of Warwick, U.K.

HAIBAO LIU, The School of Engineering and Materials Science, Queen Mary University of London, U.K.

LINGBO LI*, The School of Engineering, University of Warwick, U.K. and State Key Laboratory for Novel Software Technology, Nanjing University, China

Background: Data Augmentation (DA) has become a critical approach in Time Series Classification (TSC), primarily for its capacity to expand training datasets, enhance model robustness, introduce diversity, and reduce overfitting. However, the current landscape of DA in TSC is plagued with fragmented literature reviews, nebulous methodological taxonomies, inadequate evaluative measures, and a dearth of accessible and user-oriented tools.

Objectives: This study addresses these challenges through a comprehensive examination of DA methodologies within the TSC domain.

Methods: Our research began with an extensive literature review spanning a decade, revealing significant gaps in existing surveys and necessitating a detailed analysis of over 100 scholarly articles to identify more than 60 distinct DA techniques. This rigorous review led to the development of a novel taxonomy tailored to the specific needs of DA in TSC, categorizing techniques into five primary categories to guide researchers in selecting appropriate methods with greater clarity. In response to the lack of comprehensive evaluations of foundational DA techniques, we conducted a thorough empirical study, testing nearly 20 DA strategies across 15 diverse datasets representing all types within the UCR time-series repository. To improve practical use, we have consolidated most of these methods into a unified Python Library, whose user-friendly interface facilitates experimenting with various augmentation techniques, offering practitioners and researchers a more convenient tool for innovation than currently available options.

Results: Using ResNet and LSTM architectures, we employed a multifaceted evaluation approach, including metrics such as Accuracy, Method Ranking, and Residual Analysis, resulting in a benchmark accuracy of 84.98 ± 16.41% in ResNet and 82.41 ± 18.71% in LSTM. Our investigation underscored the inconsistent efficacies of DA techniques, for instance, methods like RGWs (with an average rank of 7.13 and average accuracy of 83.42 ± 17.53% in LSTM) and Random Permutation significantly improved model performance, whereas others, like EMD, were less effective. Furthermore, we found that the intrinsic characteristics of datasets significantly influence the success of DA methods, leading to targeted recommendations based on empirical evidence to help practitioners select the most suitable DA techniques for specific datasets.

Conclusions: In essence, this research presents an integrative perspective on the contemporary landscape of data augmentation for time series classification, combining theoretical frameworks with empirical evidence. The revelations and resources introduced herein are positioned to catalyze continued progress in this domain, fortifying machine learning models against the challenges posed by data limitations, and enhancing their generalizability and robustness.

JAIR Track: Surveys

1 Introduction

In machine learning, the pursuit of robust model performance often depends on the availability of extensive datasets, a challenge exacerbated by the frequent scarcity of data [1]. Data augmentation addresses this issue by generating additional, analogous data to expand limited datasets, thereby mitigating data insufficiency constraints. This technique, renowned for its cost-effectiveness and efficacy, has gained immense traction in various machine learning disciplines, significantly improving model accuracy in environments characterized by limited data. For instance, in computer vision, image augmentation is relatively straightforward, with a variety of simple yet effective techniques capable of producing discernible images even when subjected to noise or partial cropping [2].

In contrast, time-series data augmentation has received less attention. Existing strategies often overlook the intrinsic properties defining time-series data, such as temporal correlation, a critical aspect that distinguishes it from other data types. The potential for transforming time-series data into frequency or time-frequency domains further complicates the application of conventional augmentation techniques, particularly those borrowed from image or speech processing, leading to potential inaccuracies in synthetic data generation. Additionally, the effectiveness of data augmentation methods is highly task-dependent, requiring a nuanced approach that addresses the unique demands of tasks like forecasting and classification within the time-series context.

This paper offers an extensive examination of the current state of Data Augmentation (DA) for Time Series Classification (TSC), exploring a wide range of techniques and methodologies that have shaped the field over the past decade, providing readers with a comprehensive understanding of its evolution.

- **Comprehensive Survey and Novel Taxonomy:** To bring clarity and structure to the diverse array of techniques in this field, we introduce a novel taxonomy specifically designed for DA in TSC. This classification system organizes methods into five fundamental categories: Transformation-Based, Pattern-Based, Generative, Decomposition-Based, and Automated Data Augmentation, enabling a systematic and streamlined approach to selecting and analyzing DA methods for TSC.

- **Detailed Evaluation:** Beyond a traditional survey and taxonomy, this study makes a significant contribution by rigorously evaluating nearly twenty DA techniques for TSC. These methods were tested using two established neural network models, providing a robust performance benchmark. Our empirical analysis offers a comprehensive perspective on the relative strengths and weaknesses of these techniques, delivering valuable insights for both academia and industry.

- **Unified Library:** To complement our efforts, we have developed a Python Library that encapsulates most of the aforementioned time-series data augmentation algorithms, designed for ease of implementation. This initiative addresses the shortcomings of existing tools, offering a more user-friendly alternative[1].

The insights gained from our research provide a broad view of current techniques while highlighting potential areas for improvement and future research. We aim to foster further innovation in DA for TSC, paving the way for the development of more sophisticated and effective methods tailored to the complexities of time-series data. The main contributions of this manuscript are as follows:

- We conduct a thorough review of advancements in Data Augmentation for Time Series Classification (DA for TSC) over the past decade;

---

[1]https://github.com/RangerG/Dataugmenter

- We advocate a novel, purpose-built taxonomy for DA for TSC, segmenting techniques into quintessential categories: Transformation-Based, Pattern-Based, Generative, Decomposition-Based, and Automated Data Augmentation;
- We perform an in-depth evaluation of nearly twenty DA for TSC techniques, utilizing two renowned neural network models, and offer an exhaustive discourse on the findings;
- We introduce a Python Library of preeminent time series data augmentation algorithms, offering a unified, user-friendly platform for the application of diverse data augmentation strategies to time series datasets.

The structure of this article is meticulously organized to facilitate a comprehensive understanding of our investigation, as depicted in Figure 1. Section 2 embarks on a thorough literature review, delineating the state of the art in Data Augmentation (DA) for Time Series Classification (TSC), establishing the foundation upon which our study is predicated. In Section 3, we elaborate on the methodological framework adopted for our survey, including the criteria for selection and analysis of relevant works, and engage in a detailed discussion of the findings derived therefrom. Section 4 introduces a novel taxonomy for DA techniques, methodically categorizing these methodologies into five distinct principal categories, thereby offering a structured lens through which the current landscape can be viewed. Section 5 is dedicated to the elucidation of our experimental design, meticulously detailing the setup employed in our empirical study and articulating the research questions poised at the outset. Subsequently, Section 6 addresses these research questions, presenting a critical analysis of the experimental outcomes and drawing insights pertinent to the field of DA for TSC. Section 7 ventures into an exploration of the limitations inherent to our study and the potential threats to the validity of our findings, fostering a transparent discourse on the scope and applicability of our conclusions. The article culminates in Section 8, where we synthesize our contributions to the domain and outline directions for future research.

2 Background and Fundamentals

Time Series Classification: The domain of Time Series Classification (TSC) represents a pivotal frontier in data mining, dedicated to categorization of time series into distinct classes predicated upon historical pattern analysis [3]. Inherent in a time series is a sequence of data points, chronologically ordered and indexed over regular intervals, underscoring the temporal dynamics integral to various practical realms, including health diagnostics, financial market analytics, among others. The intricate nature of time series data, marked by heterogeneous lengths, noise interferences, and non-linear trajectories, poses substantial analytical challenges.

Data Augmentation: Within the spheres of machine learning and deep learning, data augmentation stands out as a strategy for amplifying the training dataset by generating modified iterations of the original data instances [4]. This process, which is applicable to various tasks such as image recognition, involves transformations such as rotation, scaling, or translation. In the context of TSC, augmentation encompasses techniques like noise injection, slicing, jittering, and scaling, serving dual core objectives: enhancing model generalization to diminish overfitting and addressing class imbalances within datasets.

Despite its established efficacy in various machine learning applications, data augmentation's integration into TSC remains a vibrant research avenue, necessitating deeper exploration. Historically, data augmentation has proven indispensable in scenarios plagued by data scarcity or imbalance, as evidenced in applications such as AlexNet for ImageNet classification [5]. Recent scholarly endeavors have extended these techniques to time series data, albeit with modifications to accommodate the unique characteristics of the data [6, 7].

The singular nature of time series data often precludes the direct application of standard augmentation methods, necessitating adaptations or entirely novel methodologies, particularly for tasks such as classification, forecasting, and anomaly detection within time series [8]. The past decade has witnessed a proliferation of TSC algorithms,
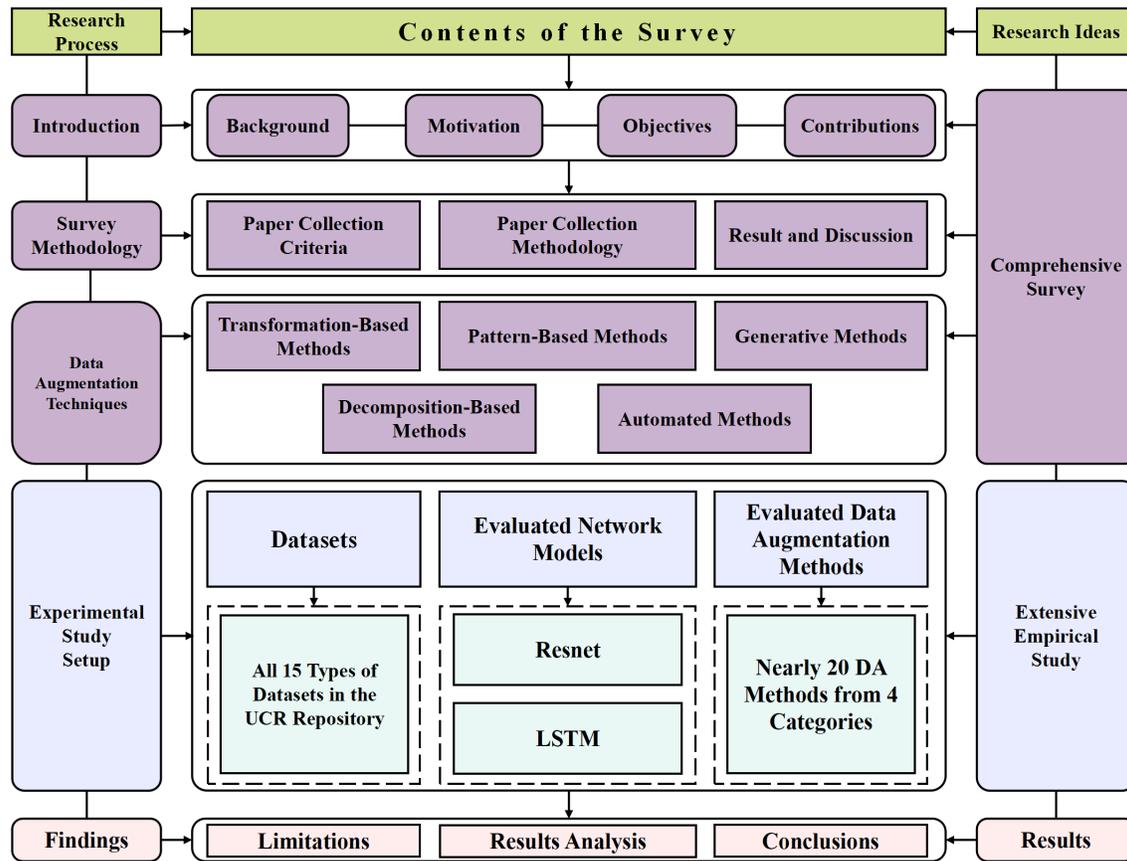
Fig. 1. The Overall Organization of this Survey Paper

driven by the increased availability of time-series data and its ubiquity across tasks that require human cognitive processes [9, 10]. The diverse types of datasets within the UCR / UEA archive further testify to the expansive applicability of TSC [11].

Addressing data imbalances, a prevalent issue in machine learning tasks, requires strategic data augmentation interventions, such as oversampling or undersampling, to rectify biases towards majority classes and enhance minority class performance. Techniques such as SMOTE (Synthetic Minority Oversampling Technique) generate synthetic examples to balance class distributions, while their counterparts reduce majority class instances [12]. However, these approaches require careful calibration to avoid overfitting or information loss.

Innovative contributions to the field include the two-layer convolutional network of Guennec et al. [13] for TSC, the time series-specific augmentation methods of Germain et al. [14], and the exploration of the augmentation of wearable sensor data [6]. Haradal et al. entered biosignal classification, employing generative adversarial networks (GANs) for the generation of synthetic time series data [15]. Fons et al. examined the classification of financial time series, evaluating various data augmentation strategies with deep learning models [16]. Nonaka et al. proposed augmentation of ECG-specific data, enhancing the classification of atrial fibrillation using single-lead

ECG data without modifying the DNN architecture [17]. Goubeaud et al. introduced White Noise Windows, a novel time series data augmentation technique, in 2021 [18].

Despite its advantages, data augmentation is not without its challenges, including the risk of overfitting and the potential ineffectiveness with high-dimensional or complex data structures [4, 19].

2.1 Relevant Survey & Study

The scholarly landscape is replete with surveys on data augmentation, spanning domains from image processing [20] to natural language processing [21], although comprehensive studies in the time series sector remain scarce. Iwana et al. conducted an empirical assessment of time series data augmentation methods using neural networks [4], their focus was classification tasks and excluding the frequency domain or externally trained data augmentation methods. Wen et al. expanded on this work, incorporating time series forecasting and anomaly detection, although their study utilized a relatively limited dataset from the Alibaba Cloud monitoring system [19].

Although these studies offer valuable insights into data augmentation techniques, there is a conspicuous absence of a holistic view that encompasses recent advances and diverse application fields. Several data augmentation techniques referenced in the literature are either antiquated or no longer align with contemporary research and practice standards, thereby failing to offer a user-friendly and accessible toolset for researchers and practitioners. This article endeavors to address this gap in the scholarly discourse by conducting a comprehensive analysis of the current landscape of data augmentation within the domain of time-series classification. It aims to serve as a pivotal reference point, guiding future research directions and fostering innovation in this rapidly evolving field.

3 Survey Methodology

This paper adopts a tripartite research methodology encompassing Survey, Evaluation, and Experimental Study Setup. The survey delineates our systematic approach to the accumulation and analysis of the literature. The evaluation introduces a novel taxonomy of time-series data augmentation, predicated on distinctive characteristics. The Experimental Study Setup discusses the authoritative datasets employed, evaluation metrics, and potential variables that influence outcomes, ensuring a rigorous analysis of time series data augmentation methodologies.

3.1 Paper Collection Criteria of the Survey

The aggregation of the literature for this study adheres to stringent inclusion criteria, encompassing:

- Papers elucidating the fundamental concepts of data augmentation for time series classification or its associated facets.
- Studies proposing enhancements to existing data augmentation methodologies, tools, or frameworks within the realm of time-series classification.
- Contributions introducing benchmarks or datasets tailored for data augmentation in time-series classification.
- Literature offering comprehensive reviews or surveys on the contemporary landscape of data augmentation for time series classification.

Exclusions pertain to studies applying relevant techniques outside the time-series context and those focusing on time-series classification, forecasting, or anomaly detection without integrating data augmentation methodologies.

3.2 Paper Collection Methodology of the Survey

Our literature retrieval process commenced with exhaustive searches across renowned scientific repositories such as Google Scholar and arXiv, utilizing precise keyword combinations. Recognizing the interdisciplinary application of time series data augmentation, we adopted a snowballing strategy for comprehensive coverage [22].

The accumulated literature was subsequently classified as Survey / Review papers, Research papers or Methods papers, following established academic conventions [23–25].

The keywords used for searching and collecting are listed below. [DA for TSC] means Data Augmentation for Time-Series Classification, which is the focus of this research project. The search strategy encompassed several iterations, as detailed below:

- DA for TSC + Survey / Review
- DA for TSC + Techniques / Methods
- DA for TSC + Evaluations / Comparison

3.3 Result and Discussion of the Survey

Table 1. Paper Query Results

Paper Query Results	Key Words	Title	Body
DA for TSC + Survey / Review	25	20	8
DA for TSC + Techniques / Methods	273	102	72
DA for TSC + Evaluations / Comparison	90	55	26
Query	-	-	106
Snowball	-	-	41
Overall	-	-	147

The terms "Key Words", "Title" and "Body" refer to the number of papers returned by the search, the number of papers that remain after title framing and the number of papers that remain after body framing, respectively. The search strategy yielded 147 relevant publications until 10th June, 2024, with a significant proportion originating from keyword searches and the remainder from snowballing. Trends in annual publication indicate a growing interest in this research area, particularly after 2019, as shown in Figure 2.

The distribution of publications, in Figure 3, across various publishers underscores the dominance of IEEE and arXiv, with significant contributions also emanating from Elsevier, MDPI, ACM, and Springer.

A geographical analysis of the publications is shown in Figure 4, which reveals a preponderance of contributions from the United States, followed by China, Japan, Germany, and the United Kingdom, with disparate representation from other global regions.

An industry-centric examination, based on the North American Industry Classification System (NAICS) is shown in Figure 5, highlights a predominant application within the Information Technology sector, followed by Healthcare, Agriculture, Astronomy, Finance, and Construction, indicating the methodologies' broad applicability.

4 Data Augmentation for Time-Series Classification Techniques

The taxonomy presented in this study for data augmentation techniques applicable to time-series classification (TSC) emerges from an exhaustive literature survey and rigorous research. It compartmentalizes the techniques into five principal categories: Transformation-Based Methods, Pattern-Based Methods, Generative Methods, Decomposition-Based Methods, and Automated Data Augmentation Methods. This categorization, meticulously refined and expanded on from previous studies [4, 8, 19], lays the groundwork for our in-depth exploration and experimentation.

Several factors underpin the formulation of this taxonomy. Primarily, it encompasses a broad spectrum of contemporary methods, thereby paving the way for domain-centric experimental undertakings in specialized
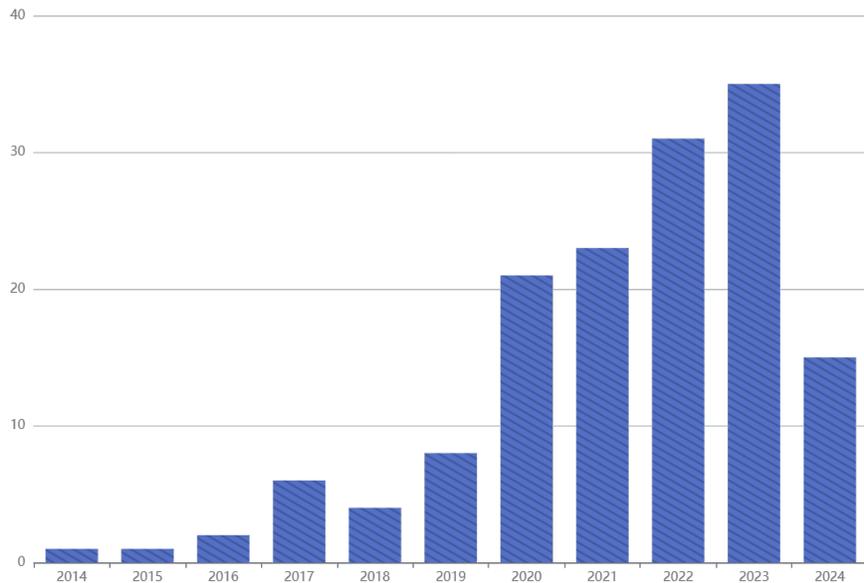
Fig. 2. Annual Publication Trends in Data Augmentation for Time-Series Classification
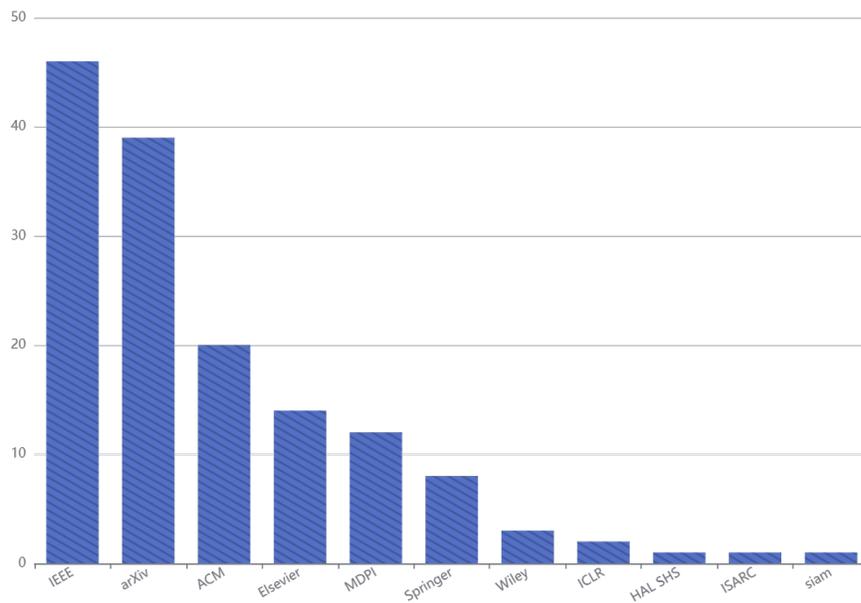

Fig. 3. Publisher Distribution of Publications in Data Augmentation for Time-Series Classification

segments like the time and frequency domains. Moreover, it offers an organized synopsis of the myriad techniques used for augmenting time-series data, thereby simplifying comparative analysis and comprehension of various
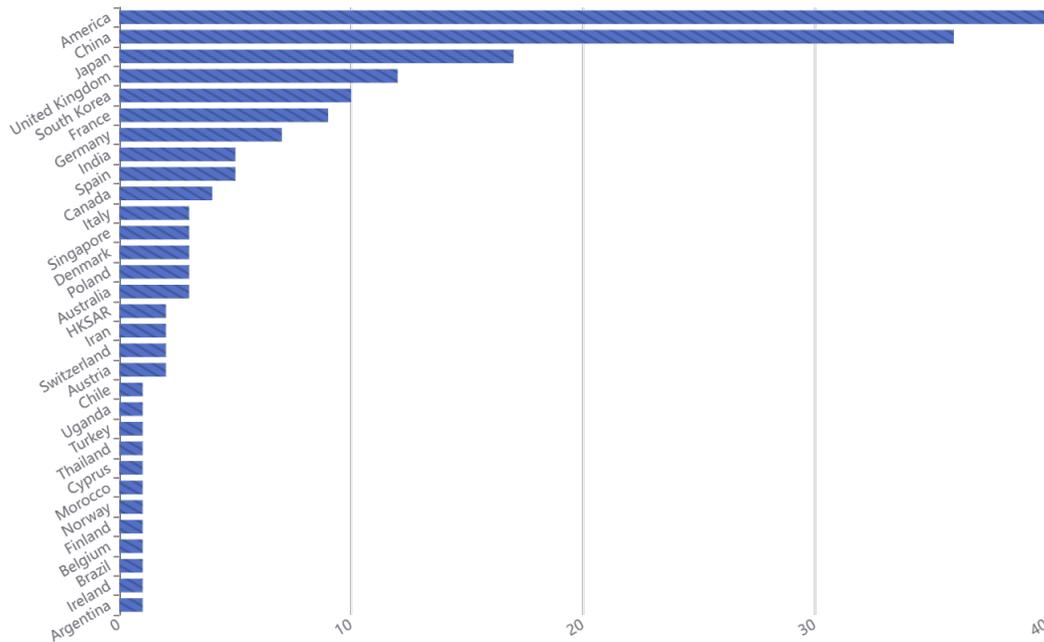
Fig. 4. Publication Count by Country in Data Augmentation for Time-Series Classification

methods. Such a structured approach is indispensable for researchers and practitioners alike, guiding them in the judicious selection of data augmentation techniques tailored to their specific requirements. Furthermore, by highlighting existing lacunas and prospects in the current scenario, it spurs the innovation of novel techniques.

The distinct categories within this taxonomy epitomize different methodologies in data augmentation. Transformation Based Methods involve the alteration of original data via various transformations. Pattern-Based Methods, on the other hand, generate new instances through the extraction and subsequent recombination of patterns inherent in the original data. Generative methods utilize generative models to synthesize new data instances. Decomposition-Based Methods, in contrast, entail the disintegration of the original time series into multiple components which are then individually augmented and reassembled. Automated Data Augmentation Methods streamline the augmentation process, often leveraging data-driven insights to determine the most effective augmentation strategy.

The ensuing subsections provide an exhaustive discourse on each category, elucidating their respective methodologies, merits, and potential use-cases. This detailed analysis serves two purposes: guiding users in the selection of appropriate data augmentation strategies and fostering the conception of avant-garde methods in forthcoming research endeavors.

4.1 Transformation-Based Methods

Transformation-based methods are based on the principle of applying a set of transformations to the original time series data. The objective is to retain the fundamental structure while introducing nuanced variations, thereby enhancing the resilience of the model. These methods, characterized by their simplicity and intuitiveness, typically encompass elementary mathematical operations, making them versatile across diverse data types and domains. The scope of transformations extends to various data facets, culminating in four distinct subcategories.

4.1.1 Magnitude Transformations.

Magnitude transformations alter the amplitude of time series data, an approach especially pertinent in contexts where the data's magnitude is informative, such as signal processing or audio analysis.

The investigation conducted by Terry et al. [6] meticulously explores the application of various data augmentation (DA) techniques—namely Jittering, Scaling, Rotation, and Magnitude-Warping—within the context of Parkinson's disease movement classification through the use of wearable sensor data. The study highlights a significant enhancement in classification performance attributable to the strategic integration of DA methods, with accuracy metrics elevating from a baseline of 77.52% to an impressive 86.88%. This advancement underscores the potential of DA in enriching the analysis of wearable sensor data. In a parallel vein, Shuang et al. [26] employ a combination of Scaling and Magnitude Warping, among other DA techniques, on I/Q signal data. This approach yields a remarkable peak recognition accuracy of 93.68%, facilitating the effective differentiation between QAM16 and QAM64 signals. Such findings corroborate the efficacy of DA techniques in enhancing signal recognition accuracy, pointing towards their vital role in the advancement of signal processing methodologies. Additionally, Goubeaud et al. [18] introduce White Noise Windows (WNW), a novel augmentation technique that fortifies classification methods and mitigates overfitting.

4.1.2 Time Transformations.

Time transformations adjust the temporal elements of the data, such as data point sequencing or timing. They are instrumental when the data's temporal dynamics are crucial, as seen in tasks like time-series classification, forecasting, or sequence prediction.
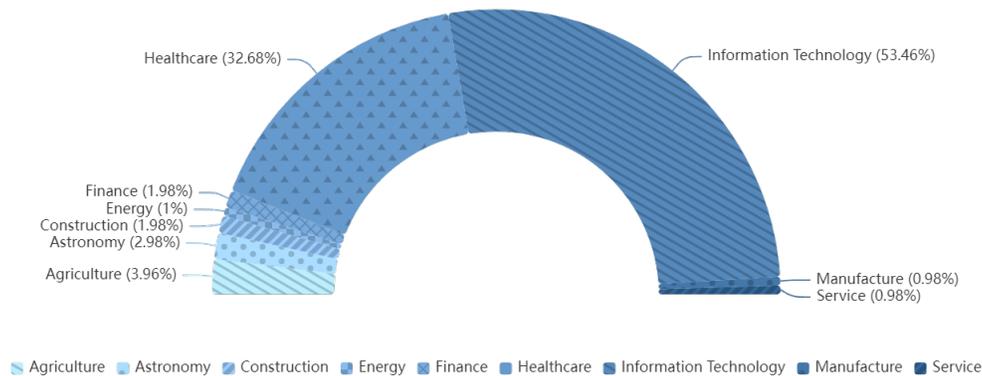

Fig. 5. Publication Count by Industry in Data Augmentation for Time-Series Classification

In addition to the aforementioned magnitude transformations, the work of Terry et al. [6] also integrates time-based data augmentation techniques such as Permutation and Time-Warping, further demonstrating their utility in enhancing Parkinson's disease movement classification through wearable sensor data. Similarly, the study by Shuang et al. [26] employs Time-Warping among its suite of techniques, showcasing its effectiveness in improving the recognition accuracy of I/Q signal data. These instances underline the broad applicability and potential of time transformations in augmenting data for varied classification tasks.

In the realm of IMU data from construction equipment, Rashid et al. [7] innovate with Window-Warping (WW), a technique that synthesizes training data, thus enhancing machine learning model proficiency in activity recognition tasks, with K-Nearest Neighbors (KNN) witnessing substantial improvements. Furthermore, Guennec
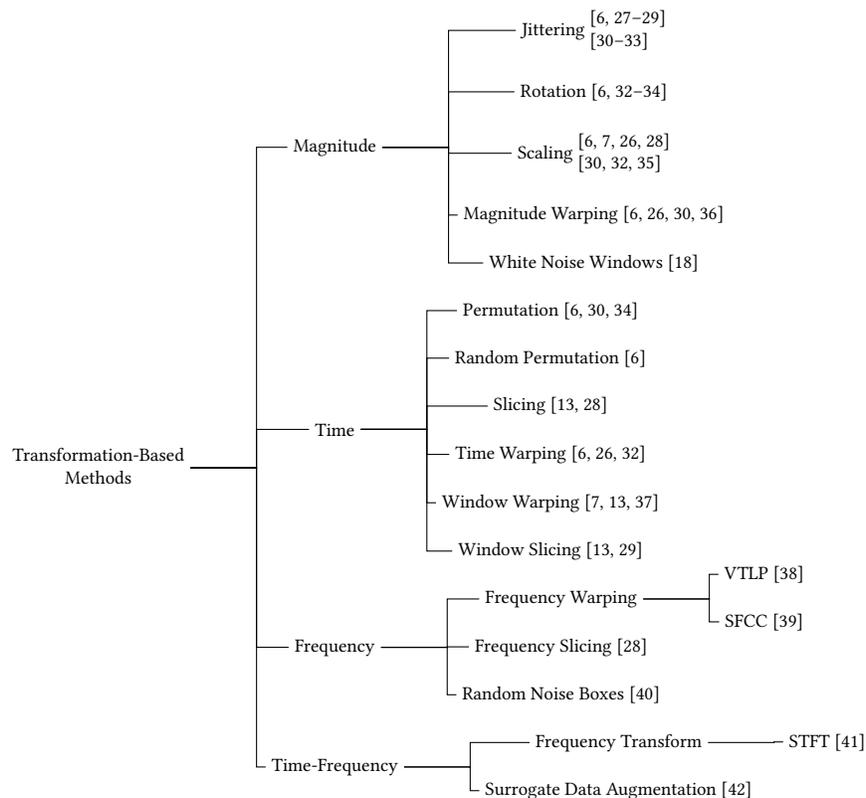
Fig. 6. Transformation-Based Methods of Data Augmentation for Time-Series Classification

et al. [13] discuss two augmentation methods for time series classification with Convolutional Neural Networks: Window Slicing (WS) and Window Warping (WW). These strategies, coupled with a semi-supervised Dataset Mixing technique, bolster classification performance, particularly in smaller datasets.

4.1.3 Frequency Transformations.

Frequency transformations tinker with the frequency components of time series data, a strategy quintessential in scenarios where the data frequency content is vital, such as in spectral or vibration analysis.

Yang et al. [39] pioneered the stratified Fourier coefficients combination (SFCC), a method that merges the original Fourier coefficients to augment time series datasets, thus boosting deep neural network performance. This efficient technique achieves state-of-the-art results. In a similar vein, Goubeaud et al. [40] reveal random noise boxes (RNBs), an innovative augmentation technique optimized for spectrogram classification. Although particularly potent for ECG and sensor datasets, its efficacy diminishes for motion datasets, underscoring the need for vigilant application to avert performance deterioration.

4.1.4 Time-Frequency Transformations.

Time-frequency transformations concurrently tweak both temporal and frequency aspects. This dual approach
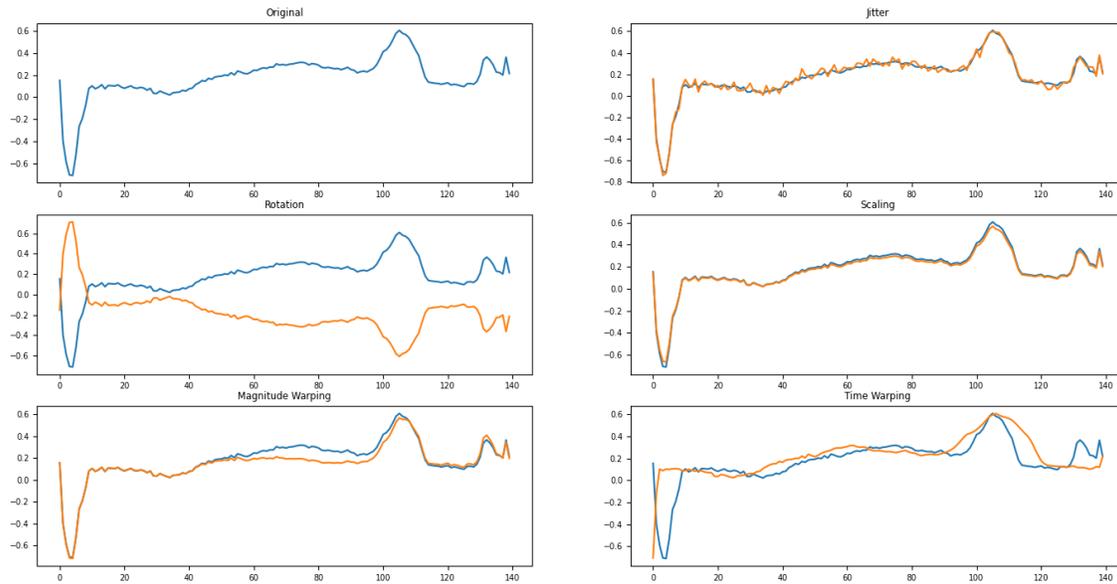
Fig. 7. Illustrative Examples of Transformation-Based Methods on the ECG5000 Dataset (Generated instances are depicted in orange)

is advantageous when both elements are pivotal, enabling the generation of new instances that preserve the original data's temporal and frequency structures.

Lee et al. [42] employ data augmentation techniques, notably the Amplitude Adjusted Fourier Transform (AAFT) and the Iterative Amplitude Adjusted Fourier Transform (IAAFT), to concoct surrogate data conducive to deep learning. These strategies maintain the time-domain statistics and power spectrum of the original time series, thereby enriching the training of pretrained Convolutional Neural Networks (CNNs). The research corroborates the efficacy of these methods in automated rehabilitation classification tasks, suggesting avenues for further exploration.

4.2 Pattern-Based Methods

Pattern-based methods focus on discerning and capitalizing on specific patterns or structures inherent in the time-series data. These patterns, local or global, simple or complex, often involve machine learning models for identification and utilization. The hallmark of this category is its emphasis on the data's intrinsic patterns, a focus that proves invaluable in contexts with pronounced structural or pattern-based components, such as electrocardiograms (ECG) or stock market data.

4.2.1 Magnitude Pattern Mining.

This niche capitalizes on significant local structures or motifs within the data, exploiting these elements to spawn new, realistic data instances.

Kamycki et al. [45] introduce SPAWNER, an innovative data augmentation method that aligns input sequences and restricts the warping path to produce synthetic examples, thus improving the performance of time series classifiers on benchmark datasets. In a parallel study, Aboussalah et al. [44] employs Recurrent In-fill Masking (RIM) in 2022, a novel augmentation method demonstrated to improve the performance of time series classification
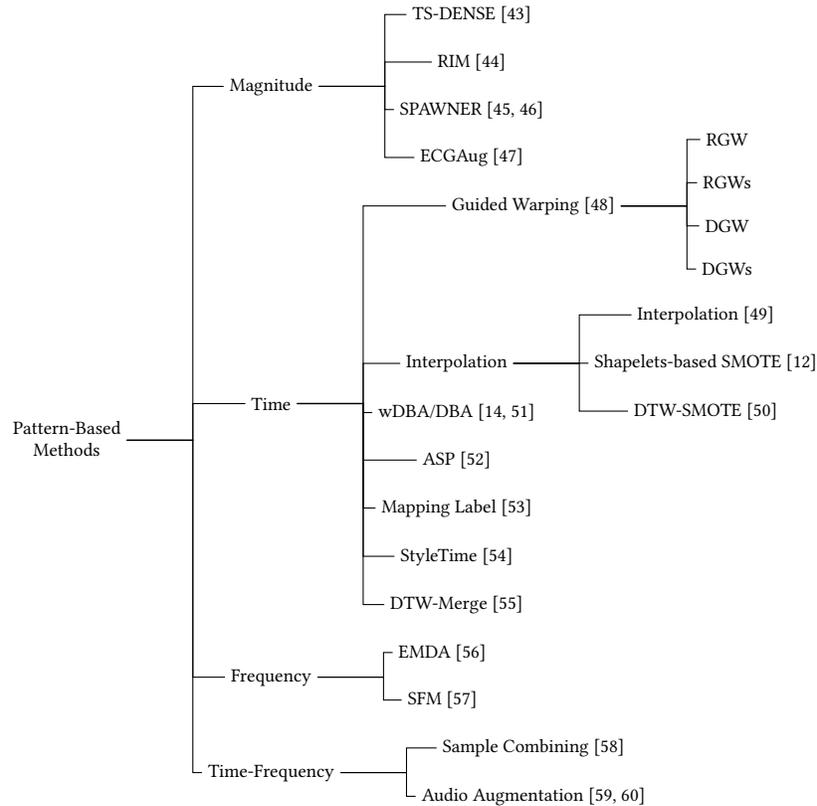
Fig. 8. Pattern-Based Methods of Data Augmentation for Time-Series Classification

tasks across synthetic and real-world datasets, outperforming traditional training and other augmentation strategies.

4.2.2 Time Pattern Mining.

These methods adjust the temporal characteristics of the data, maintaining the temporal structure of the original data while generating new instances.

Forestier et al. [14] advocate a data augmentation method that uses dynamic time warping (DTW) to synthesize examples by averaging time series sets, a method especially potent in scenarios with limited example availability. The study confirms that the expansion of full datasets typically enhances accuracy, with notable improvements in 56 of the 85 datasets tested. Similarly, Iwana et al. [48] and Cheolhwan et al. [49] introduce novel methods that take advantage of time warping and cubic interpolation, respectively, to generate additional training samples effectively, thus optimizing model performance across various datasets.

4.2.3 Frequency Pattern Mining.

Frequency Pattern Mining methods focus on manipulating the frequency components of time series data. These methods are pivotal when the data's frequency content is crucial, enabling the generation of new instances that preserve the original data's frequency structure.
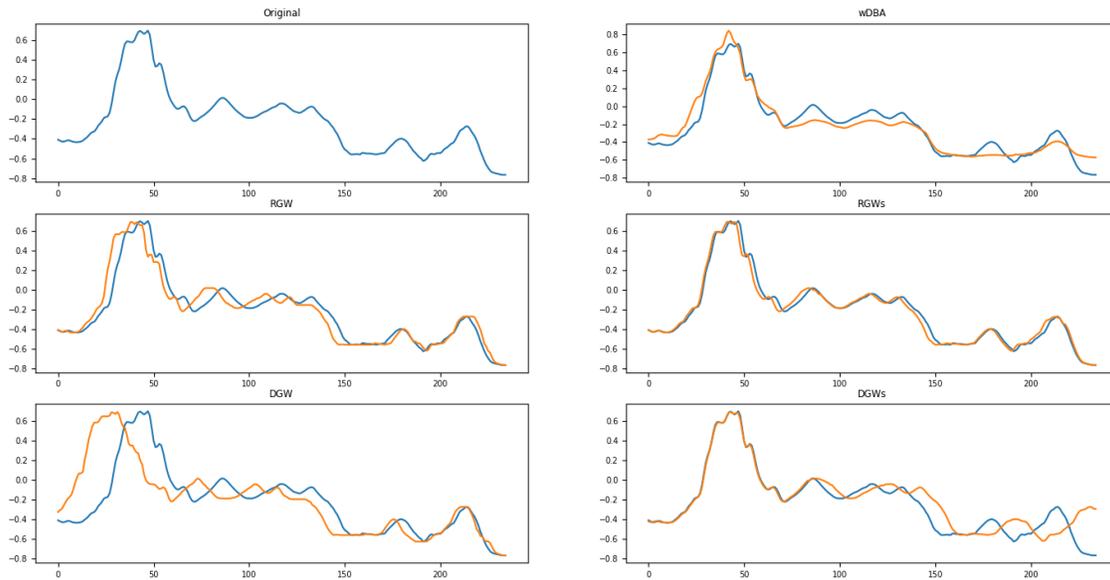
Fig. 9. Illustrative Examples of Pattern-Based Methods on the Strawberry Dataset (Generated instances are depicted in orange)

Loris et al. [56] introduce a technique known as EMDA, or Equalized Mixture Data Augmentation, a method premised on spectrograms. EMDA operates by forging a new spectrogram through the weighted average of two randomly selected spectrograms from the same class, incorporating perturbations like time-delay and equalizer functions. This strategy, as per the authors, enhances the diversity and robustness of audio signals, thereby fortifying classification performance. Conversely, Cui et al. [57] advocate for stochastic feature mapping (SFM), a data augmentation method tailored for deep neural network acoustic modeling. SFM, a label-preserving transformation, statistically transmutes one speaker's speech data into another's within a specified feature space. The essence of SFM is to amplify pattern variations via transformations, thereby bolstering the neural networks' classification invariance and generalization capabilities.

4.2.4 Time-Frequency Pattern Mining.

Time-Frequency Pattern Mining methods adjust both time and frequency components of the data. This dual-pronged approach is beneficial when both the data's temporal dynamics and frequency content are integral, as seen in domains like speech recognition or music analysis.

Zhao et al. [58] propose a unique method that transitions time domain data into the frequency domain via a discrete cosine transform (DCT), subsequently generating artificial data by merging different frequency bands. These synthesized signals are then reverted to the time domain through an inverse DCT, with experimental results underscoring the method's efficacy. In a related study, Nanni et al. [59] employ data augmentation strategies to enhance the accuracy of animal sounds classification using convolutional neural networks (CNN). These strategies, which encompasses transformations of audio signals and their spectrograms, utilize techniques such as pitch shift, noise addition, and equalized mixture. The findings confirm that data augmentation enhances the accuracy and generalizability of CNNs, with the fusion of different CNNs yielding optimal results.

4.3 Generative Methods

Generative Methods constitute a separate class within the taxonomy of time series data augmentation techniques. These methods leverage generative models to craft new time series instances, diverging from transformation-based and pattern-based methods by incorporating a learning phase. During this phase, the model discerns the data's underlying distribution, subsequently spawning new instances from this distribution. This category is invaluable when the objective is to generate diverse, realistic synthetic instances that mirror the original data's statistical attributes.
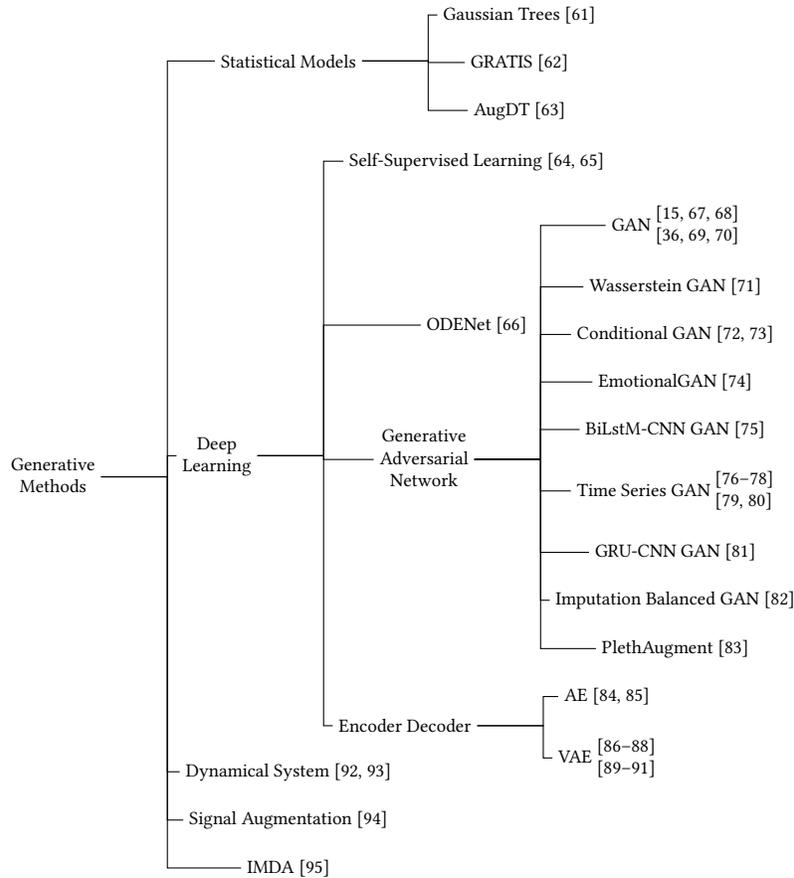
Fig. 10. Generative Methods of Data Augmentation for Time-Series Classification

4.3.1 Statistical Models.

Statistical Models utilize generative models with pre-established structures to produce new time series instances. These models are advantageous when the data's underlying distribution is known or can be reasonably inferred, facilitating the generation of new instances that adhere to the original data's distribution.

Cao et al. [61] present a parsimonious mixture of Gaussian trees (2MoGT) model, designed for data augmentation in scenarios with imbalanced and multimodal time series classification. This model employs Chow-Liu
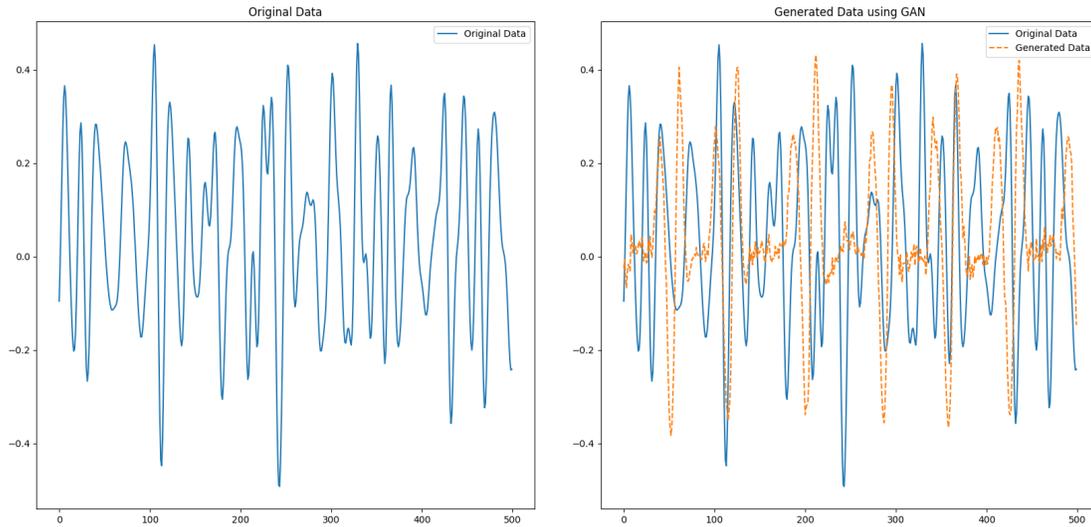
Fig. 11. Illustrative Examples of Generative Methods (GAN) on the FordB Dataset (Generated instances are depicted in orange)

trees to encapsulate dependencies among time-series features and a Gaussian mixture model to accommodate multimodality. It surpasses alternative oversampling methods, especially in datasets with multimodal minority class distributions, and minimizes storage complexity. Additionally, Kang et al. [62] introduce GRATIS, a data augmentation technique that generates time series with adjustable characteristics. Using a mixture autoregressive (MAR) model, attuned to specific features, it spawns diverse time series. The resulting series exhibit diversity, attributable to the mixtures' adaptability and random parameter configurations, which proves especially beneficial for analyzing collections of time series with varying lengths, scales, and characteristics.

4.3.2 Deep Learning Models.

Deep Learning Models employ generative models without predefined forms to create new time series instances. These models are beneficial when dealing with data containing complex, nonlinear, or multidimensional patterns, as they can generate new instances that encapsulate these intricate patterns.

In 2018, Haradal et al. [15] presented a data augmentation strategy based on Generative Adversarial Networks (GAN) for biosignal classification. This method incorporates a recurring neural network (RNN) with LSTM units for the generation of time series data. Its efficacy is validated by an uptick in classification accuracy, underscoring its viability for synthesizing biosignals and enhancing biosignal classification. In contrast, Garcia et al. [68] applied GANs for data augmentation, synthesizing light curves from variable stars. This methodology, combined with an innovative resampling technique and evaluation metric, markedly elevates the classification accuracy of variable stars, which proved particularly potent for unbalanced datasets and in pinpointing GAN-overfitting scenarios.

Several researchers have also harnessed specialized GANs for data augmentation. For example, Ehrhart et al. [73] used a conditional Generative Adversarial Network (cGAN) with a diversity term for data augmentation, using an LSTM for the generator and an FCN for the discriminator. This method synthesized moments of

stress from physiological sensor data, bolstering classifier performance and emerging as a promising technique for stress detection in wearable physiological sensor data. Moreover, Yang et al. [76] introduced a time series Generative Adversarial Network (TS-GAN) for augmenting sensor-based health data, generating synthetic time series data that preserve the temporal dependencies of the original data, thus enhancing model performance. The study asserts that TS-GAN surpasses conventional augmentation methods in maintaining data consistency. Furthermore, Deng et al. [82] implemented a new data augmentation strategy, IB-GAN, merging imputation and balancing techniques. It boosts classification performance by synthesizing synthetic samples for underrepresented classes, averting the mode collapse often seen in standard GANs. This approach records substantial performance enhancements against cutting-edge baselines. In 2019, Chen et al. [74] proposed an emotion state identification framework and an ECG signal augmentation algorithm based on GAN [74], which is known as EmotionalGAN. It is used to synthesize synthetic ECG data, which is then amalgamated with the original training set, improving the classification of emotion states. The findings indicate a notable improvement in classification performance with the integration of EmotionalGAN-synthesized data.

In their 2021 study, Jiang et al. [64] embarked on an exploration of self-supervised contrastive learning methodologies aimed at augmenting the accuracy of sleep stage classification through electroencephalogram (EEG) signal analysis. The cornerstone of their approach was the utilization of transformed EEG signal pairs as a pretext task, culminating in an empirical demonstration of enhanced classification performance, evidenced by an accuracy rate of 88.16% and an F1 score of 81.96% on the Sleep-edf dataset. This investigation not only substantiates the viability of self-supervised learning frameworks in the efficient processing of EEG signals but also illuminates its potential to mitigate the challenges associated with limited labeled datasets. In a parallel inquiry, Banville et al. [65] scrutinized the application of self-supervised learning (SSL) strategies for the derivation of discriminative EEG signal representations. The study's outcomes elucidated that SSL-derived features empower linear classifiers to achieve superior performance relative to supervised deep neural network models, particularly in contexts constrained by limited labeled data availability. These findings articulate the transformative impact of SSL on the efficacy of deep learning models deployed for EEG data analysis, heralding a new era of methodological advancements in EEG-based diagnostic and sleep stage delineation tasks.

Models based on the encoder-decoder structure have also gained traction in Data Augmentation for time-series classification. Moreno et al. [89] employ variable autoencoders (VAE) and generative adversarial networks (GANs) for data augmentation, improving prediction accuracy in small datasets, with substantial gains realized through computational effort. The study also introduces alterations to the balance class representation in synthetic samples. Zha et al. [84] unveils a straightforward but potent self-supervised model for time series generation, termed the Masked Autoencoder with Extrapolator (ExtraMAE). The synthetic time series crafted by ExtraMAE exhibit remarkable practicality and fidelity, eclipsing all benchmarks while remaining more compact and expedient in a minimalist design. Furthermore, Hsu et al. [91] introduce two Variational Autoencoder (VAE)-based data augmentation strategies for unsupervised domain adaptation in Automatic Speech Recognition (ASR). These methods, encompassing nuisance attribute replacement and soft latent nuisance subspace perturbation, alter nuisance attributes of speech, achieving approximately a 35% absolute Word Error Rate (WER) reduction.

4.4 Decomposition-Based Methods

Decomposition-Based Methods disassemble the original time series into multiple components, augment each component independently, and subsequently reassemble them to produce new instances. This unique tactic of deconstructing time series data and individually addressing each component proves advantageous when various components of the data necessitate distinct augmentation techniques.

In 2022, Li et al. [96] present a data augmentation strategy based on empirical mode decomposition (EMD) and an integrated LSTM neural network. Raw hand motor signals are fragmented by EMD, followed by the
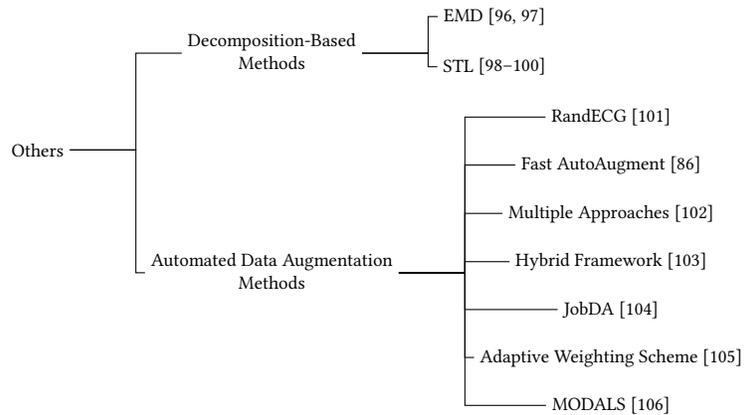
Fig. 12. Decomposition-Based and Automated Methods of Data Augmentation for Time-Series Classification

augmentation of the decomposed signals through LSTM. The method, tested in the public dataset NinaWeb, demonstrated a 5.2% ascension in classification accuracy compared to benchmark results. Wen et al. [98] introduce RobustSTL, a pioneering algorithm for time series decomposition that addresses anomalies, sudden trend shifts, seasonality alterations, and noise. It employs bilateral filtering for noise eradication, LAD regression for trend extraction, and non-local seasonal filtering for seasonality determination. The algorithm surpasses existing solutions in the management of complex time series data. Wen et al. [100], presents an extension of RobustSTL, designed to accommodate multiple seasonalities in time series data, utilizing a specialized generalized ADMM algorithm. This method outperforms leading algorithms like MSTL, STR, and TBATS in terms of effectiveness and computational efficiency, with future explorations geared towards multiresolution seasonal-trend decomposition for further enhancements.

4.5 Automated Data Augmentation Methods

Automated Data Augmentation Methods strive to automate data augmentation processes, often discerning the optimal augmentation strategy directly from the data. This category distinguishes itself with its focus on automation and adaptability, ideal for circumstances where manual selection and adjustment of augmentation techniques are either impractical or not optimal.

Park et al. [86] utilize PoseVAE and Fast AutoAugment for data augmentation, with PoseVAE rejuvenating skeleton data based on initial human poses, thus generating stable and novel data. Fast AutoAugment fine-tunes the augmentation policy employing RNN-based reinforcement learning. The research concludes that these methodologies can effectively augment skeleton data, boosting model performance.

In addition, Fons et al. [105], utilizing the W-Augment and α-trimmed Augment methods, present an adaptive weighting scheme for automatic time series data augmentation. These techniques improve the performance of time series classification tasks, especially in financial markets, by enhancing the accuracy of stock trend predictions. The authors ascertain that their methods are computationally efficient and effective on large datasets. Furthermore, Cheung et al. [106] introduce MODALS, a modality-agnostic automated data augmentation method that operates in latent space.

It fine-tunes four universal data transformation operations to conform to various data modalities, creating challenging examples, and employing additional loss terms to refine label-preserving transformations. It manifests effectiveness in a variety of modalities, including text, tabular, time series, and images.
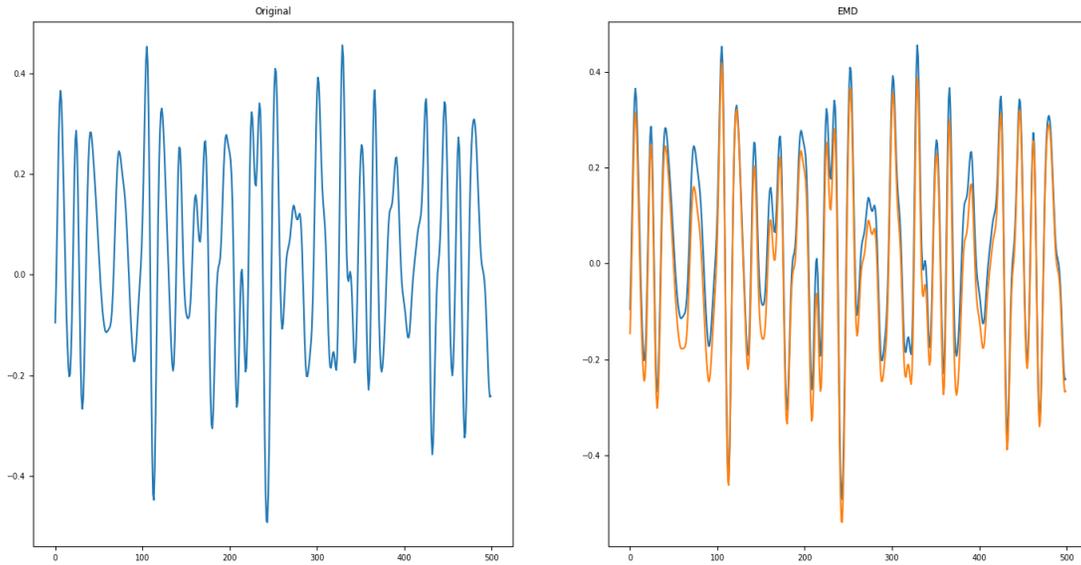
Fig. 13. Illustrative Examples of Decomposition-Based Methods (EMD) on the FordB Dataset (Generated instances are depicted in orange)

5 Experimental Study Setup

In this section, we detail the comprehensive comparative evaluations performed, using various data enhancement techniques. These evaluations are designed to provide an empirical comparison between different data augmentation strategies. The primary focus of the comparison is to assess the influence of these methods on the performance of neural network models, especially in the context of time series datasets. This approach allows for a deeper understanding of how these techniques can be optimized to enhance model performance, offering a broader view of their potential impact across various applications.

5.1 Datasets

This research encompasses the application of several time series classification techniques to various datasets from the University of California, Riverside (UCR) time series classification repository [11]. The UCR repository is a rich collection of time series datasets, extensively utilized within the academic community for benchmarking time series classification algorithms. For this study, fifteen representative categories were chosen, each representing a unique type of time series data. By including these datasets, we cover all fifteen types of classification datasets within the UCR repository, namely Device, ECG, EOG, EPG, Hemodynamics, HRM, Image, Motion, Power, Sensor, Simulated, Spectro, Spectrum, Traffic, and Trajectory. These types encompass a diverse range of domains, such as medical signals, motion capture, sensor readings, and simulated environments, among others.

The UCR repository is widely recognized for its extensive collection of time-series datasets, each belonging to a distinct category that reflects different real-world applications. This ensures that our selections are not only classic and representative but also demonstrate the broad applicability and generalizability of our experiments across diverse domains. Our goal was to delve deeper into the nuanced relationships between DA methods and datasets. Selecting fewer, but highly representative datasets, allows us to achieve this objective more effectively.

Table 2. Selected datasets from the 2018 UCR Time Series Classification Archive

Dataset	Type	Train	Test	Class	Length	Reference
CBF	Simulated	30	900	3	128	[107]
ECG5000	ECG	500	4500	5	140	[108]
FordB	Sensor	3636	810	2	500	[109]
GunPointAgeSpan	Motion	135	316	2	150	[11]
ScreenType	Device	375	375	3	720	[11]
Strawberry	Spectro	613	370	2	235	[110]
Yoga	Image	300	3000	2	426	[11]
EOGHorizontalSignal	EOG	362	362	12	1250	[111]
Fungi	HRM	18	186	18	201	[112]
GestureMidAirD1	Trajectory	208	130	26	Vary	[113]
InsectEPGRegularTrain	EPG	62	249	3	601	[114]
MelbournePedestrian	Traffic	1194	2439	10	24	[11]
PigCVP	Hemodynamics	104	208	52	2000	[11]
PowerCons	Power	180	180	2	144	[11]
SemgHandMovementCh2	Spectrum	450	450	6	1500	[115]

In Table 2, the term "Class" refers to the number of distinct categories or labels within the dataset. It reflects the granularity and diversity of events or phenomena that the dataset captures. A higher number of classes generally suggests a more complex classification task. "Length" denotes the duration of each individual time series in the dataset, representing the number of time points or observations for each sequence. This factor is crucial as longer sequences might capture more detailed dynamics but also introduce challenges in computational complexity and potential for overfitting.

All datasets were firstly rescaled to ensure that the minimum value in the training set is -1 and the maximum is 1. This normalization step helps preserve relative values and maintain consistency across different datasets. Additionally, we address missing values, if present, by substituting them with zero, a standard approach that maintains the dimensional integrity of the data while ensuring that missing values do not influence the learning process.

5.2 Evaluated Network Models

For our evaluation, we use two deep learning models: the deep convolutional neural network (CNN) model ResNet and the Long Short-Term Memory (LSTM) model. ResNet is particularly well-suited for time series data due to its ability to capture hierarchical features through convolutional layers and its robustness to overfitting via residual connections. LSTM, on the other hand, excels at learning temporal dependencies and capturing long-term trends in sequential data due to its unique memory cell architecture, which can retain information

over long periods. Both ResNet and LSTM have been widely adopted in the field of data augmentation for time series classification, serving as classic models with extensive usage and validation.

Many studies have demonstrated their superior performance compared to other models, highlighting their effectiveness in this task. In our survey, over one-third of the papers utilized ResNet, LSTM, or both to evaluate their proposed algorithms, including studies such as [3, 30, 36, 39, 49, 50, 55, 116–118]. This widespread use underscores the robustness and reliability of these models in benchmarking DA methods for TSC. Our decision to use both ResNet and LSTM stems from their prevalent usage and proven success in many renowned studies. Additionally, the unique characteristics of these models make them particularly suited for time series classification. By employing these models, we aim to provide a reliable evaluation of DA methods and are confident they will demonstrate excellent results with the methods we have collected.

Hyperparameters for the 1D ResNet used in this evaluation were informed by research conducted by Fawaz et al. 2019 and Iwana et al. 2021, where they noted improvements in time series classification using data augmentation. Similarly, the LSTM configurations were guided by studies [4, 119, 120] that have shown its effectiveness in handling temporal dependencies in time series data.

5.2.1 LSTM.

As shown in Figure 14, our implementation of the LSTM network is based on the architecture proposed by Karim et al. 2017. It incorporates both LSTM and convolutional layers to capture temporal dependencies and extract meaningful features from the time-series data. Although this setup uses LSTM in a non-standard way, it has been shown to be more effective than traditional configurations [4, 121].

The LSTM component begins by permuting the input shape to align with the multivariate time-step structure required by the LSTM layer. This layer consists of 128 units, followed by a dropout layer with a dropout rate of 0.8, which helps prevent overfitting and enhances generalization. The convolutional component consists of three 1D convolutional layers. The first layer applies 128 filters with a kernel size of 8 and uses the He uniform initializer for weights. The second layer increases the number of filters to 256, with a kernel size of 5. The third layer reduces the filter count to 128, using a kernel size of 3. Each convolutional layer is followed by batch normalization [122] and ReLU activation functions, ensuring stable training and promoting non-linearity. A Global Average Pooling (GAP) layer is used to flatten the output of the convolutional layers, and the results are then concatenated with the output from the LSTM layer. This combination allows the model to leverage both sequential information from the LSTM and spatial features extracted by the convolutional layers. The final output layer uses softmax activation to produce class probabilities, making the network suitable for multi-class time-series classification tasks.

5.2.2 ResNet.

As shown in Figure 15, our implementation of ResNet diverges from the original model [123] by incorporating only three residual blocks with varying filter lengths and omitting max pooling. The first block employs three layers of 64 1D convolutions, the second block contains three layers of 128 1D convolutions, and the third block features three layers of 128 convolutions. Each residual block has three convolutional layers with filter sizes of 8, 5, and 3, respectively. The batch normalization functions [122] and the ReLU activation functions follow each convolution. Additionally, the residual connections merge the input of each residual block with the input of the subsequent block through an addition operation. The final two layers of ResNet comprise a Global Average Pooling (GAP) layer and an output layer with softmax activation.

All network parameters were initialized using the Glorot uniform method [124]. We trained the network using the Adam optimizer [125], with an initial learning rate of 0.001 and exponential decay rates for the first and second momentum estimates at 0.9 and 0.999, respectively, following Fawaz et al.'s methodology [3]. To maintain
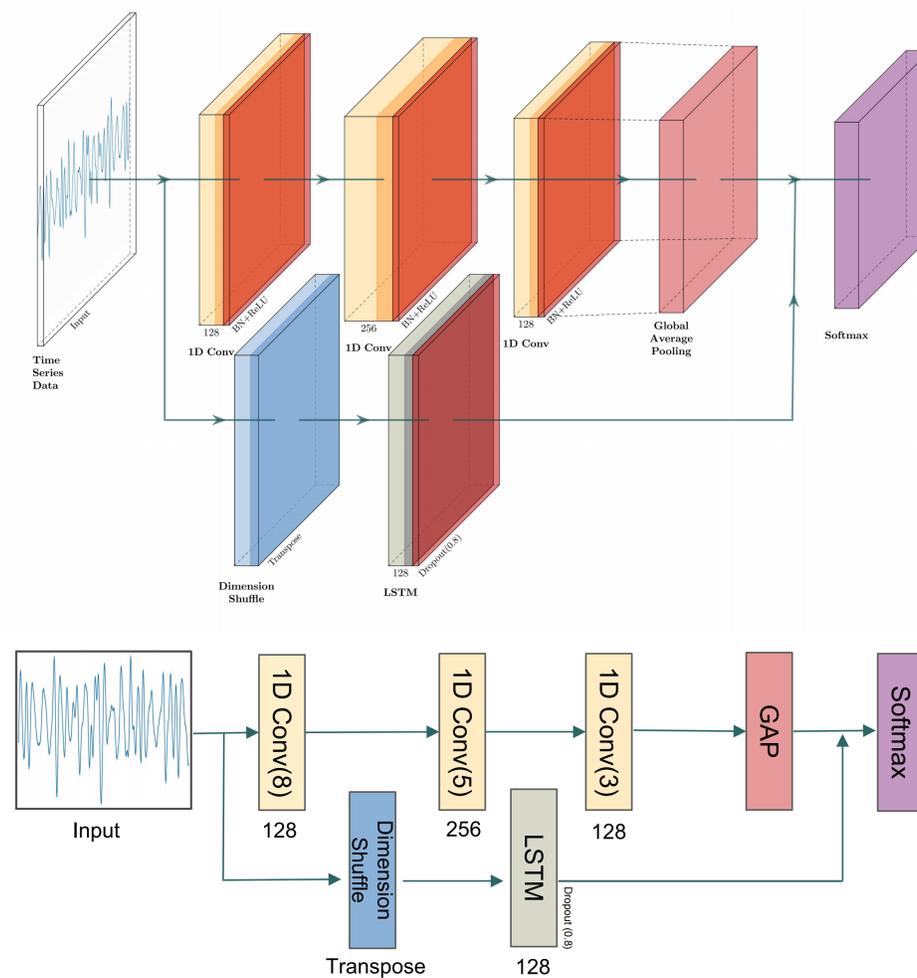
Fig. 14. The architectures of the neural networks (LSTM Fully Convolutional Network) used to evaluate the data augmentation methods

consistency in the comparison of data augmentation methods, we ensured that each model underwent the same 10,000 iterations, reducing the learning rate by 0.1 on plateaus of 500 iterations. The number of patterns in the training dataset was quadrupled compared to the original dataset. The choice of 10,000 iterations stemmed from observations that indicated that the training loss for the model on each dataset typically converged before this point.

Moreover, as the number of iterations remained constant, the control experiment without augmentation effectively repeated the same patterns to match the augmented experiments. Regarding hyperparameter selection, we adhered to the best practices documented in the literature as closely as possible, avoiding excessive adjustments to maintain a fair comparison of the data augmentation effects.
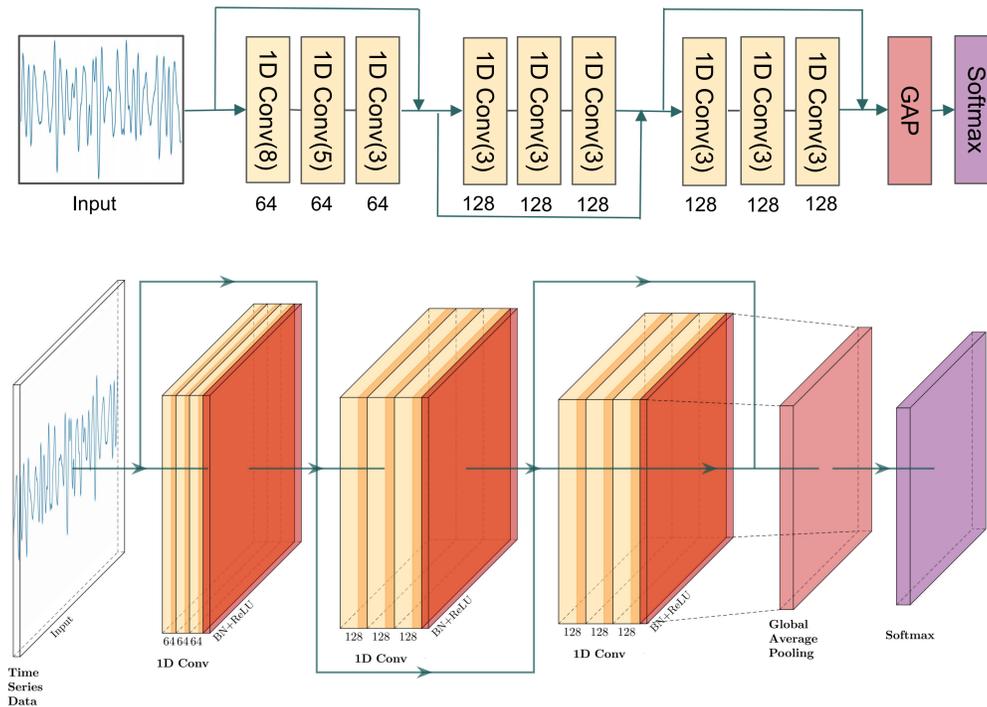
Fig. 15. The architectures of the neural networks (ResNet) used to evaluate the data augmentation methods

5.3 Evaluated Data Augmentation Methods

The data augmentation methods scrutinized in this study are delineated in Table 3. The criterion for their selection was predominantly their recurrent citation in the scholarly literature and the accessibility of their open source implementations, which facilitates reproducibility and comparative scrutiny. Amidst a plethora of techniques, some entailing elaborate combinations and complexities, the endeavor was to encapsulate those epitomizing foundational and seminal contributions in the field. This strategic choice ensures the study's resonance, relevance, and alignment with contemporary scholarly discourse and benchmarks. Each method under evaluation was configured using parameters from their respective seminal papers, thereby preserving the authenticity of their original design and ensuring an equitable evaluation landscape. These methods were evaluated based on their efficacy in enhancing the performance of both the ResNet and LSTM models across various time series datasets.

We excluded Automated Methods due to their complexity and the need for extensive parameter tuning. These methods often integrate multiple algorithms, adding complexity beyond our study's scope, which focuses on foundational algorithms. The lack of open-source resources also makes reproducibility difficult, potentially compromising reliability. Our aim is to evaluate foundational data augmentation techniques that serve as the building blocks for more advanced models, providing a comprehensive understanding of their capabilities and limitations.

Many generated data augmentation techniques require external training and pre-training, adding significant complexity to the evaluation process. Each dataset often requires tailored training, making it resource-intensive and challenging to maintain consistency across diverse datasets. Evaluating these methods also involves careful hyperparameter tuning and managing potential variability in outcomes, affecting reproducibility and reliability.

Table 3. Evaluated Data Augmentation Methods and Their Categories

No.	Data Augmentation Method	Category	Reference
1	None	-	-
2	Jittering	Transformation-Based	[6]
3	Rotation	Transformation-Based	[6]
4	Scaling	Transformation-Based	[6]
5	Magnitude Warping	Transformation-Based	[6]
6	Permutation	Transformation-Based	[6]
7	Random Permutation	Transformation-Based	[6]
8	Time Warping	Transformation-Based	[6]
9	Window Slicing	Transformation-Based	[13]
10	Window Warping	Transformation-Based	[13]
11	SFCC	Transformation-Based	[39]
12	SPAWNER	Pattern-Based	[45]
13	wDBA	Pattern-Based	[14]
14	RGW	Pattern-Based	[48]
15	RGWs	Pattern-Based	[48]
16	DGW	Pattern-Based	[48]
17	DGWs	Pattern-Based	[48]
18	DTW-Merge	Pattern-Based	[55]
19	GAN	Generative	[70]
20	EMD	Decomposition-Based	[97]

Due to these challenges and the impracticality of including all methods, we selected GANs as a representative of generated methods, as they offer a manageable balance of complexity within our evaluation framework.

5.4 Evaluation Metrics

In addressing the intricate task of evaluating algorithm reliability within the domain of time series analysis, we acknowledge the absence of a singular metric universally applicable across all conceivable applications. This context underpins our selection of evaluation metrics, aiming to bridge the gap between general applicability and the specific demands of our study. This research used a multifaceted evaluation metric system to assess the effectiveness of data augmentation methods. Each metric contributes a unique evaluative perspective, cumulatively offering a holistic insight into each method's efficacy. The metrics incorporated are as follows:

1. Accuracy (Acc): Accuracy, an intuitive and globally recognized metric, gauges the ratio of correctly predicted instances in all prognostications. We have opted for Acc as a primary metric due to its relevance in scenarios where the correct identification of both true positives and true negatives is of paramount importance. This choice is informed by the nature of our balanced dataset, where equal importance is placed on each class and the distinct types of classification errors are deemed equally significant. In scenarios where false negatives and false positives are critically important, alternative metrics such as the F1-score might be preferred; however, given the balanced

nature of our datasets and the equivalence of misclassification impacts, Acc was deemed the most appropriate metric for our analysis of data augmentation in time-series classification tasks.

2. Method Ranking: To facilitate a more nuanced comparative analysis of the data augmentation methods under investigation, we have adopted a ranking paradigm based on accuracy scores. This approach not only simplifies the identification of the most effective strategies but also establishes a clear hierarchy of performance across different datasets. A lower ranking is indicative of superior methodological performance, offering a structured framework to assess the relative success of each data augmentation strategy.

3. Residual Analysis: Residual plots were instrumental in juxtaposing the performance of each method with the baseline ("None"). Herein, a residual denotes the deviation between the actual and forecast accuracy. Positive residuals signify outperformance over the baseline, whereas negative residuals indicate underperformance. These plots are particularly suited to our study as they provide a visual conduit for assessing the relative performance increments or decrements attributable to each data augmentation strategy. By visualizing these differences, we can identify patterns and trends in the data that may not be apparent through traditional accuracy metrics alone.

This multimetric evaluation approach underpins the study's commitment to a nuanced, rigorous, and exhaustive assessment, fortifying the conclusions regarding the effectiveness of each data augmentation strategy.

5.5 Research Questions and Objectives

Navigating the complexities of time-series data augmentation necessitates a profound understanding of the subtleties and ramifications of various methods. This study is predicated on the pursuit of clarity in this domain, aiming to elucidate the efficacy and implications of various data augmentation strategies through the following research questions.

5.5.1 RQ 1: What is the impact of data augmentation on model performance?

This pivotal question seeks to explore the influence of data augmentation on model efficacy, probing the disparities in outcomes engendered by different techniques. It bifurcates into more specific inquiries:

(1) *RQ 1.1: What is the extent of the universality of data augmentation strategies?*

This inquiry is dedicated to discerning the consistency of data augmentation's effectiveness across various contexts, highlighting any discrepancies and underlying rationales for performance enhancement or degradation.

(2) *RQ 1.2: Which data augmentation strategy is the most effective for TSC?*

This probe focuses on identifying the preeminent strategy among the evaluated techniques, offering critical insights to practitioners in the time-series domain.

5.5.2 RQ 2: How do intrinsic dataset characteristics influence the effectiveness of data augmentation and subsequent recommendations?

This exploration delves into the nuanced interplay between dataset attributes and augmentation efficacy, aiming to extrapolate actionable insights for optimal method application based on specific dataset traits. The objective is to equip practitioners with informed and customized augmentation strategies, thereby maximizing model performance and circumventing the pitfalls of indiscriminate augmentation application.

The ensuing chapter, "Results Analysis", will systematically address these queries, contributing to a robust, comprehensive understanding of time-series data augmentation and its diverse implications in machine learning.

6 Results Analysis

In this section, we undertake a methodical examination of the research questions posited at the outset. Our first point of inquiry assesses the universality and optimality of various data augmentation techniques (RQ 1). Subsequently, we explore the nuanced interdependencies between the attributes of the dataset and the efficacy of these augmentation methods, providing customized guidance for practitioners (RQ 2). This discourse contributes a lucid and empirically grounded perspective on the dynamic between data augmentation strategies and time series classification.

6.1 Answer to RQ 1.1: How universal are data augmentation techniques?

Table 4. Accuracy Performance of Different Data Augmentation Methods (%)

Data Augmentation Method	ResNet (%)	Ranking	LSTM (%)	Ranking
RGWs	85.69 ± 15.12	1	83.42 ± 17.53	2
DGWs	85.65 ± 15.19	2	82.24 ± 18.63	11
Window Warping	85.55 ± 15.32	3	82.71 ± 18.13	6
Random Permutation	85.44 ± 16.02	4	83.61 ± 16.83	1
SFCC	85.43 ± 17.08	5	83.21 ± 17.38	4
DGW	85.32 ± 15.28	6	82.38 ± 18.94	10
Permutation	85.26 ± 16.58	7	83.24 ± 17.19	3
RGW	85.21 ± 15.85	8	82.61 ± 18.15	7
DTW-Merge	85.19 ± 15.67	9	82.59 ± 18.93	8
Window Slicing	85.17 ± 15.62	10	82.13 ± 18.61	12
GAN	85.15 ± 15.22	11	83.04 ± 18.57	5
None	**84.98 ± 16.41**	**12**	**82.41 ± 18.71**	**9**
wDBA	84.78 ± 16.36	13	81.48 ± 20.25	13
Time Warping	84.49 ± 15.39	14	81.47 ± 17.94	14
Scaling	84.15 ± 17.27	15	79.81 ± 23.51	15
Jitter	83.95 ± 17.52	16	78.23 ± 24.55	17
SPAWNER	83.26 ± 17.07	17	76.77 ± 24.95	18
Magnitude Warping	83.17 ± 17.75	18	79.34 ± 23.37	16
Rotation	80.21 ± 19.93	19	76.61 ± 23.46	19
EMD	76.57 ± 26.16	20	73.84 ± 28.69	20

The empirical results of diverse data augmentation strategies are encapsulated in Table 4. In the absence of any data augmentation methodology (labeled "None"), the model records an average accuracy of 84.98 ± 16.41% and 82.41 ± 18.71% on ResNet and LSTM respectively, serving as our performance standard. Intriguingly, specific strategies such as Window Warping, DGWs, and RGWs exhibit a slight enhancement in performance compared to this reference point in the ResNet model. In the LSTM model, methods such as Random Permutation, Permutation, and SFCC also demonstrate commendable performance. Notably, the RGWs method performs exceptionally well across both of the models. Conversely, certain approaches, notably EMD, incur a significant decrease in efficacy.
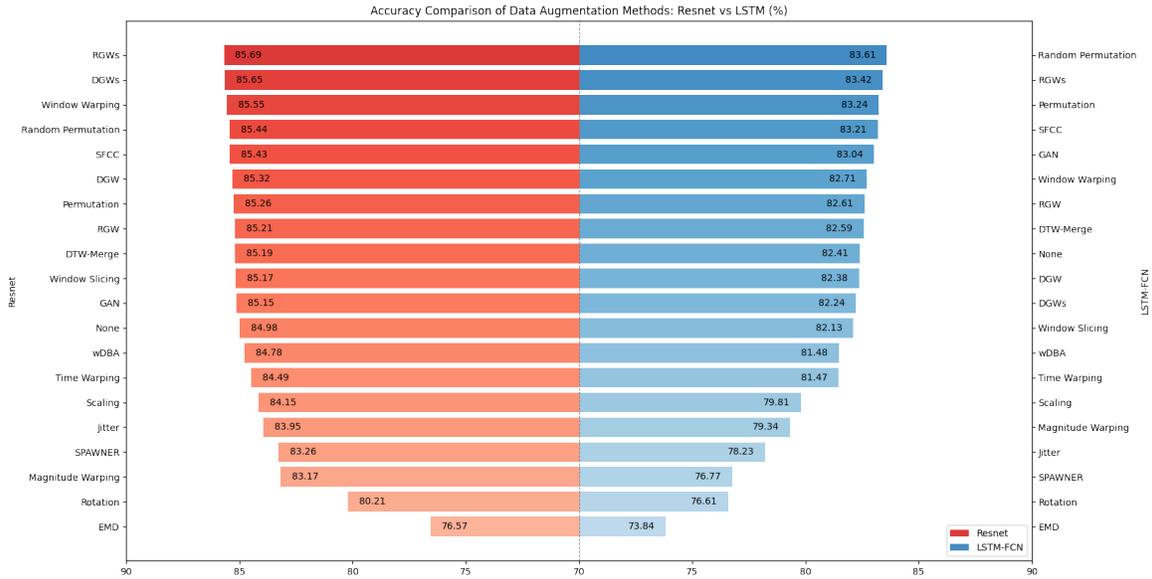
Fig. 16. Accuracy Comparison of Data Augmentation Methods: ResNet vs LSTM (%)

This observation emphatically accentuates that the benefits of augmentation are not universally applicable, emphasizing the necessity for judicious selection of data augmentation techniques. While a subset of data augmentation strategies marginally outperforms the established baseline, the distinction in efficacy between several methods and the standard remains negligible. For instance, methodologies like Window Slicing, DTW-Merge, and GAN in ResNet show marginal deviations, suggesting that the advantages conferred by some data augmentation techniques may be nuanced and potentially contingent on specific contexts.

6.2 Answer to RQ 1.2: Which data augmentation technique is optimal for TSC?

This subsection first reports the empirical ranking of twenty data augmentation techniques on ResNet and LSTM and then sketches the defining traits of each method family—Transformation, Pattern, Generative, and Decomposition—to orient the detailed discussion that follows.

6.2.1 Transformation-Based Methods.

Transformation-Based methods concentrate on subtle modifications to the original data's structure, avoiding the creation of entirely new instances. Techniques encompassing Permutation, Scaling, Warping, and Rotation are categorized within this domain.

Referring to Table 4, both Permutation and Random Permutation diversify the dataset by rearranging time series segments, avoiding extreme modifications that could distort the data's essential structure. Similarly, Magnitude Warping and Scaling perform elastic transformations, ensuring that the fundamental data features remain unaltered. The success of these strategies reaffirms the criticality of preserving the innate temporal dynamics in time-series data while introducing nuanced variations.

Further experimental results in Figure 16 reinforce the efficacy of the Transformation-Based methods. In the ResNet model, methods such as Window Warping, Random Permutation, SFCC, Permutation, and Window Slicing all surpass the baseline, with Window Warping, Random Permutation, and SFCC achieving accuracies of 85.55%,

Table 5. Average Ranking of Data Augmentation Methods in each dataset

No.	Data Augmentation Method	Category	ResNet	Best	LSTM	Best
1	Permutation	Transformation-Based	7.07	2	7.33	2
2	Random Permutation	Transformation-Based	7.93	0	7.13	1
3	DGWs	Pattern-Based	8.40	0	9.40	1
4	SFCC	Transformation-Based	8.60	4	7.93	1
5	RGWs	Pattern-Based	9.07	0	7.13	0
6	None	-	9.11	2	8.93	2
7	Window Warping	Transformation-Based	9.47	1	8.87	0
8	DTW-Merge	Pattern-Based	9.47	0	8.93	3
9	DGW	Pattern-Based	10.07	0	9.80	0
10	Scaling	Transformation-Based	10.73	0	10.20	0
11	RGW	Pattern-Based	10.80	1	10.40	0
12	Jittering	Transformation-Based	11.00	0	13.27	1
13	GAN	Generative	11.07	0	7.53	2
14	Window Slicing	Transformation-Based	11.40	1	11.27	2
15	Magnitude Warping	Transformation-Based	11.80	0	10.40	0
16	wDBA	Pattern-Based	12.20	0	11.87	0
17	Rotation	Transformation-Based	12.47	2	15.27	0
18	Time Warping	Transformation-Based	13.13	0	11.87	0
19	SPAWNER	Pattern-Based	13.20	1	16.67	0
20	EMD	Decomposition-Based	14.67	1	15.80	0

85.44%, and 85.43% respectively, placing them among the top five techniques. These methods, specifically Window Warping, Random Permutation, and Permutation, belong to the "Time" branch of Transformation-Based methods, illustrating their capacity to adaptively modify time series data while maintaining its intrinsic properties.

In the LSTM model, the Transformation-Based methods continue to demonstrate robust performance, with Random Permutation leading at an accuracy of 83.61%, followed by Permutation and SFCC, all exceeding the performance of the baseline. The success of SFCC, categorized under the "Frequency" branch, highlights the potential of frequency-based transformations to enhance model training without compromising the integrity of the time series data.

These findings indicate that while traditional time manipulation techniques remain effective, incorporating frequency-based augmentations like SFCC can provide a complementary approach to enriching the data representation for time series classification.

6.2.2 Pattern-Based Methods.

Pattern-Based methods focus on extracting or amalgamating patterns from the original data, generating new instances that may introduce novel structural elements.

Despite their potential, Pattern-Based Methods have shown mixed performance in our evaluation. Remarkably, in the ResNet model, the "Time" branch of Pattern-Based methods such as RGWs and DGWs achieved the highest accuracies, ranking first and second with scores of 85.69% and 85.65% respectively. This superior performance underscores the effectiveness of these methods in leveraging temporal patterns without distorting the fundamental

time series structure. Additionally, DTW-Merge also performed above the baseline, further highlighting the potential of Pattern-Based methods to enhance classification accuracy through sophisticated pattern synthesis.

However, the SPAWNER technique from the "Magnitude" branch, which merges patterns and infuses noise, although creative, has shown limitations, ranking fourth from last in the ResNet model. This outcome could be attributed to its approach potentially introducing excessive convolution to the data, which might obscure rather than clarify the underlying patterns in the time series.

In the LSTM model, the situation is somewhat different. RGWs again performed exceptionally well, securing the second position with an accuracy of 83.42%. In contrast, methods like DGW and DGWs did not surpass the baseline in the LSTM model. This suggests that Pattern-Based methods, which modify sequence structures, may disrupt the continuity that LSTM networks rely on to capture long-term dependencies. LSTM are sensitive to the order and coherence of time-series data, and altering these patterns could obscure crucial information, leading to reduced accuracy. This indicates that while effective in ResNet, Pattern-Based methods may not align as well with the sequential learning nature of LSTM architectures. This variability suggests that these methods might be sensitive to the specific dynamics and characteristics of the dataset, such as inherent correlations or unique motifs, which could be disrupted by the more aggressive transformations employed by some Pattern-Based methods.

The relatively lower performance of SPAWNER, ranking third from last in the LSTM model, corroborates the notion that the synthesis of patterns, while innovative, risks introducing artifacts that deviate too far from the original data configurations, potentially complicating the model's ability to learn the essential features of the data.

These findings suggest a nuanced view of Pattern-Based methods: while they can offer substantial benefits in certain contexts, particularly in models like ResNet that may be better able to handle complex transformations, their application needs to be carefully tailored to the specific characteristics of the data and the model architecture to avoid detrimental effects on performance.

6.2.3 Generative Methods.

Generative methods, including models such as GANs, are used to generate new samples that, while structurally divergent from the original data, are designed to be statistically consistent.

In our analysis, GAN, representing Generative Methods, achieved a middle-ground ranking. Despite the ability of GANs to generate diverse datasets, the intricacies of time-series data present a formidable challenge, particularly in maintaining temporal coherence while fostering variations. In the ResNet model, GAN performed only slightly better than the "None" baseline, ranking just one position higher. This modest improvement suggests that while GANs can introduce beneficial variability, the general-purpose GAN model used in our experiment may not be finely tuned to the specific demands of time-series data in the ResNet architecture.

Conversely, in the LSTM model, GAN achieved a more favorable position, ranking 5th with an accuracy of 83.04%. This result could be attributed to LSTM's inherent capability to handle sequential data, where the diversity introduced by GAN might have helped in capturing more nuanced patterns, thus enhancing performance.

However, it is important to note that due to resource constraints, the GAN model we employed was a relatively generic implementation. The performance of GANs can be significantly influenced by the extent of task-specific training and fine-tuning, particularly for time-series data. Specialized GAN models, as highlighted in our taxonomy, have shown promise for time-series generation, but they require extensive optimization to achieve peak performance. Despite the current limitations, GAN remains a highly innovative and promising method, particularly for its potential in scenarios with limited data availability.

6.2.4 Decomposition-Based.

Empirical Mode Decomposition (EMD) is an adaptive method designed to handle nonlinear and non-stationary

signals by decomposing them into a finite number of Intrinsic Mode Functions (IMFs) and a residual. However, in our study, the performance of EMD was suboptimal, ranking the lowest in both the ResNet and LSTM models with accuracies of 76.57% and 73.84%, respectively.

Theoretically, EMD's adaptive nature should make it particularly suitable for complex time series signals. Indeed, in applications, such as noise classification, EMD has demonstrated effective performance. However, its performance in time series classification tasks appears to be limited by several factors. Firstly, while the decomposition process can effectively remove high-frequency noise, it may also result in the loss of useful information, especially when critical features are decomposed into higher-order IMFs. In our implementation, the strategy of selecting the first two IMFs may lead to information loss in certain cases, thereby affecting the model's performance.

Moreover, the signals generated by EMD are a combination of multiple IMFs, and the interactions between these IMFs can lead to augmented data that significantly deviates from the original data. This deviation could reduce the generalization capability of the classifier. Overall, while EMD shows potential in processing complex signals, its performance in time series classification is constrained by several factors. Future work could explore combining EMD with other signal processing techniques or improving the EMD implementation to enhance its applicability in time series data augmentation.

6.2.5 Detailed Analysis of Top-Performing and Under-Performing Data Augmentation Techniques.

In this subsection, we provide an in-depth examination of the data augmentation techniques that performed exceptionally well, as well as those that underperformed in our experiments. This analysis aims to highlight the strengths and limitations of each method, offering insights into their applicability and effectiveness in time series classification tasks.

(1) Top-Performing Data Augmentation Methods

Our findings, as shown in Table 5 of the average ranking and Figure 17 of the result's residual analysis, reveal that Permutation, Random Permutation, SFCC, and RGWs surpass our benchmark performance well.

The consistent presence of these methods at the top is not solely a testament to their effectiveness but also indicative of their inherent capability to maintain and potentially amplify crucial data features during augmentation. These methods introduce coherent variations aligned with the underlying data structure, thereby enhancing model training fidelity without straying from the original dataset's essence. Their reliability suggests broader applicability across various datasets and experimental conditions and their significance in achieving stable and predictable results.

In particular, SFCC, which ranks fourth overall with an average ranking of 8.60, is notable for its innovation in augmenting time series data by combining stratified Fourier coefficients. This method excels in maintaining the integrity of the original data's frequency components while introducing variability, leading to significant performance improvements. For instance, SFCC achieved the best accuracy in 4 out of 15 datasets, the highest count among all methods.

RGWs also stands out as a top performer, ranking first and second on average with a strong presence in ResNet and LSTM models. This method's focus on preserving the shape of the time series rather than exact temporal alignment ensures that the augmented data remains true to the original structure, which is critical for time series classification.

Permutation and Random Permutation, which can be considered as variants of a similar strategy, demonstrate their strength by ranking first and second in average rankings across all the datasets as shown in Table 5, with 7.07 and 7.93 in ResNet, as well as 7.33 and 7.13 in LSTM. These methods excel at diversifying the dataset by
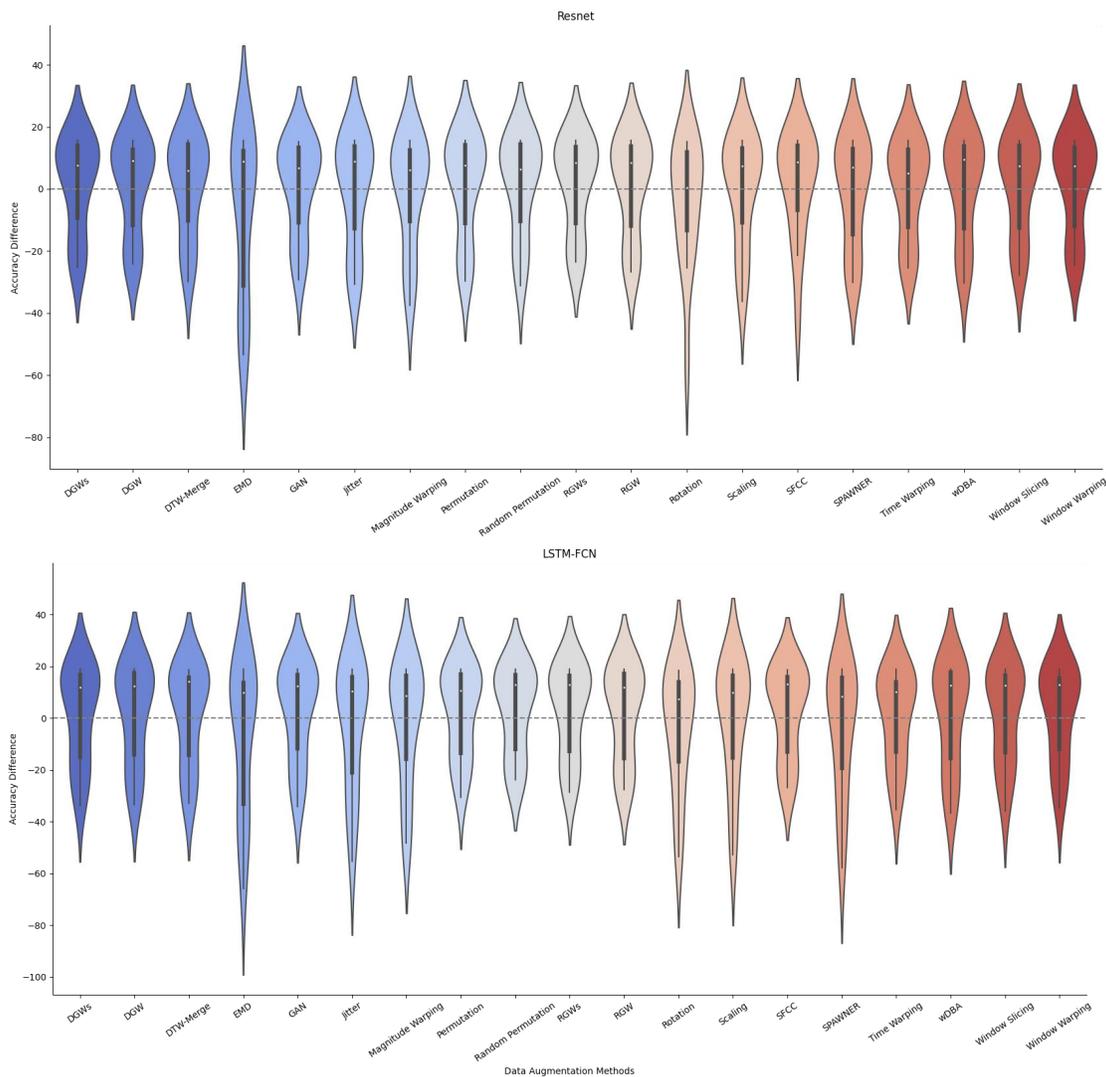
Resnet

LSTM-FCN

Fig. 17. Residual Analysis (Compared to "None") for each Data Augmentation Method

rearranging time series segments in a way that retains the sequence's essential characteristics. By doing so, they enhance the model's generalization capability and reduce the risk of overfitting, which is crucial for achieving robust performance across various tasks.

The success of these top-performing methods underscores the importance of strategies that balance the introduction of variability with the preservation of fundamental data characteristics. These methods not only improve model accuracy but also offer insights into how different types of augmentations can be optimally applied to various time series classification tasks.

(2) Under-Performing Data Augmentation Methods

Conversely, Rotation, SPAWNER, and EMD are identified as underperforming data augmentation techniques across the models we evaluated. These methods not only exhibited lower accuracy but also demonstrated significant variability in performance, indicating a lack of robustness and generalizability across different datasets.

Rotation and EMD consistently ranked among the lowest-performing techniques in both the ResNet and LSTM models. Their substantial fluctuation in performance, as revealed by residual analysis, suggests that these methods might struggle to maintain data integrity across diverse datasets. The inconsistent results indicate that these methods may not generalize well and might be more suitable for very specific use cases. Rotation, for example, could be more effective in the domain of image data, where altering orientation might introduce beneficial variability. However, in the context of time series classification, the disruption of natural sequence order may obscure temporal dependencies that are critical for accurate modeling. Similarly, EMD, which aims to decompose the signal into intrinsic mode functions, might inadvertently disrupt the underlying structure of the time series data, leading to a degradation in model performance. The high variability observed with EMD further emphasizes its potential instability, possibly due to its sensitivity to the characteristics of individual datasets.

SPAWNER, while slightly outperforming EMD, still exhibits considerable underperformance. Its hybrid approach, which merges patterns and introduces noise, may result in overly complex or artificial sequences that do not align well with the natural data patterns. This complexity can create challenges for the model in learning meaningful patterns, thus leading to poorer classification outcomes.

These underperforming techniques highlight the importance of selecting appropriate data augmentation strategies that are well-aligned with the nature of the time series data. Their limited success underscores the necessity of careful consideration and potentially more sophisticated design or tuning when applying these methods in time series classification tasks.

In response to RQ 1 on the universality and optimality of data augmentation techniques for time series classification: The performance of data augmentation methods is diverse, with RGWs and Random Permutation emerging as consistently strong performers. These methods have proven effective across various datasets, with SFCC also showing notable performance, particularly excelling in four datasets where it achieved the best results. On the other hand, techniques like Rotation, SPAWNER, and EMD significantly underperformed, with high variability and poor generalization across different datasets. While some of these methods offer robust improvements in certain dataset, others require more cautious application, often needing to be tailored to specific dataset characteristics to be effective.

6.3 Answer to RQ 2: Influence of dataset characteristics on augmentation effectiveness and user recommendations

The realm of time series classification presents perennial challenges, particularly given the heterogeneous nature of datasets. Our experimental phase has shed light on the nuanced responses of various UCR datasets to different augmentation techniques. This discussion aims to unravel the complex interplay between the attributes of the dataset and the efficacy of the augmentation strategies.

6.3.1 Dataset Characteristics Analysis.

This subsection delves into the impact of specific dataset characteristics on the performance of different data augmentation techniques, offering insights into how these factors influence the effectiveness of various methods.
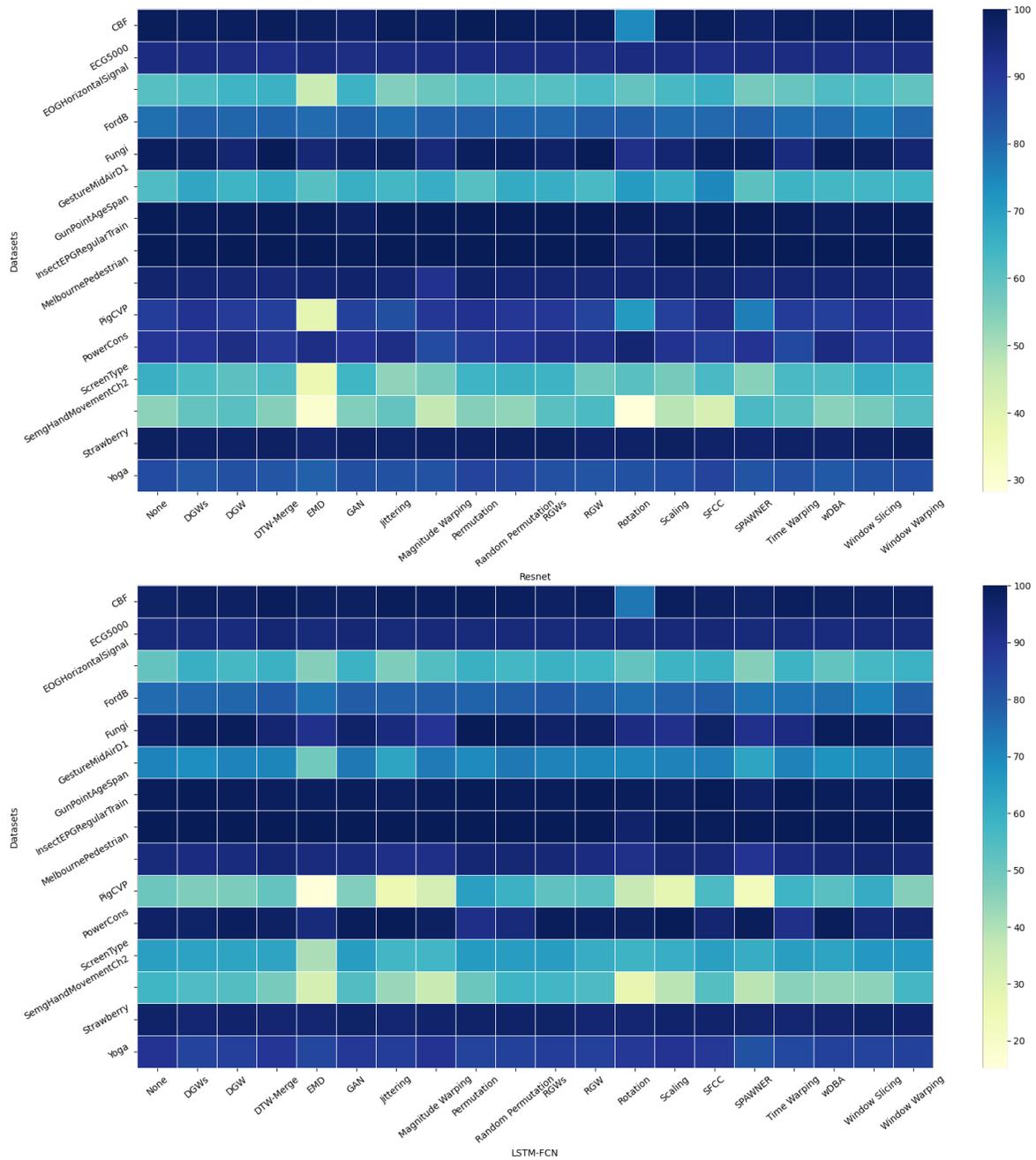
Fig. 18. Accuracy Heat Map of Data Augmentation Techniques across all the Datasets

(1) Length of Time Series

Datasets with extended time series, such as PigCVP, with the longest time series among the three datasets, reveals a notable difference in data augmentation method performance between the ResNet and LSTM models. In the ResNet model, most data augmentation methods, including SFCC, DTW-Merge, and Window Warping, outperform the baseline, with SFCC achieving the highest accuracy at 92.98%. For longer time series, ResNet benefits from data augmentation techniques that introduce variability while preserving temporal coherence. On the other hand, the LSTM model shows a more mixed performance. While Permutation and Random Permutation significantly outperform the baseline, methods like DTW-Merge and SFCC still surpass the baseline but with a smaller margin.

A similar trend is observed in the SemgHandMovementCh2 dataset, where ResNet again leverages data augmentation methods effectively, but LSTM struggles more. Despite methods like RGWs and Random Permutation outperforming the baseline, others such as SFCC and DTW-Merge fall short, likely due to LSTM's sensitivity to sequence disruptions in longer, complex time series. This suggests that LSTM, which heavily relies on sequential data, might be less robust to disruptions in temporal coherence introduced by certain data augmentation methods, especially in long time series.

The MelbournePedestrian dataset, with its short time series, allows various data augmentation methods to perform consistently well. The reduced length of the sequences likely diminishes the risk of overfitting, enabling data augmentation methods to introduce variability effectively while maintaining the integrity of the data structure. Consequently, both models achieve high accuracy, reflecting the ease with which temporal coherence can be preserved in shorter sequences.

The performance differences across these datasets highlight the significant impact of time series length on the effectiveness of data augmentation methods. Longer time series, like those in PigCVP and SemgHandMovementCh2, benefit more consistently from data augmentation methods due to the increased complexity and variability these methods introduce. Conversely, shorter time series, such as in the MelbournePedestrian dataset, show more uniform benefits across data augmentation techniques, suggesting that time series length is a crucial factor in determining the effectiveness of data augmentation strategies.

(2) Dataset Type and Inherent Noise

The performance of different datasets in ResNet and LSTM models reveals important insights into how the type of dataset and inherent noise characteristics influence the effectiveness of data augmentation methods.

As shown in Table 6, datasets with highly regular patterns, such as GunPointAgeSpan and CBF, generally achieve high classification accuracy across both models, often exceeding 97%. These datasets benefit from their structured nature, which contains clear, distinguishable patterns with minimal noise. The predictability and consistency in these datasets make them less vulnerable to disruptions caused by data augmentation methods.

In contrast, datasets derived from biometric or health-related domains, such as ECG5000 and EOGHorizontalSignal, present a more complex challenge. While ECG5000 still maintains high accuracy due to its regularities, EOGHorizontalSignal struggles, particularly in the LSTM model, where accuracy significantly drops. This discrepancy is likely due to the higher noise levels and variability inherent in these datasets, which can complicate the model's ability to capture consistent temporal patterns, particularly for methods that rely heavily on sequence integrity.

FordB and ScreenType datasets, representative of sensor-based and device-collected data, exhibit moderate performance across the models. Sensor data, by nature, can be prone to noise and irregularities, which may obscure the underlying patterns that models need to learn. The moderate accuracy observed in these datasets

suggests that the inherent variability and potential noise from sensors challenge the models, making it harder to maintain high accuracy even with the application of data augmentation methods.

SemgHandMovementCh2 and GestureMidAirD1, datasets involving motion or trajectory data, show varied performance, particularly in LSTM, where SemgHandMovementCh2 records notably low accuracy. Motion and trajectory data typically involve complex temporal and spatial dependencies, which can be difficult for models to process effectively, especially when disrupted by noise or sequence alterations. This complexity and potential noise make these datasets challenging for accurate classification.

Lastly, the MelbournePedestrian dataset, classified under traffic data, consistently shows high performance in both ResNet and LSTM models. The regular temporal patterns present in traffic data, due to the structured nature of pedestrian movement, likely contribute to this high accuracy. Despite the potential introduction of noise, the strong and consistent underlying signal in this dataset enables models to maintain high accuracy, highlighting the importance of regularity in the data for successful application of data augmentation methods.

The characteristics of the dataset, such as regularity, noise levels, and temporal complexity, play a crucial role in determining the success of data augmentation techniques. Datasets with structured, predictable patterns generally allow for better performance across different models and data augmentation methods, while those with high variability and noise require more careful consideration in the application of augmentation strategies.

Table 6. Average Accuracy Performance of Selected Datasets from the 2018 UCR Time Series Classification Archive

No.	Dataset	Type	ResNet Accuracy	LSTM Accuracy
1	InsectEPGRegularTrain	EPG	**99.82%**	99.77%
2	GunPointAgeSpan	Motion	**99.61%**	99.16%
3	PowerCons	Power	91.14%	**97.13%**
4	CBF	Simulated	**97.67%**	97.10%
5	Strawberry	Spectro	**97.87%**	96.34%
6	Fungi	HRM	**97.55%**	95.80%
7	ECG5000	ECG	93.68%	**93.98%**
8	MelbournePedestrian	Traffic	**95.94%**	93.48%
9	Yoga	Image	84.95%	**86.68%**
10	FordB	Sensor	**80.39%**	77.03%
11	GestureMidAirD1	Trajectory	65.28%	**69.16%**
12	ScreenType	Device	60.22%	**61.61%**
13	EOGHorizontalSignal	EOG	**60.50%**	55.48%
14	SemgHandMovementCh2	Spectrum	**53.24%**	47.69%
15	PigCVP	Hemodynamics	**85.60%**	45.44%

(3) Number of Classes

The diversity of classes within a dataset also impacts the augmentation options. From our analysis, datasets with fewer classes generally exhibit better performance across both of the models. This trend is particularly evident in datasets like InsectEPGRegularTrain, GunPointAgeSpan, and Strawberry, all of which contain fewer than three classes and consistently rank among the top three in these models. The simplicity in class structure likely allows these models to focus more effectively on distinguishing the limited number of patterns, leading to higher accuracy.

Conversely, datasets with a larger number of classes tend to perform less effectively. GestureMidAirD1 and PigCVP, which have the highest number of classes among the datasets analyzed, show notably lower accuracies. Both models struggle with GestureMidAirD1, achieving less than 70% accuracy, which indicates a difficulty in distinguishing between the numerous and potentially overlapping classes. The complexity introduced by a high number of classes may dilute the effectiveness of augmentation techniques, making it harder for models to learn distinct patterns within each class. Interestingly, the PigCVP dataset presents an anomaly. While the LSTM model performs poorly on this dataset, with an accuracy around 45%, the ResNet model achieves a significantly higher accuracy of approximately 85% (although ranked 10th). This stark contrast could be attributed to the inherent differences in how these models process temporal data. ResNet, with its ability to capture hierarchical patterns through convolutional layers, might be better suited to handle the diverse and complex patterns within the PigCVP dataset, even when the number of classes is high. In contrast, the LSTM model, which relies heavily on sequential data processing, might struggle to maintain consistency across such a varied and class-heavy dataset, leading to its lower performance.

These findings underscore the impact of the number of classes on model performance and highlight the need for tailored data augmentation strategies that consider the class distribution within the dataset.

6.3.2 Recommendations for Data Augmentation Practice.

This subsection sketches a pragmatic escalation strategy for data augmentation, advising practitioners to start with lightweight, structure-preserving techniques, adopt richer synthetic methods only when data complexity or scarcity requires them, and reserve resource-intensive or automated pipelines for specialized expert-level scenarios.

(1) Transformation-Based Methods

Transformation-Based methods are advisable for datasets where preserving the inherent data structure is critical. These methods maintain the core attributes of the original data while introducing controlled alterations that do not disrupt the fundamental sequence or pattern integrity.

These methods are not only critical for preserving pattern-centric time series data but are also relatively simple and intuitive to apply. Their operation principles are well-established, particularly in both image and time series domains, where they have gained widespread application. These methods are computationally efficient, offering quick execution times with minimal resource consumption, making them an excellent starting point for practitioners.

For users, there are many readily available tools and libraries beyond those presented in this paper, which include these fundamental algorithms, making them easy to implement. Based on our experimental results, methods like Permutation, Random Permutation, and SFCC consistently demonstrated strong performance, which showed average accuracy improvements of 3-5% over the baseline, and should be considered as primary candidates for initial attempts. However, users should note that certain methods, like Rotation, which are effective

in computer vision, may not translate well to time series data due to the nature of sequential dependencies. Additionally, frequency domain methods may not be suitable for non-periodic time series, which requires careful consideration during application.

(2) Pattern-Based Methods

Pattern-Based methods are invaluable for datasets that are feature-rich or when the introduction of new structural elements could meaningfully enhance the dataset. These methods excel at synthesizing variations that can introduce new, potentially beneficial patterns into the data, offering the model additional perspectives that it might not encounter through Transformation-Based methods alone.

Practitioners should carefully assess the complexity of their dataset before applying Pattern-Based methods. For datasets with already rich feature sets, such as those with intricate temporal or spatial structures, these methods can be particularly powerful. Techniques like RGWs, which performed in the top 30% on 80% of evaluated datasets, and DGWs, which excel at introducing elastic deformations, were shown to enhance performance by enabling models to become more robust to slight variations and distortions in time series data. However, practitioners should be cautious of overcomplicating simpler datasets with these methods, as adding unwarranted complexity could lead to overfitting or a dilution of the signal's clarity.

They are especially beneficial when the goal is to augment the data with synthetic variations that maintain the underlying structure while offering new insights. Given their computational demands, it may be advisable to use these methods selectively, perhaps after initial trials with simpler methods like Transformation-Based techniques. Furthermore, it is important to note that the effectiveness of Pattern-Based methods can vary greatly depending on the inherent noise and variability within the dataset. In noisy datasets or those with highly variable patterns, these methods may need careful calibration to avoid introducing excessive complexity that could confuse the model rather than enhance its learning.

(3) Generative Methods

Generative Methods, such as GANs and VAEs, are powerful tools for augmenting datasets lacking in volume or diversity. These techniques can generate new samples, potentially expanding the dataset significantly. However, it's essential to ensure that the generated samples align with the original data's statistical properties.

While Generative Methods can offer substantial benefits by significantly expanding the dataset, their application is not as straightforward or universally applicable as other augmentation techniques. These methods often require extensive external training and fine-tuning tailored to the specific characteristics of the target dataset. For instance, in our experiments, we utilized a standard, generalized GAN, which, while effective to some extent, did not achieve the level of performance that might be possible with a more specialized generative model.

For specialized tasks, where the generative model is meticulously designed and trained, these methods can outperform traditional augmentation techniques, providing innovative solutions that can produce results unattainable by simpler methods. However, the complexity of Generative Methods cannot be understated. The generative model itself may require extensive computational resources, prolonged training periods, and sophisticated model design. Moreover, the availability of ready-to-use tools for these methods is limited, often necessitating custom implementations that are not as user-friendly or accessible as Transformation-Based or Pattern-Based methods.

Practitioners should consider starting with simpler augmentation techniques and move to generative methods as their understanding of the dataset deepens. For specialized tasks, where tailored generative models are crucial, these methods can provide unparalleled results, though they come with higher complexity and resource demands.

(4) Decomposition-Based Methods

Decomposition-Based Methods, despite its potential, is not as widely used as other methods due to its complexity and specific requirements for the dataset. Our experiments with a general EMD configuration demonstrated subpar performance, ranking last with average ranks of 14.67 and 15.80, suggesting that this method may not be well-suited for broad application unless there is a particular need or a deep familiarity with methods like EMD.

Given the relatively poor results observed in our study, we recommend practitioners to approach Decomposition-Based Methods with caution. It may be more suitable for niche applications where the decomposition of signals is critical, but for most time series tasks, simpler and more established methods are likely to yield better results with less complexity. Unless there is a clear justification for its use, or a strong understanding of how to optimize those for a specific dataset, it may be advisable to prioritize other data augmentation techniques that offer greater flexibility and proven effectiveness across a wider range of tasks.

(5) Automated Data Augmentation Methods

Automated Data Augmentation Methods often involve combining or stacking multiple basic augmentation techniques to optimize performance. These methods are not covered in our experiments, as we focused on evaluating individual methods. Automated approaches typically require setting weights and fine-tuning various sub-methods to achieve the best results, demanding a deep understanding of each data augmentation technique. Therefore, they may not be suitable for beginners or those who aim to use data augmentation methods without extensive customization.

While these methods hold significant potential, especially for complex and large-scale tasks, they are best employed by users with substantial expertise in data augmentation techniques. For those dealing with particularly intricate datasets, the investment in designing and fine-tuning an automated approach may be justified. However, for most standard tasks, simpler, well-established methods are likely more practical and efficient.

6.3.3 Navigating Complex Scenarios in Data Augmentation Selection.

In addressing complex scenarios encountered by practitioners when selecting appropriate data augmentation methods for time series classification, it becomes evident that multifaceted dataset characteristics introduce additional layers of uncertainty. This complexity can sometimes lead to conflicts or confusion when attempting to apply the guidelines provided earlier in our study. In such intricate cases, our initial recommendation leans towards the employment of Transformation-Based methods as an initial attempt. These methods have demonstrated a robust and generally effective approach accompanied by minimal resource and time consumption across various datasets in our experiments, making them a reliable starting point.

However, when faced with datasets that present unique challenges, such as a high number of classes coupled with short series length, the choice between preserving pattern integrity and capturing class-specific patterns requires careful consideration. For datasets where the "Length of Time Series" is a distinctive and critical characteristic, prioritizing this feature may guide the selection of data augmentation techniques that respect temporal dynamics without compromising data integrity.

Practitioners are encouraged to experiment with more advanced strategies, such as DGW for nuanced class differentiation, or even generative methods like GANs for enriching dataset diversity. It is crucial, however, to assess the inherent traits of the dataset and determine which aspects are most vital for the task at hand. For instance, if class distinction is paramount, methods enhancing inter-class variability should be prioritized, whereas datasets with a pronounced "Length of Time Series" might benefit more from techniques that efficiently utilize the available temporal information without distorting the original patterns.

The selection of data augmentation methods in complex scenarios demands a balanced approach, integrating empirical evidence from experiments with a nuanced understanding of the dataset's characteristics. By initially leaning on the more conservative, transformation-based techniques and gradually integrating more specialized methods, practitioners can navigate the uncertainties inherent in complex data augmentation landscapes, optimizing their approach to suit the unique demands of their datasets.

In response to RQ 2 on the influence of dataset characteristics on augmentation effectiveness and user recommendations: Our analysis shows that the length of time series, the number of classes, and the inherent noise in datasets significantly impact the performance of data augmentation methods, with ResNet and LSTM models responding differently to these factors.

We recommend that users tailor their augmentation strategies to the specific characteristics of their datasets, favoring Transformation-Based Methods for maintaining data integrity, Pattern-Based Methods for feature-rich datasets, Generative Methods for expanding limited datasets with caution, and Decomposition-Based Methods only when there is specific expertise or necessity. Automated Data Augmentation, while powerful, should be reserved for complex, large-scale tasks where fine-tuning and expert knowledge are available.

7 Limitations and Threats to Validity

Every empirical investigation, regardless of the meticulousness in its execution, has certain limitations. Acknowledging these constraints not only bolsters the study's credibility, but also paves the way for subsequent, more refined research. In this discourse, we deliberate on the inherent limitations of our study and the strategies used to minimize their implications.

7.1 Internal Validity Threats

This subsection scrutinizes potential biases and confounding factors intrinsic to our experimental setup, ranging from dataset coverage and technique breadth to hyper-parameter choices and model dependence—and explains the safeguards we adopted to curb each risk. Clarifying these internal constraints helps delineate the conditions under which our empirical claims remain sound.

7.1.1 Limitation of Data Sources.

The primary resource for our study was the UCR time series classification archive. Although this repository is comprehensive and esteemed within scholarly circles, it encompasses merely a fraction of all conceivable time series datasets. To counterbalance this limitation, our selection of datasets from the UCR archive was intentionally diverse, spanning various categories of time series data, from sensor readings to medical diagnostic data, contains all 15 different types in the database.

7.1.2 Scope of Data Augmentation Techniques.

Our evaluation encompassed several data augmentation strategies, yet the expanse of existing and emerging methods is vast. In order to address this, our research prioritized methods that have gained recognition and widespread application. Furthermore, we adopted an inclusive approach by utilizing open-source implementations where feasible, establishing a unified framework conducive to future expansions and adaptations.

7.1.3 Influence of Hyperparameter Configurations.

The efficacy of data augmentation strategies is often dependent on hyperparameter configurations. Our adherence to parameters endorsed in the scholarly literature does not exclude the potential for alternative configurations to engender different outcomes. To counterbalance this limitation, we adopted hyperparameters validated by

previous empirical studies, ensuring that our findings are anchored in established methodologies and facilitating comparability with existing literature.

7.1.4 Model Specificity.

Our analyses relied heavily on the LSTM and the ResNet architectures. The implications of using alternative models remain an open question. The selection of ResNet as well as LSTM was informed by its documented efficacy in time series classification tasks. We recognize the plethora of alternative models and advocate for additional studies to determine the impact of various data augmentation techniques across diverse neural network architectures.

7.2 External Validity Threats

This subsection delineates the principal factors that could constrain the transferability of our findings to other datasets, tasks, and temporal contexts. By foregrounding these external validity threats, we equip readers with a clear framework for judging when and how the reported insights can be responsibly generalized.

7.2.1 Generalizability Concerns.

The applicability of our conclusions beyond the datasets and methods employed in our study may be subject to scrutiny. In order to address this, our methodological design, characterized by a diverse selection of datasets and a wide range of data augmentation techniques, was geared toward maximizing the generalizability of our findings. However, we advise prudence in extending these conclusions to substantially dissimilar scenarios.

7.2.2 Temporal Context of Tools and Methodologies.

The machine learning landscape is in a state of perpetual evolution, with constant refinements in tools, methodologies, and best practices. This study captures the state of the art at the time of writing. While our approach was exhaustive given the current landscape, impending advancements may necessitate new interpretations or adaptations of our findings.

7.2.3 Potential Publication Bias.

Our literature review may reflect a propensity towards methodologies that have demonstrated success or novelty, inadvertently marginalizing equally potent but underreported techniques. To counterbalance this limitation, we endeavored for a holistic literature review, although we concede the potential for publication bias. Therefore, we endorse the exploration of lesser-known methodologies and gray literature in future scholarly endeavors.

In summary, our study, despite its inherent limitations, was conducted with a rigorous adherence to scholarly standards. By candidly addressing potential validity threats, we aspire to furnish a robust platform for subsequent inquiries in the realm of time series data augmentation.

8 Conclusions and Prospects for Future Research

Building upon the analyses presented earlier, this chapter first distills the principal conclusions drawn from our systematic evaluation of time-series data augmentation, and then sketches promising avenues for future inquiry. Together, these discussions furnish researchers with a concise roadmap that bridges current insights and forthcoming challenges in the field.

8.1 Conclusions

This investigation embarked on a meticulous examination of data augmentation strategies, specifically curated for time series classification. Upon an exhaustive retrospection spanning the past 10 years, we discerned a plethora of techniques, accumulating over 60 distinct methods from more than 100 scholarly articles. These methods were subsequently assimilated into an innovative taxonomy, crafted with a focus on time-series classification. This taxonomy segregates the techniques into five quintessential categories: Transformation-Based, Pattern-Based, Generative, Decomposition-Based, and Automated Data Augmentation, thereby providing a structured reference point for scholars and industry practitioners.

Our inquiry transcended theoretical postulation and went on to practical assessments. We applied a variety of nearly 20 prevalent data augmentation strategies to all 15 types categories of datasets derived from the esteemed UCR time series classification archive. The decision to employ ResNet and LSTM as the evaluative model was strategic, motivated by their documented success in analogous research contexts, thus ensuring solid experimental groundwork. Our multifaceted evaluation approach, encompassing Accuracy, Method Ranking, and Residual Analysis, yielded insights into the nuanced performance dynamics of each technique.

The experimental phase spanned approximately two months, culminating in a baseline model accuracy of $84.98 \pm 16.41\%$ in ResNet and $82.41 \pm 18.71\%$ in LSTM. In particular, the RGWs, which ranked top 2 in both of the models, substantially improved the performance of the model, while the EMD method was detrimental. Other strategies, including Window Slicing and GAN, manifested negligible deviations, underscoring the situational specificity of these methods. Transformation-Based Methods, in particular, have consistently exhibited their merit, potentially attributable to their finesse in maintaining and subtly modifying the inherent structures of time-series data.

Our analysis, extended across various UCR datasets, revealed that the intrinsic characteristics of each dataset significantly influence the success of augmentation strategies. Based on our research and experimental findings, we provide detailed, actionable recommendations and method suggestions for researchers. These insights are designed to equip practitioners with well-informed and tailored augmentation strategies, ensuring they maximize model performance while avoiding the risks associated with indiscriminate application of augmentation techniques.

To contribute tangibly to the field, we have integrated most of these strategies into a comprehensive Python Library. This utility, engineered for user-friendliness, simplifies the application of diverse augmentation strategies, thereby catalyzing further experimental endeavors and innovations.

In essence, this research presents an integrative perspective on the contemporary landscape of data augmentation for time series classification, combining theoretical frameworks with empirical evidence. The revelations and resources introduced herein are positioned to catalyze continued progress in this domain, fortifying machine learning models against the challenges posed by data limitations, and enhancing their generalizability and robustness.

8.2 Future Research Directions

The domain of data augmentation for time-series classification is expansive and continually evolving. The milestones achieved in this research constitute a foundational layer with expansive potential for further exploration and refinement. The following are prospective directions for subsequent inquiries.

8.2.1 Exploration of Diverse Neural Network Architectures.

While the present study was anchored around a select group of established neural network models, the dynamic nature of deep learning architectures intimates the potential advantages of a more extensive exploration. In particular, the adaptability of emerging models, especially those designed for sequential data like Transformer models, warrants comprehensive investigation.

8.2.2 Extension to an Expanded Dataset Spectrum.

The scope of our study was limited to a specific selection of datasets, representing only a segment of the myriad applications of time series classification. Future research could broaden this scope to include datasets from diverse fields such as finance, environmental science, or social dynamics, thus testing the universality and domain-specific nuances of the augmentation techniques. Additionally, a promising direction is the creation of synthetic datasets with controlled inner characteristics. For instance, generating time series data from ARMA processes could offer a structured approach to examining how different data augmentation methods influence time-series classification across varied data types. By manipulating these synthetic datasets, researchers can gain deeper insights into the specific strengths and weaknesses of each augmentation method, ultimately enhancing their application to real-world datasets.

8.2.3 Advanced Augmentation Selection and Combination.

Within the domain of data augmentation, the synergistic integration of disparate techniques emerges as a pivotal factor in augmenting the efficacy of model training. However, this integration necessitates a methodical selection process to mitigate potential conflicts and eliminate redundancy. Our empirical analysis elucidates that specific technique combinations may either synergize or antagonize, thereby exerting a differential impact on model efficacy. Illustratively, the amalgamation of noise injection with geometric transformations might veil essential data patterns, whereas the concurrent application of jittering and scaling could effectively replicate real-world data variabilities. This underscores the imperative for a data-driven approach to data augmentation, predicated on a nuanced comprehension of dataset characteristics. Furthermore, the conduct of preliminary experiments or small-scale pilot studies is posited as a crucial step towards garnering insights into the compatibility and efficacy of various augmentation technique combinations. Such investigative endeavors facilitate the elucidation of optimal strategies that not only maximize model performance but also safeguard the integrity of the fundamental information content inherent within the data.

Innovative methods such as RandECG [101], which judiciously integrate multiple enhancements through transparent and interpretable hyperparameters, herald a promising future for augmentation synthesis. Adapting such strategies for time-series data, potentially with the incorporation of advanced frameworks like reinforcement learning or meta-learning, could optimize augmentation processes. Furthermore, ensuring the computational efficiency of these algorithms, especially in scenarios with extensive datasets or limited processing capabilities, remains an essential consideration.

Acknowledgments

References

[1] Connor Shorten and Taghi M Khoshgoftaar. 2019. A survey on image data augmentation for deep learning. *Journal of big data* 6, 1 (2019), 1–48.

[2] A Jung et al. 2017. imgaug: Image augmentation for machine learning experiments. *p. Accessed* 3 (2017), 977–997.

[3] G Forestier, J Weber, L Idoumghar, PA Muller, et al. 2019. Deep learning for time series classification: a review. *Data Min. Knowl. Discov.* 33, 4 (2019), 917–963.

[4] Brian Kenji Iwana and Seiichi Uchida. 2021. An empirical survey of data augmentation for time series classification with neural networks. *Plos one* 16, 7 (2021), e0254841.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25 (2012).

[6] Terry T Um, Franz MJ Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić. 2017. Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. In *Proceedings of the 19th ACM international conference on multimodal interaction*. 216–220.

[7] Khandakar M Rashid and Joseph Louis. 2019. Window-warping: a time series data augmentation of IMU data for construction equipment activity identification. In *ISARC. Proceedings of the international symposium on automation and robotics in construction*, Vol. 36. IAARC Publications, 651–657.

[8] Edgar Talavera, Guillermo Iglesias, Ángel González-Prieto, Alberto Mozo, and Sandra Gómez-Canaval. 2022. Data augmentation techniques in time series domain: A survey and taxonomy. *arXiv preprint arXiv:2206.13508* 2206.13508 (2022).

[9] Anthony Bagnall, Jason Lines, Aaron Bostrom, James Large, and Eamonn Keogh. 2017. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data mining and knowledge discovery* 31 (2017), 606–660.

[10] John Cristian Borges Gamboa. 2017. Deep learning for time-series analysis. *arXiv preprint arXiv:1701.01887* 1701.01887 (2017).

[11] Hoang Anh Dau, Anthony Bagnall, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanama-hatana, and Eamonn Keogh. 2019. The UCR time series archive. *IEEE/CAA Journal of Automatica Sinica* 6, 6 (2019), 1293–1305.

[12] Peiyu Li, Soukaïna Filali Boubrahimi, and Shah Muhammad Hamdi. 2021. Shapelets-based Data Augmentation for Time Series Classification. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 1373–1378.

[13] Arthur Le Guennec, Simon Malinowski, and Romain Tavenard. 2016. Data augmentation for time series classification using convolutional neural networks. In *ECML/PKDD workshop on advanced analytics and learning on temporal data*.

[14] Germain Forestier, François Petitjean, Hoang Anh Dau, Geoffrey I Webb, and Eamonn Keogh. 2017. Generating synthetic time series to augment sparse datasets. In *2017 IEEE international conference on data mining (ICDM)*. IEEE, 865–870.

[15] Shota Haradal, Hideaki Hayashi, and Seiichi Uchida. 2018. Biosignal data augmentation based on generative adversarial networks. In *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 368–371.

[16] Elizabeth Fons, Paula Dawson, Xiao-jun Zeng, John Keane, and Alexandros Iosifidis. 2020. Evaluating data augmentation for financial time series classification. *arXiv preprint arXiv:2010.15111* 2010.15111 (2020).

[17] Naoki Nonaka and Jun Seita. 2020. Data augmentation for electrocardiogram classification with deep neural network. *arXiv preprint arXiv:2009.04398* 2009.04398 (2020).

[18] Maxime Goubeaud, Philipp Joußen, Nicolla Gmyrek, Farzin Ghorban, and Anton Kummert. 2021. White Noise Windows: Data Augmentation for Time Series. In *2021 7th International Conference on Optimization and Applications (ICOA)*. IEEE, 1–5.

[19] Qingsong Wen, Liang Sun, Fan Yang, Xiaomin Song, Jingkun Gao, Xue Wang, and Huan Xu. 2020. Time series data augmentation for deep learning: A survey. *arXiv preprint arXiv:2002.12478* 2002.12478 (2020).

[20] Humza Naveed, Saeed Anwar, Munawar Hayat, Kashif Javed, and Ajmal Mian. 2021. Survey: Image mixing and deleting for data augmentation. *arXiv preprint arXiv:2106.07085* 2106.07085 (2021).

[21] Steven Y Feng, Varun Gangal, Jason Wei, Sarath Chandar, Soroush Vosoughi, Teruko Mitamura, and Eduard Hovy. 2021. A survey of data augmentation approaches for NLP. *arXiv preprint arXiv:2105.03075* 2105.03075 (2021).

[22] Claes Wohlin. 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering*. 1–10.

[23] Wayne C Booth, Gregory G Colomb, and Joseph M Williams. 2003. *The craft of research*. University of Chicago press.

[24] John W Creswell and J David Creswell. 2017. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

[25] Barbara Gastel and Robert A Day. 2022. *How to write and publish a scientific paper*. ABC-CLIO.

[26] Shuang Pi, Shuanggen Zhang, Shumin Wang, Bochi Guo, and Wei Yan. 2022. Improving Modulation Recognition Using Time Series Data Augmentation via a Spatiotemporal Multi-Channel Framework. *Electronics* 12, 1 (2022), 96.

[27] DongShi Zhang, Jun Liu, and ChangHao Liang. 2017. Perspective on how laser-ablated particles grow in liquids. *Science China Physics, Mechanics & Astronomy* 60 (2017), 1–16.

[28] Zhuoran Cai, Wenxuan Ma, Xinrui Wang, Hanhong Wang, and Zhongming Feng. 2023. The Performance Analysis of Time Series Data Augmentation Technology for Small Sample Communication Device Recognition. *IEEE Transactions on Reliability* 72, 2 (2023), 574–585. https://doi.org/10.1109/TR.2022.3178707

[29] Gyu-Il Kim and Kyungyong Chung. 2024. Augmented and End-to-End Models for Defect Classification of Structures. In *Proceedings of the 18th ACM International Conference on Distributed and Event-based Systems*. 183–184.

[30] Bo Liu, Zhenguo Zhang, and Rongyi Cui. 2020. Efficient time series augmentation methods. In *2020 13th international congress on image and signal processing, BioMedical engineering and informatics (CISP-BMEI)*. IEEE, 1004–1009.

[31] Kyutae Kim and Jongpil Jeong. 2020. Deep learning-based data augmentation for hydraulic condition monitoring system. *Procedia Computer Science* 175 (2020), 20–27.

[32] Khandakar M Rashid and Joseph Louis. 2019. Times-series data augmentation and deep learning for construction equipment activity recognition. *Advanced Engineering Informatics* 42 (2019), 100944.

[33] Liang Huang, Weijian Pan, You Zhang, Liping Qian, Nan Gao, and Yuan Wu. 2019. Data augmentation for deep learning-based radio modulation classification. *IEEE access* 8 (2019), 1498–1506.

[34] Chao Li, Korkut Kaan Tokgoz, Masamoto Fukawa, Jim Bartels, Takumi Ohashi, Ken-ichi Takeda, and Hiroyuki Ito. 2021. Data augmentation for inertial sensor data in CNNs for cattle behavior classification. *IEEE Sensors Letters* 5, 11 (2021), 1–4.

[35] Pouya Hosseinzadeh, Soukaina Filali Boubrahimi, and Shah Muhammad Hamdi. 2024. Improving solar energetic particle event prediction through multivariate time series data augmentation. *The Astrophysical Journal Supplement Series* 270, 2 (2024), 31.

[36] Edmund Do, Jack Boynton, Byung Suk Lee, and Daniel Lustgarten. 2022. Data augmentation for 12-lead ecg beat classification. *SN Computer Science* 3 (2022), 1–17.

[37] Dawid Warchoł and Mariusz Oszust. 2022. Efficient Augmentation of Human Action Recognition Datasets with Warped Windows. *Procedia Computer Science* 207 (2022), 3018–3027.

[38] Navdeep Jaitly and Geoffrey E Hinton. 2013. Vocal tract length perturbation (VTLP) improves speech recognition. In *Proc. ICML Workshop on Deep Learning for Audio, Speech and Language*, Vol. 117. 21.

[39] Wenbo Yang, Jidong Yuan, and Xiaokang Wang. 2023. SFCC: data augmentation with stratified Fourier coefficients combination for time series classification. *Neural Processing Letters* 55, 2 (2023), 1833–1846.

[40] Maxime Goubeaud, Nicolla Gmyrek, Farzin Ghorban, Lucas Schelkes, and Anton Kummert. 2021. Random Noise Boxes: Data Augmentation for Spectrograms. In *2021 IEEE International Conference on Progress in Informatics and Computing (PIC)*. IEEE, 24–28.

[41] Odongo Steven Eyobu and Dong Seog Han. 2018. Feature representation and data augmentation for human activity classification based on wearable IMU sensor data using a deep LSTM neural network. *Sensors* 18, 9 (2018), 2892.

[42] Tracey Eileen KM Lee, YL Kuah, Kee-Hao Leo, Saeid Sanei, Effie Chew, and Ling Zhao. 2019. Surrogate rehabilitative time series data for image-based deep learning. In *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 1–5.

[43] Rodrigo H Zanella, Lucas A de Castro Coelho, and Vinicius MA Souza. 2022. TS-DENSE: Time Series Data Augmentation by Subclass Clustering. In *2022 26th International Conference on Pattern Recognition (ICPR)*. IEEE, 1800–1806.

[44] Amine Mohamed Aboussalah, Minjae Kwon, Raj G Patel, Cheng Chi, and Chi-Guhn Lee. 2022. Recursive Time Series Data Augmentation. In *The Eleventh International Conference on Learning Representations*.

[45] Krzysztof Kamycki, Tomasz Kapuscinski, and Mariusz Oszust. 2019. Data augmentation with suboptimal warping for time-series classification. *Sensors* 20, 1 (2019), 98.

[46] Dawid Warchoł and Mariusz Oszust. 2022. Augmentation of Human Action Datasets with Suboptimal Warping and Representative Data Samples. *Sensors* 22, 8 (2022), 2947.

[47] Hans Friedrich Stabenau, Christopher P Bridge, and Jonathan W Waks. 2021. ECGAug: A novel method of generating augmented annotated electrocardiogram QRST complexes and rhythm strips. *Computers in biology and medicine* 134 (2021), 104408.

[48] Brian Kenji Iwana and Seiichi Uchida. 2021. Time series data augmentation for neural networks by time warping with a discriminative teacher. In *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 3558–3565.

[49] Cheolhwan Oh, Seungmin Han, and Jongpil Jeong. 2020. Time-series data augmentation based on interpolation. *Procedia Computer Science* 175 (2020), 64–71.

[50] Xinyu Yang, Zhenguo Zhang, Xu Cui, and Rongyi Cui. 2021. A time series data augmentation method based on dynamic time warping. In *2021 International Conference on Computer Communication and Artificial Intelligence (CCAI)*. IEEE, 116–120.

[51] Zhifeng Li, Yaqin Song, Runchen Li, Sen Gu, and Xuze Fan. 2022. A Novel Data Augmentation Method for Improving the Accuracy of Insulator Health Diagnosis. *Sensors* 22, 21 (2022), 8187.

[52] Pin Liu, Xiaohui Guo, Pengpeng Chen, Bin Shi, Tianyu Wo, and Xudong Liu. 2022. Adaptive Shapelets Preservation for Time Series Augmentation. In *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

[53] Ebrahim Khalili and Babak Mohammadzadeh Asl. 2021. Automatic sleep stage classification using temporal convolutional neural network and new data augmentation technique from raw single-channel EEG. *Computer Methods and Programs in Biomedicine* 204 (2021), 106063.

[54] Yousef El-Laham and Svitlana Vyetrenko. 2022. StyleTime: Style Transfer for Synthetic Time Series Generation. In *Proceedings of the Third ACM International Conference on AI in Finance*. 489–496.

[55] Mohammad Akyash, Hoda Mohammadzade, and Hamid Behroozi. 2021. DTW-Merge: A Novel Data Augmentation Technique for Time Series Classification. arXiv:2103.01119 [cs.LG]

[56] Nanni Loris and Maguolo Gianluca. 2020. Paci Michelangelo. *Data augmentation approaches for improving animal audio classification. Ecological Informatics* 57, 101084 (2020), 10–1016.

[57] Xiaodong Cui, Vaibhava Goel, and Brian Kingsbury. 2015. Data augmentation for deep neural network acoustic modeling. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 23, 9 (2015), 1469–1477.

[58] Xuyang Zhao, Jordi Solé-Casals, Binghua Li, Zihao Huang, Andong Wang, Jianting Cao, Toshihisa Tanaka, and Qibin Zhao. 2020. Classification of epileptic IEEG signals by CNN and data augmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 926–930.

[59] Loris Nanni, Gianluca Maguolo, and Michelangelo Paci. 2020. Data augmentation approaches for improving animal audio classification. *Ecological Informatics* 57 (2020), 101084.

[60] Daniel S Park, William Chan, Yu Zhang, Chung-Cheng Chiu, Barret Zoph, Ekin D Cubuk, and Quoc V Le. 2019. Specaugment: A simple data augmentation method for automatic speech recognition. *arXiv preprint arXiv:1904.08779* 1904.08779 (2019).

[61] Hong Cao, Vincent YF Tan, and John ZF Pang. 2014. A parsimonious mixture of Gaussian trees model for oversampling in imbalanced and multimodal time-series classification. *IEEE transactions on neural networks and learning systems* 25, 12 (2014), 2226–2239.

[62] Yanfei Kang, Rob J Hyndman, and Feng Li. 2020. GRATIS: GeneRAting TIme Series with diverse and controllable characteristics. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 13, 4 (2020), 354–376.

[63] Sagheer Khan, Aaesha Alzaabi, Tharmalingam Ratnarajah, and Tughrul Arslan. 2024. Novel statistical time series data augmentation and machine learning based classification of unobtrusive respiration data for respiration Digital Twin model. *Computers in Biology and Medicine* 168 (2024), 107825.

[64] Xue Jiang, Jianhui Zhao, Bo Du, and Zhiyong Yuan. 2021. Self-supervised Contrastive Learning for EEG-based Sleep Staging. arXiv:2109.07839 [cs.LG]

[65] Hubert Banville, Omar Chehab, Aapo Hyvärinen, Denis-Alexander Engemann, and Alexandre Gramfort. 2021. Uncovering the structure of clinical EEG signals with self-supervised learning. *Journal of Neural Engineering* 18, 4 (2021), 046020.

[66] Anindya Sarkar, Anirudh Sunder Raj, and Raghu Sesha Iyengar. 2020. Neural Data Augmentation Techniques for Time Series Data and its Benefits. In *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Vol. 107–114. IEEE.

[67] Faezeh Nejati Hatamian, Nishant Ravikumar, Sulaiman Vesal, Felix P Kemeth, Matthias Struck, and Andreas Maier. 2020. The effect of data augmentation on classification of atrial fibrillation in short single-lead ECG signals using deep neural networks. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1264–1268.

[68] Germán García-Jara, Pavlos Protopapas, and Pablo A Estévez. 2022. Improving astronomical time-series classification via data augmentation with generative adversarial networks. *The Astrophysical Journal* 935, 1 (2022), 23.

[69] Justin Albert, Pawel Glöckner, Bjarne Pfitzner, and Bert Arnrich. 2021. Data augmentation of kinematic time-series from rehabilitation exercises using GANs. In *2021 IEEE International Conference on Omni-Layer Intelligent Systems (COINS)*. IEEE, 1–6.

[70] Nicolas Morizet, Matteo Rizzato, David Grimbert, and George Luta. 2022. A Pilot Study on the Use of Generative Adversarial Networks for Data Augmentation of Time Series. *AI* 3, 4 (2022), 789–795.

[71] Huan Lou, Zongfeng Qi, and Jianxun Li. 2018. One-dimensional data augmentation using a Wasserstein generative adversarial network with supervised signal. In *2018 Chinese Control And Decision Conference (CCDC)*. IEEE, 1896–1901.

[72] Giorgia Ramponi, Pavlos Protopapas, Marco Brambilla, and Ryan Janssen. 2018. T-cgan: Conditional generative adversarial network for data augmentation in noisy time series with irregular sampling. *arXiv preprint arXiv:1811.08295* 1811.08295 (2018).

[73] Maximilian Ehrhart, Bernd Resch, Clemens Havas, and David Niederseer. 2022. A Conditional GAN for Generating Time Series Data for Stress Detection in Wearable Physiological Sensor Data. *Sensors* 22, 16 (2022), 5969.

[74] Genlang Chen, Yi Zhu, Zhiqing Hong, and Zhen Yang. 2019. EmotionalGAN: Generating ECG to enhance emotion state classification. In *Proceedings of the 2019 International conference on artificial intelligence and computer science*. 309–313.

[75] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Rezende, and Daan Wierstra. 2015. Draw: A recurrent neural network for image generation. In *International conference on machine learning*. PMLR, 1462–1471.

[76] Zhenyu Yang, Yantao Li, and Gang Zhou. 2023. TS-GAN: Time-series GAN for Sensor-based Health Data Augmentation. *ACM Transactions on Computing for Healthcare* 4, 2 (2023), 1–21.

[77] Xiaomin Li, Anne Hee Hiong Ngu, and Vangelis Metsis. 2022. Tts-cgan: A transformer time-series conditional gan for biosignal data augmentation. *arXiv preprint arXiv:2206.13676* 2206.13676 (2022).

[78] Kaleb E Smith and Anthony O Smith. 2021. A spectral enabled GAN for time series data generation. *arXiv preprint arXiv:2103.01904* 2103.01904 (2021).

[79] Zheng Li, Chao Ma, Xiaochuan Shi, Dian Zhang, Wei Li, and Libing Wu. 2021. Tsa-gan: A robust generative adversarial networks for time series augmentation. In *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

[80] Kaleb E Smith and Anthony O Smith. 2020. Conditional GAN for timeseries generation. *arXiv preprint arXiv:2006.16477* 2006.16477 (2020).

[81] Dongsheng Liu, Yuting Wu, Deyan Hong, and Siting Wang. 2022. Time series data augmentation method of small sample based on optimized generative adversarial network. *Concurrency and Computation: Practice and Experience* 34, 27 (2022), e7331.

[82] Grace Deng, Cuize Han, Tommaso Dreossi, Clarence Lee, and David S Matteson. 2022. Ib-gan: A unified approach for multivariate time series classification under class imbalance. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*. SIAM, 217–225.

[83] Dani Kiyasseh, Girmaw Abebe Tadesse, Louise Thwaites, Tingting Zhu, David Clifton, et al. 2020. PlethAugment: GAN-based PPG augmentation for medical diagnosis in low-resource settings. *IEEE journal of biomedical and health informatics* 24, 11 (2020), 3226–3235.

[84] Mengyue Zha, SiuTim Wong, Mengqi Liu, Tong Zhang, and Kani Chen. 2022. Time series generation with masked autoencoder. *arXiv preprint arXiv:2201.07006* 2201.07006 (2022).

[85] Terrance DeVries and Graham W Taylor. 2017. Dataset Augmentation in Feature Space. *arXiv e-prints* 1702.05538, 1702.05538 (2017), arXiv–1702.

[86] Sung-Soo Park, Hye-Jeong Kwon, Ji-Won Baek, and Kyungyong Chung. 2022. Dimensional Expansion and Time-Series Data Augmentation Policy for Skeleton-Based Pose Estimation. *IEEE Access* 10 (2022), 112261–112272.

[87] Luay Alawneh, Tamam Alsarhan, Mohammad Al-Zinati, Mahmoud Al-Ayyoub, Yaser Jararweh, and Hongtao Lu. 2021. Enhancing human activity recognition using deep learning and time series augmented data. *Journal of Ambient Intelligence and Humanized Computing* 12, 12 (2021), 10565–10580.

[88] Qiankun Feng, Yue Li, and Hongzhou Wang. 2021. Intelligent random noise modeling by the improved variational autoencoding method and its application to data augmentation. *Geophysics* 86, 1 (2021), T19–T31.

[89] Francisco J Moreno-Barea, José M Jerez, and Leonardo Franco. 2020. Improving classification accuracy using data augmentation on small data sets. *Expert Systems with Applications* 161 (2020), 113696.

[90] Guillermo Iglesias, Edgar Talavera, Ángel González-Prieto, Alberto Mozo, and Sandra Gómez-Canaval. 2023. Data Augmentation techniques in time series domain: a survey and taxonomy. *Neural Computing and Applications* 35, 14 (2023), 10123–10145.

[91] Wei-Ning Hsu, Yu Zhang, and James Glass. 2017. Unsupervised domain adaptation for robust speech recognition via variational autoencoder-based data augmentation. In *2017 IEEE automatic speech recognition and understanding workshop (ASRU)*. IEEE, 16–23.

[92] Ludovico Minati, Chao Li, Jim Bartels, Parthojit Chakraborty, Zixuan Li, Natsue Yoshimura, Mattia Frasca, and Hiroyuki Ito. 2023. Accelerometer time series augmentation through externally driving a non-linear dynamical system. *Chaos, Solitons & Fractals* 168 (2023), 113100.

[93] Yonatan Sanz Perl, Carla Pallavicini, Ignacio Perez Ipiña, Morten Kringelbach, Gustavo Deco, Helmut Laufs, and Enzo Tagliazucchi. 2020. Data augmentation based on dynamical systems for the classification of brain states. *Chaos, Solitons & Fractals* 139 (2020), 110069.

[94] Jordan Yeomans, Simon Thwaites, William SP Robertson, David Booth, Brian Ng, and Dominic Thewlis. 2019. Simulating time-series data for improved deep neural network performance. *IEEE Access* 7 (2019), 131248–131255.

[95] Zihao Wang, Youli Qu, Junru Tao, and Yudan Song. 2019. Image-mediated data augmentation for low-resource human activity recognition. In *Proceedings of the 2019 3rd International Conference on Compute and Data Analysis*. 49–54.

[96] Chenguang Li, Hongjun Yang, Long Cheng, and Fubiao Huang. 2022. A Time-Series Augmentation Method Based on Empirical Mode Decomposition and Integrated LSTM Neural Network. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 333–336.

[97] Gue-Hwan Nam, Seok-Jun Bu, Na-Mu Park, Jae-Yong Seo, Hyeon-Cheol Jo, and Won-Tae Jeong. 2020. Data augmentation using empirical mode decomposition on neural networks to classify impact noise in vehicle. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 731–735.

[98] Qingsong Wen, Jingkun Gao, Xiaomin Song, Liang Sun, Huan Xu, and Shenghuo Zhu. 2019. RobustSTL: A robust seasonal-trend decomposition algorithm for long time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 5409–5416.

[99] Christoph Bergmeir, Rob J Hyndman, and José M Benítez. 2016. Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. *International journal of forecasting* 32, 2 (2016), 303–312.

[100] Qingsong Wen, Zhe Zhang, Yan Li, and Liang Sun. 2020. Fast RobustSTL: Efficient and robust seasonal-trend decomposition for time series with complex patterns. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2203–2213.

[101] Naoki Nonaka and Jun Seita. 2021. RandECG: Data Augmentation for Deep Neural Network based ECG classification. In *Annual Conference of the Japanese Society for Artificial Intelligence*. Springer, 178–189.

[102] Chao Li, Ludovico Minati, Korkut Kaan Tokgoz, Masamoto Fukawa, Jim Bartels, A Sihan, Ken-Ichi Takeda, and Hiroyuki Ito. 2022. Integrated data augmentation for accelerometer time series in behavior recognition: roles of sampling, balancing, and fourier surrogates. *IEEE Sensors Journal* 22, 24 (2022), 24230–24241.

[103] Haiping Wu and Hui Liu. 2021. Non-intrusive load transient identification based on multivariate LSTM neural network and time series data augmentation. *Sustainable Energy, Grids and Networks* 27 (2021), 100490.

[104] Qianli Ma, Zhenjing Zheng, Jiawei Zheng, Sen Li, Wanqing Zhuang, and Garrison W Cottrell. 2021. Joint-label learning by dual augmentation for time series classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 8847–8855.

[105] Elizabeth Fons, Paula Dawson, Xiao-jun Zeng, John Keane, and Alexandros Iosifidis. 2021. Adaptive weighting scheme for automatic time-series data augmentation. *arXiv preprint arXiv:2102.08310* 2102.08310 (2021).

[106] Tsz-Him Cheung and Dit-Yan Yeung. 2020. Modals: Modality-agnostic automated data augmentation in the latent space. In *International Conference on Learning Representations*.

[107] Naoki Saito. 1994. *Local feature extraction and its applications using a library of bases*. Yale University.

[108] Yanping Chen, Yuan Hao, Thanawin Rakthanmanon, Jesin Zakaria, Bing Hu, and Eamonn Keogh. 2015. A general framework for never-ending learning from time series streams. *Data mining and knowledge discovery* 29 (2015), 1622–1664.

[109] Anthony Bagnall, Luke Davis, Jon Hills, and Jason Lines. 2012. Transformation based ensembles for time series classification. In *Proceedings of the 2012 SIAM international conference on data mining*. SIAM, 307–318.

[110] James K Holland, E Katherine Kemsley, and Reginald H Wilson. 1998. Use of Fourier transform infrared spectroscopy and partial least squares regression for the detection of adulteration of strawberry purees. *Journal of the Science of Food and Agriculture* 76, 2 (1998),

263–269.

[111] Fuming Fang and Takahiro Shinozaki. 2018. Electrooculography-based continuous eye-writing recognition system for efficient assistive communication systems. *PloS one* 13, 2 (2018), e0192684.

[112] Sha Lu, Gordana Mirchevska, Sayali S Phatak, Dongmei Li, Janos Luka, Richard A Calderone, and William A Fonzi. 2017. Dynamic time warping assessment of high-resolution melt curves provides a robust metric for fungal identification. *PLoS One* 12, 3 (2017), e0173320.

[113] Fabio M Caputo, Pietro Prebianca, Alessandro Carcangiu, Lucio D Spano, and Andrea Giachetti. 2018. Comparing 3D trajectories for simple mid-air gesture recognition. *Computers & Graphics* 73 (2018), 17–25.

[114] Denis S Willett, Justin George, Nora S Willett, Lukasz L Stelinski, and Stephen L Lapointe. 2016. Machine learning for characterization of insect vector feeding. *PLoS computational biology* 12, 11 (2016), e1005158.

[115] Christos Sapsanis, George Georgoulas, Anthony Tzes, and Dimitrios Lymberopoulos. 2013. Improving EMG based classification of basic hand movements using EMD. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 5754–5757.

[116] Yang Zhou, Haoxi Zhang, Yuan Li, and Guangjian Ning. 2020. ECG heartbeat classification based on ResNet and Bi-LSTM. In *IOP Conference Series: Earth and Environmental Science*, Vol. 428. IOP Publishing, 012014.

[117] Xinjie Xu, Zhuangzhi Chen, Dongwei Xu, Huaji Zhou, Shanqing Yu, Shilian Zheng, Qi Xuan, and Xiaoniu Yang. 2022. Mixing signals: Data augmentation approach for deep learning based modulation recognition. *arXiv preprint arXiv:2204.03737* (2022).

[118] Huiyuan Yang, Han Yu, and Akane Sano. 2022. Empirical evaluation of data augmentations for biobehavioral time series data with deep learning. *arXiv preprint arXiv:2210.06701* (2022).

[119] Fazle Karim, Somshubra Majumdar, and Houshang Darabi. 2019. Insights into LSTM fully convolutional networks for time series classification. *Ieee Access* 7 (2019), 67718–67725.

[120] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. 2017. LSTM fully convolutional networks for time series classification. *IEEE access* 6 (2017), 1662–1669.

[121] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Samuel Harford. 2019. Multivariate LSTM-FCNs for time series classification. *Neural networks* 116 (2019), 237–245.

[122] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*. pmlr, 448–456.

[123] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[124] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 249–256.

[125] Diederik P. Kingma and Jimmy Ba. 2017. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs.LG]