ARTICLE

Check for updates

Journal of
Sensor Science and Technology

The Korean Sensors Society

SAPF: A Similarity-Based Pre-Evaluation Framework for Determining Effective Augmentation Ratios in Time-Series Data

Gyu-Li Kim[1] and Kwangjae Lee[2,+]

[1]*Department of Battery Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Republic of Korea*
[2]*Department of AI Mobility Engineering, Sangmyung University, 31 Sangmyeongdae-gil, Dongnam-gu, Cheonan, 31066, Republic of Korea*

Cite This: *J. Sens. Sci. Technol. Vol. 34, No. 6 (2025) 630-637*

https://doi.org/10.46670/JSST.2025.34.6.630

ABSTRACT: In deep learning, data augmentation is key for enhancing model performance and improving generalization. However, its effectiveness depends not only on the applied augmentation method but also on the augmentation ratio, which determines the scale of the generated data. Conventional approaches either arbitrarily set this ratio or use repetitive model training under all conditions, resulting in high computational costs and limited practicality. To overcome these limitations, we propose a Similarity-based Augmentation Performance Framework (SAPF) that predicts the optimal augmentation ratio without iterative training. SAPF extracts embedding vectors from a Bidirectional Gated Recurrent Unit(Bi-GRU) model trained on the original dataset and quantifies the distributional difference between the original and augmented data using the Wasserstein Distance (WD). By analyzing the WD growth pattern across augmentation ratios (×2 to ×100), SAPF identifies the optimal ratio as the saturation point where further augmentation yields negligible distributional change. Experimental results show that WD and classification accuracy increased rapidly at lower ratios (×2 to ×25) and then saturated, with a strong positive correlation (ρ = 0.94–1.00) between WD and accuracy. Furthermore, SAPF maintained comparable performance while reducing the training time by more than 99.28% compared to conventional methods, thereby demonstrating its efficiency and practicality in designing effective augmentation strategies.

KEYWORDS: *Data augmentation, Gas classification, Wasserstein distance, Similarity, Deep learning*

1. INTRODUCTION

The performance of deep-learning models largely depends on the availability of large-scale, high-quality training data. However, collecting and accurately labeling data in real-world environments is costly and time-consuming, and is often constrained by factors such as data accessibility and ethical limitations [1-4]. Data scarcity is particularly common in fields such as medical imaging, rare disease diagnosis, and IoT sensor–based monitoring, where it hinders model generalization and increases the risk of overfitting [5-7]. Data augmentation has been widely used to overcome these challenges. This technique transforms existing data in various ways to increase the quantity and diversity of training samples, thereby enhancing the model expressiveness and generalization. It has been applied across diverse domains, including images, text, and time-series data [8-11].

Conventional research on data augmentation has mainly focused on the design and application of various augmentation techniques and verification of their effectiveness in improving model generalization [12,13]. However, both the type of augmentation method and the augmentation ratio, which determine the amount of data generated, have a strong impact on the performance. However, many studies have arbitrarily selected ratios without establishing clear standards. Excessive augmentation may produce redundant data, whereas insufficient augmentation may limit learning diversity, both of which can negatively affect the efficiency and model performance [14,15]. Recognizing these issues, several studies have attempted to analyze the effects of augmentation ratios on performance systematically. For example, Li *et al.* [16] compared the model performance using different augmentation ratios (1:10, 1:100, and 1:1000) for automatic medical report labeling, whereas Corrado and Hanna [17] investigated the relationship between

[+]Corresponding author: begleam@smu.ac.kr
Received : Oct. 23, 2025, Accepted : Oct. 28, 2025

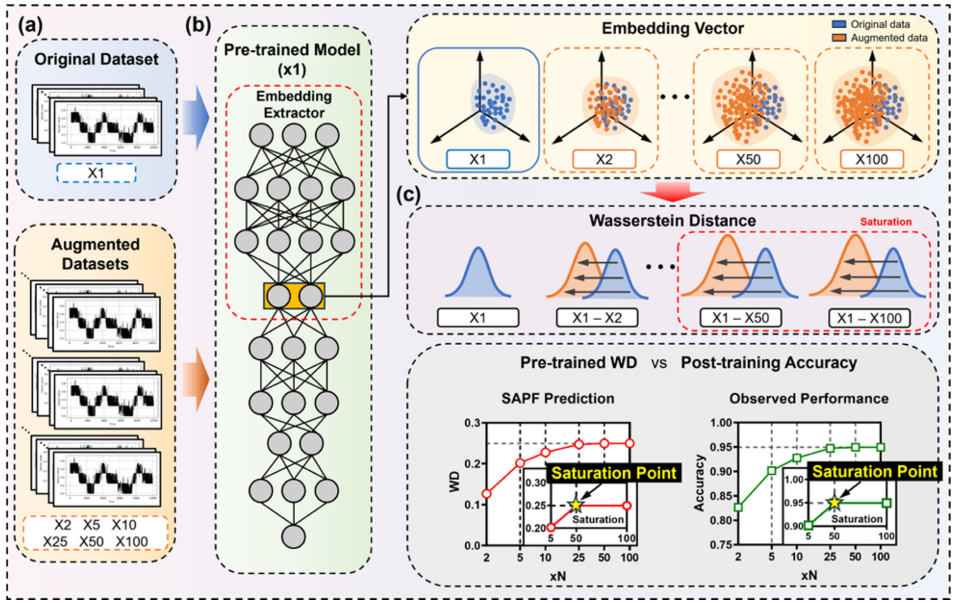

Fig. 1. Schematic overview of the SAPF. (a) Data preparation using the original (×1) and augmented (×2 to ×100) datasets generated by SpecAugment and SpecSwap. (b) Embedding extraction by training a Bi-GRU once on the ×1 dataset and obtaining frozen embeddings. (c) Similarity analysis using the WD to identify the optimal augmentation ratio.

data efficiency and performance by adjusting the augmentation ratio in reinforcement learning. However, this approach requires repeated model training for each condition, and excessive time and computational resources [18].

This problem is more pronounced for time-series data. Because of their inherent characteristics such as temporal order, seasonality, and trends, it is difficult to directly apply augmentation techniques that are effective for image and text data [19-21]. Furthermore, time-series models are often structurally complex and computationally intensive, making it impractical to perform repeated training across multiple augmentation ratios [22]. Nevertheless, studies providing prior evaluation criteria for efficiently determining augmentation ratios in time-series data remain limited.

To address this gap, this study proposes a Similarity-based Pre-Evaluation Framework (SAPF) that enables the pre-evaluation of augmentation effects without repeated model training. The SAPF utilizes the embedding vectors extracted from a Bidirectional Gated Recurrent Unit(Bi-GRU) model trained on the original dataset. The distributional difference between the original and augmented data was measured using the Wasserstein Distance (WD), thereby indirectly estimating the effect of augmentation ratio on model performance [23]. This approach is valuable because it allows the early identification of performance saturation points before training, which helps prevent unnecessary augmentation and repeated learning, and ultimately reduces the overall training time.

2. EXPERIMENTAL

2.1 Proposed Framework

The proposed SAPF was designed to evaluate the effectiveness of data augmentation quantitatively without iterative model training. The overall process is illustrated in Fig. 1 and consists of the following four steps.

Step 1: Data preparation. Time-series sensor data were converted into mel-spectrograms to represent their characteristics in the frequency–time domain. An original dataset (×1) and several augmented datasets (×2 to ×100) were constructed. Although the augmentation method is not limited to a specific technique, this study used SpecAugment and SpecSwap for a consistent performance evaluation [24,25].

Step 2: Embedding vector extraction. At this stage, a neural network model such as Bi-GRU was trained using the original dataset (×1) to establish a stable representational space. During this process, the model learns parameters that reflect the statistical structure of the original data. One of the intermediate layers was defined as the embedding layer, which served as a shared feature space for both the original and augmented datasets. The intermediate outputs were extracted as embedding vectors when the augmented data passed through the same model. These embeddings enable a quantitative comparison of the distributional differences between datasets, allowing the statistical similarity between the original and augmented data to be evaluated without retraining multiple models.

Step 3: Similarity analysis. The difference between the two datasets in the embedding space was measured using the WD, which quantifies the distance between data distributions in the embedding space. As the augmentation ratio increases, the data distribution gradually diverges from the original distribution and saturates after a certain point. This saturation point represents the stage at which the similarity between datasets no longer changes and additional training results in a negligible improvement in performance.

Through these steps, the SAPF captures changes in data distributions with different augmentation ratios and provides an efficient and practical method for predicting the optimal augmentation ratio before model training.

2.2 Construction of Original and Augmented Datasets

To validate the SAPF, we used a gas sensor array time-series dataset from the UCI machine-learning repository [26]. The dataset contains measurements from 72 metal oxide sensors collected under various gas exposure conditions. Each sample consists of 72-dimensional time-series signals measured at 100 Hz for approximately 260 s. A total of 14,400 time-series samples were analyzed in this study. Among the ten gases, carbon monoxide (450 samples) and butanol (1,500 samples) were excluded because of insufficient sample counts, leaving eight gases: acetaldehyde, benzene, ammonia, acetone, ethylene, methane, methanol, and toluene. All sensor values were normalized to the range [0, 1] using MinMaxScaler. Linear interpolation based on window warping was applied to account for differences in sequence length, and all sequences were standardized to 10,000 data points. Each preprocessed sample contained five components: time series data (X), voltage, flow, position, and label. A detailed description of the components is provided in Table 1.

The time-series data were first converted into mel spectrograms and reconstructed as a 72-channel input. SpecAugment randomly masks regions along the frequency axis. In this study, two masks with a width of five frequency bins were used. SpecSwap exchanges random frequency intervals using a swap width of seven bins and two swaps per sample. Both techniques were independently applied to each sensor channel to produce diverse transformations without significantly distorting the original data distribution.

The dataset was structured to evaluate the augmentation effect using the SAPF framework, as shown in Table 2. The entire dataset was divided into subsets for SAPF evaluation, consisting of 90% (11,520 samples) for analysis and 10% (2,880 samples) for verification. From the evaluation subset, three groups of 3,600, 7,200, and 11,520 samples were selected to examine how the augmentation effect varied with different dataset sizes. The augmentation ratios were set to ×2, ×5, ×10, ×25, ×50, and ×100. Each augmented dataset was configured to maintain the same conditional distributions (voltage, flow rate, position, and label) as the original dataset. To achieve this, the original data were first uniformly sampled to match the target quantity for each condition and the remaining samples were supplemented with augmented data. This procedure was applied consistently across all the experimental groups to prevent imbalances from influencing the results. For example, when the original sample size was 3,600, the ×5 dataset included 14,400 augmented samples, which was five times the number of the original samples, resulting in 18,000 samples.

Table 1. Components of the preprocessed samples used in SAPF evaluation.

Component	Description
X	Time-series measurements from 72 sensors
Voltage	Heater voltage condition
Flow	Inflow velocity (airflow) condition
Position	Sensor position
Label	Gas type label

Table 2. Composition of the original and augmented datasets across augmentation ratios (×1 to ×100).

Original Sample Size	Aug. Ratio	Sample Size of Aug.	Total Sample Size
3,600	×1	0	3,600
	×2	3,600	7,200
	×5	14,400	18,000
	×10	32,400	36,000
	×25	86,400	90,000
	×50	176,400	180,000
	×100	356,400	360,000
7,200	×1	0	7,200
	×2	7,200	14,400
	×5	28,800	36,000
	×10	64,800	72,000
	×25	172,800	180,000
	×50	352,800	360,000
	×100	712,800	720,000
11.520	×1	0	11,520
	×2	11,520	23,040
	×5	46,080	57,600
	×10	103,680	115,200
	×25	276,480	288,000
	×50	564,480	576,000
	×100	1,140,480	1,152,000

Note. Aug.: Augmentation.

2.3 Similarity Evaluation Criteria

This section introduces the concept of embedding-based similarity analysis used in the SAPF to quantitatively evaluate the effects of data augmentation.

Conventional approaches determine the optimal augmentation ratio by training multiple models for each augmentation condition and comparing their classification performances. However, this process requires substantial computational resources and a long training time, making it inefficient for large-scale evaluations.

To address this issue, this study proposes an embedding-based similarity analysis that enables the quantitative pre-evaluation of augmentation effects without repeated model retraining. Instead of relying on accuracy comparisons across multiple trained models, the SAPF measures how the feature distributions of the original and augmented data differ within a shared representational space. Embedding vectors were extracted from an intermediate layer of a deep learning model trained on the original dataset, providing a consistent reference for evaluating the distributional changes caused by augmentation. As the reliability of this representational space depends on the stability of the learned parameters, the model must first be sufficiently trained to update its parameters to reflect the statistical structure of the original data.

In this study, a deep-learning model based on the Bi-GRU architecture was implemented as a Pre-trained Model (×1). The network integrates time-series inputs (128×72) with conditional inputs representing voltage, flow, and position. Time-series inputs were sequentially processed using two Bi-GRU layers, followed by a dense layer of 128 units. During the training stage, the model learned the representative statistical structure of the original dataset and established a stable representational space for subsequent analyses. The overall architecture of the Pre-trained Model (×1) is shown in Fig. 2.

After the model is sufficiently trained, one of its intermediate layers, specifically the dense layer output obtained before concatenation with the conditional input, is defined as the embedding layer. When the augmented datasets were passed through this Pre-trained Model, the intermediate outputs were extracted as embedding vectors. These vectors served as the basis for similarity analysis in the SAPF, allowing the distributional differences between the original and augmented datasets to be quantitatively compared within the same representational space.

The differences between the embedding distributions of the two datasets were quantified using a first-order WD. The WD is derived from the optimal transport (OT) theory and measures the minimal cost required to transport the mass between two data distributions. Unlike conventional metrics, which only consider differences in the mean or variance, the

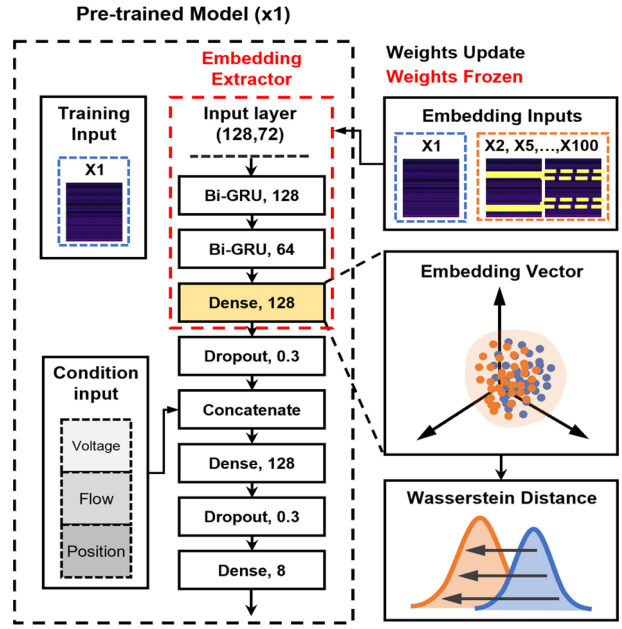

Fig. 2. Architecture of the Bi-GRU model used as the embedding extractor in SAPF. The model is trained once using the original dataset (×1) to learn representative features and is subsequently frozen to extract intermediate dense-layer embeddings from both the original and augmented datasets (×1 to ×100) for WD comparison.

WD can capture discrepancies in the overall distribution shape [27]. Specifically, for two embedding matrices $A, B \in \mathbb{R}^{N\times d}$, where N denotes the number of samples and d represents the embedding dimension, the one-dimensional WD, $W_1(A_i,B_i)$, is calculated for each dimension $i = 1,\ldots,d$, and the overall distance is defined as the average across all dimensions. The complete formulation is given as

$$WD(A,B) = \frac{1}{d}\sum_{i=1}^{d} W_1(A_i,B_i), \tag{1}$$

where A_i and B_i denote the i-th column of the embedding matrices A and B, respectively, and W_1 is WD between two one-dimensional real-valued distributions. The WD is used as the similarity criterion in SAPF because it measures the minimal distributional shift between the original and augmented data. This metric provides an interpretable basis for quantitative comparison.

3. RESULTS AND DISCUSSION

3.1 Experimental Setup

All the experiments were conducted using Python 3.10, TensorFlow 2.10, and the main analyses were performed using NumPy, SciPy, scikit-learn, and Matplotlib. All the computations were performed on a workstation equipped with

an NVIDIA GeForce RTX 5070 Ti GPU (16 GB VRAM), an Intel® Core™ Ultra 7 265 K CPU, and 64 GB RAM. The primary objective of this experiment was to verify whether the SAPF could reliably determine the optimal augmentation ratio before model training. Three procedures were conducted. First, the WD was calculated for each augmentation ratio (×2 to ×100) using a Bi-GRU embedding extractor trained on the original (×1) dataset, and the results were used to estimate the optimal augmentation ratio. Second, the same augmented datasets were trained using the Bi-GRU and Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) models to verify whether the predicted optimal ratio was valid in terms of classification performance, ensuring that the SAPF results were generalizable and independent of specific model architectures. Third, the Spearman correlation coefficient was computed between the similarity and performance metrics to quantitatively assess their relationship, and the training time was compared with that of conventional approaches to evaluate the practical efficiency of the SAPF. This validation process proved the SAPF to be not only a distribution comparison tool but also a reliable methodology capable of identifying the optimal augmentation ratio in advance while maintaining both model performance and training efficiency.

3.2 Similarity and Performance Evaluation Results

Fig. 3 shows the variation in the WD with respect to the augmentation ratio for the SpecAugment and SpecSwap techniques. The results indicate that the WD steadily increased as the augmentation ratio increased, indicating that the distributional difference between the original and augmented data gradually increased. However, the rate of increase declined sharply at approximately ×50, and the WD values saturated in the ×25 to ×100 range.

To further examine this trend, the WD_{slope} was calculated for each section. The slope between the two augmentation ratios x_a and x_b is defined in Eq. (2), where x_a and x_b denote the augmentation ratios at the beginning and end of the comparison interval, respectively. For example, x_a = 2 and x_b = 25 correspond to the WD_{slope} between ×2 and ×25, respectively.

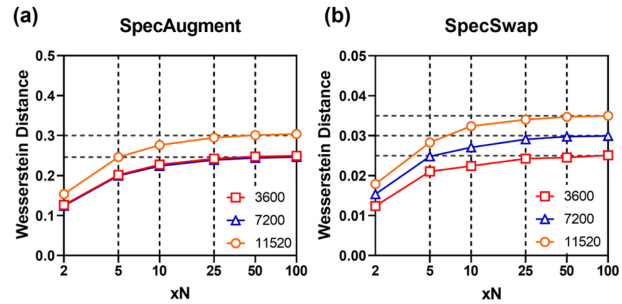

Fig. 3. Variation in the WD with respect to the augmentation ratio for the two augmentation methods. (a) SpecAugment showing a monotonic increase in WD with increasing augmentation ratio. (b) SpecSwap showing a similar trend, where the growth rate sharply decreases beyond ×50, indicating a saturation region with negligible change by further augmentation.

$$WD_{slope} = \frac{\Delta WD}{\Delta x} = \frac{WD(x_b) - WD(x_a)}{x_b - x_a} \quad (2)$$

As shown in Table 3, the WD_{slope} between ×2 and ×25 was relatively large for both SpecAugment and SpecSwap, whereas the slope between ×50 and ×100 was approximately two orders of magnitude smaller. This indicates that, although data diversity expands rapidly during the initial stages of augmentation, little new information is introduced beyond a certain point, and only the repetition of existing patterns is reinforced, effectively reaching a saturation state.

In addition, the change in classification accuracy between the two augmentation ratios is defined as ΔAcc. ΔAcc represents the accuracy difference between x_a and x_b and is calculated in the same manner as the WD_{slope}. The ΔAcc analysis likewise showed a distinct improvement between ×2 and ×25; however, the improvement was marginal beyond ×50. Accordingly, SAPF identified ×50 as the optimal augmentation ratio from a practical standpoint, based on a comprehensive consideration of both WD_{slope} and ΔAcc.

The classification results were consistent with the SAPF

Table 3. WD_{slope} ($\times 10^{-3}$) and accuracy difference (ΔAcc) between augmentation intervals for SpecAugment and SpecSwap.

Original Sample Size	Data Augment Method	WD_{slope} ($\times 10^{-3}$) (×2 to ×25)	WD_{slope} ($\times 10^{-3}$) (×50 to ×100)	ΔAcc (×2 to ×25)	ΔAcc (×50 to ×100)
3,600	SpecAugment	5.03	0.04	0.0868	0.0059
	SpecSwap	0.51	0.01	0.0750	0.0167
7,200	SpecAugment	4.97	0.05	0.0528	0.0059
	SpecSwap	0.60	0.01	0.0677	0.0059
11,520	SpecAugment	6.13	0.02	0.0487	0.0004
	SpecSwap	0.70	0.01	0.0583	0.0052

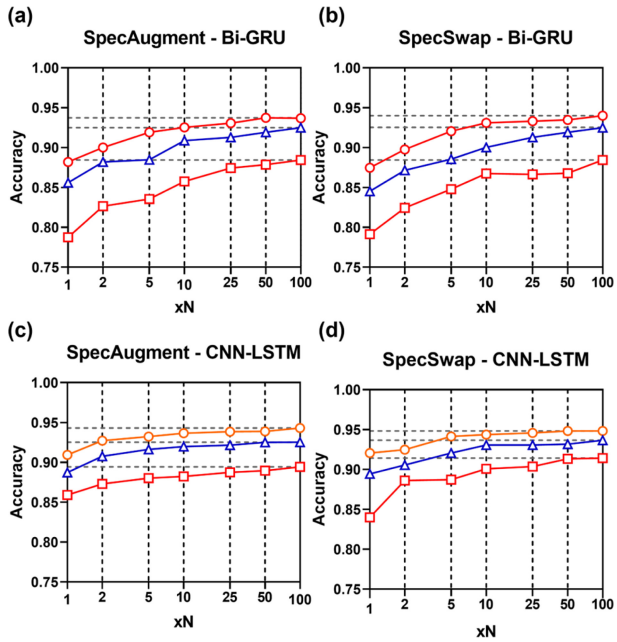

Fig. 4. Classification accuracy of Bi-GRU and CNN-LSTM models under different augmentation ratios. (a) Bi-GRU (SpecAugment), (b) Bi-GRU (SpecSwap), (c) CNN-LSTM (SpecAugment), and (d) CNN-LSTM (SpecSwap). Accuracy rapidly improves between ×2 and ×25 but saturates beyond ×50, with differences ≤0.01 between ×50 and ×100, verifying that the performance saturation predicted by SAPF corresponds to the experimentally observed trend.

predictions. As shown in Fig. 4 and Table 3, the accuracy of both the Bi-GRU and CNN-LSTM models improved significantly during the early augmentation stage (×2 to ×25) but slowed thereafter and remained nearly unchanged beyond ×50. In particular, the ΔAcc between ×50 and ×100 for both Bi-GRU and CNN-LSTM was ≤0.01, confirming that the optimal ratio (×50) and the performance saturation range (×50 to ×100) suggested by SAPF were consistent with the actual learning outcomes.

These findings suggest that SAPF is not merely a tool for quantifying distributional differences but is also a reliable indicator capable of predicting model performance in advance.

3.3 Analysis of Similarity–Performance Correlation and Training Efficiency

The Spearman rank correlation coefficient was calculated to quantitatively examine the relationship between similarity and the performance metrics [28]. The Spearman correlation coefficient measures the strength and direction of a monotonic relationship between two variables, where ρ represents the correlation strength. A value close to +1 indicates a strong positive correlation, meaning that both metrics increase simultaneously. Statistical significance was evaluated using the p-value, with correlations considered significant at p-value <0.05. The experimental results are presented in Table 4.

Table 4. Spearman correlation (ρ) and significance (p-value) between WD and classification accuracy across sample sizes and augmentation methods.

Original Sample Size	Data Augment Method	ρ	p-value
3,600	SpecAugment	1	0
	SpecSwap	0.9643	0.0045
7,200	SpecAugment	1	0
	SpecSwap	1	0
11,520	SpecAugment	0.9643	0.0045
	SpecSwap	1	0

The analysis confirmed a very strong positive correlation between WD and accuracy under all conditions (ρ = 0.94–1.00), with statistical significance achieved at p-value <0.05. This indicates that accuracy increases with higher augmentation ratios, demonstrating strong consistency between the two metrics. The Spearman correlation analysis between WD and accuracy is presented in Table 4. These findings strongly suggest that the similarity metric alone can reliably predict performance trends to a considerable degree.

To evaluate the training cost reduction effect quantitatively, two methods were compared: conventional training and SAPF. The conventional training approach follows the conventional procedure of training a separate model for each augmentation ratio (×1 to ×100). As the size of the augmented dataset increases, the training costs increase exponentially. In contrast, the SAPF was trained only on the original dataset (×1), computed a pre-evaluation index, and derived the optimal augmentation ratio accordingly. Actual model training was performed once.

Consequently, SAPF maintains a constant training time regardless of the number of augmentation ratios and maximizes resource efficiency by eliminating redundant training. The training time (T) and computational cost (floating-point operations (FLOPs)) were used as evaluation metrics. The training time reduction ratio is defined as

$$Reduction\ Ratio_{time}\,(\%) = \frac{T_{baseline} - T_{SAPF}}{T_{baseline}} \times 100, \tag{3}$$

where $T_{baseline}$ denotes the total training time required to train across all augmentation ratios in the conventional training method and T_{SAPF} represents the training time required to train only on the original dataset (×1) using the SAPF approach. The reduction rate was obtained by dividing the difference between these two values by $T_{baseline}$. The computational resource reduction rate is defined as

Table 5. Comparison of training time and computational cost (FLOPs) between the conventional training and SAPF methods.

Original Sample Size	Conventional training Time (h)	SAPF Time (h)	Reduction Ratio (Time, %)	Conventional training (FLOPs)	SAPF (FLOPs)	Reduction Ratio (%)
3,600	49.06	0.27	99.45	1.96E+15	0.01E+15	99.42
7,200	73.95	0.53	99.28	3.91E+15	0.02E+15	99.42
11,520	128.55	0.85	99.34	6.26E+15	0.03E+15	99.42

$$Reduction\ Ratio_{FLOPs}\ (\%) = \frac{F_{baseline} - F_{SAPF}}{F_{baseline}} \times 100\,, \qquad (4)$$

where $T_{baseline}$ denotes the total computational cost required by conventional training, and T_{SAPF} denotes that of SAPF. FLOPs represent the number of floating-point operations performed during model training, and serve as a direct measure of computational complexity. Thus, a comparison of the FLOPs provides a clear quantification of the efficiency of the proposed method in achieving the same outcome with fewer operations. The reduction effects on the training time and computational cost are presented in Table 5. Across all dataset sizes, the SAPF achieved a reduction of more than 99.28% in training time and 99.42% in computational cost compared with conventional training. Specifically, when the original sample size was 3,600, the total training time decreased from 49.06 h to 0.27 h (99.45% reduction), and the computational cost decreased from 1.96×10^{15} to 0.01×10^{15} FLOPs (99.42% reduction). For 7,200 samples, SAPF reduced the training time from 73.95 h to 0.53 h (99.28% reduction), and FLOPs from 3.91×10^{15} to 0.02×10^{15} (99.42% reduction). Similarly, for 11,520 samples, the time decreased from 128.55 h to 0.85 h (99.34% reduction), and the computational cost decreased from 6.26×10^{15} to 0.03×10^{15} FLOPs (99.42% reduction).

These consistent results across all dataset sizes clearly demonstrate that the SAPF drastically reduces both the training time and computational demand by more than two orders of magnitude while maintaining comparable accuracy. This confirms that the SAPF is not only efficient but also practically scalable for large-scale augmentation evaluations.

4. CONCLUSIONS

This study proposed a SAPF approach to pre-evaluate the effectiveness of data augmentation in deep learning models and validated its feasibility using time-series gas sensor data. The SAPF quantitatively evaluates the distributional similarity between the original and augmented datasets without repeated model training using the WD computed from the Bi-GRU embeddings trained on the original dataset (×1). The experimental results showed that as the augmentation ratio increased, both the WD and accuracy exhibited substantial improvement in the early stage (×2 to ×25), whereas the slope decreased by more than two orders of magnitude in the later stage (×50 to ×100), indicating performance saturation. Based on this trend, SAPF identified ×50 as the optimal augmentation ratio. In the actual Bi-GRU and CNN-LSTM training, the difference in accuracy between ×50 and ×100 was less than 0.01, which was consistent with the SAPF prediction. The Spearman correlation coefficient between the WD and accuracy ranged from 0.94 to 1.00 under all conditions, confirming that the similarity metric alone can reliably predict performance trends.

In terms of training cost, the SAPF achieved a reduction of over 99.28% in training time and 99.42% in computational cost compared to conventional methods without any loss of performance. Therefore, the SAPF serves not only as a distribution comparison tool but also as a practical framework that enables efficient and reliable determination of the optimal augmentation ratio, reducing unnecessary data generation and repetitive training in the pre-training stage SAPF is applicable beyond gas-sensing datasets, and future extensions to diverse datasets and augmentation techniques are expected to further enhance their versatility and applicability to time-series data analysis.

CRediT Authorship Contribution Statement

Gyu-Li Kim: Conceptualization, Methodology, Data curation, Formal analysis, Investigation, Software, Visualization, Writing – original draft. **Kwangjae Lee:** Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that may have influenced the work reported in this paper.

Acknowledgements

No external funding was received for the study.

REFERENCES

[1] Y. Li, X. Yu, N. Koudas, Data acquisition for improving machine learning models, ArXiv., https://arxiv.org/abs/

2105.14107 (2021).
[2] S.W. Park, H.M. Joe, CNN-based Fall Detection Model for Humanoid Robots, J. Sens. Sci. Technol. 33 (2024) 18–23.
[3] G.L. Kim, S.J. Ro, K. Lee, A Multi-Sensor Fire Detection Method based on Trend Predictive BiLSTM Networks, J. Sens. Sci. Technol. 33 (2024) 248–254.
[4] S.J. Ro, K. Lee, Early Fire Detection System for Embedded Platforms: Deep Learning Approach to Minimize False Alarms, J. Sens. Sci. Technol. 33 (2024) 298–304.
[5] F. Bargagna, L.A. De Santi, N. Martini, D. Genovesi, B. Favilli, G. Vergaro, et al., Bayesian convolutional neural networks in medical imaging classification: a promising solution for deep learning limits in data scarcity scenarios, J. Digit. Imaging 36 (2023) 2567–2577.
[6] J. Banerjee, J.N. Taroni, R.J. Allaway, D.V. Prasad, J. Guinney, C.S. Greene, Machine learning in rare disease, Nat. Methods 20 (2023) 803–814.
[7] C. Li, T. Denison, T. Zhu, A survey of few-shot learning for biomedical time series, IEEE Rev. Biomed. Eng. 18 (2024) 192–210.
[8] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (2019) 60.
[9] A. Mumuni, F. Mumuni, Data augmentation: a comprehensive survey of modern approaches, Array 16 (2022) 100258.
[10] Y. Wang, G. Huang, S. Song, X. Pan, Y. Xia, C. Wu, Regularizing deep networks with semantic data augmentation, IEEE Trans. Pattern Anal. Mach. Intell. 44 (2020) 3733–3748.
[11] B.S. Park, S.M. Im, H. Lee, Y.T. Lee, C. Nam, S. Hong, et al., Visual and tactile perception techniques for braille recognition, Micro Nano Syst. Lett. 11 (2023) 23.
[12] L. Taylor, G. Nitschke, Improving deep learning with generic data augmentation, Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 2018, pp. 1542–1547.
[13] E.K. Kim, H. Lee, J.Y. Kim, S. Kim, Data augmentation method by applying color perturbation of inverse PSNR and geometric transformations for object recognition based on deep learning, Appl. Sci. 10 (2020) 3755.
[14] A.L.C. Ottoni, R.M. de Amorim, M.S. Novo, D.B. Costa, Tuning of data augmentation hyperparameters in deep learning to building construction image classification with small datasets, Int. J. Mach. Learn. Cybern. 14 (2023) 171–186.
[15] W. Oronowicz-Jaskowiak, Empirical verification of the suggested hyperparameters for data augmentation using the fast.ai library, Mach. Learn. Appl. 7 (2022) 100222.
[16] T. Li, Y. Zhang, D. Su, M. Liu, M. Ge, L. Chen, et al., Knowledge graph-based few-shot learning for label of medical imaging reports, Acad. Radiol. 32 (2025) 4206–4220.
[17] N.E. Corrado, J.P. Hanna, Understanding when dynamics-invariant data augmentations benefit model-free reinforcement learning updates, ArXiv., https://arxiv.org/abs/2310.17786 (2023).
[18] D. Wagner, F. Ferreira, D. Stoll, R.T. Schirrmeister, S. Müller, F. Hutter, On the importance of hyperparameters and data augmentation for self-supervised learning, ArXiv., https://arxiv.org/abs/2207.07875 (2022).
[19] G. Iglesias, E. Talavera, Á. González-Prieto, A. Mozo, S. Gómez-Canaval, Data augmentation techniques in time series domain: a survey and taxonomy, Neural Comput. Appl. 35 (2022) 10123–10145.
[20] B.K. Iwana, S. Uchida, An empirical survey of data augmentation for time series classification with neural networks, Plos one 16 (2021) e0254841.
[21] K. Kamycki, T. Kapuscinski, M. Oszust, Data augmentation with suboptimal warping for time-series classification, Sensors 20 (2019) 98.
[22] I. Naiman, N. Berman, I. Pemper, I. Arbiv, G. Fadlon, O. Azencot, Utilizing image transforms and diffusion models for generative modeling of short and long time series, Adv. Neural Inf. Process. Syst. 37 (2024) 121699–121730.
[23] S. Yang, S. Guo, J. Zhao, F. Shen, Investigating the effectiveness of data augmentation from similarity and diversity: an empirical study, Pattern Recognit. 148 (2024) 110204.
[24] D.S. Park, W. Chan, Y. Zhang, C.C. Chiu, B. Zoph, E.D. Cubuk, et al., SpecAugment: a simple data augmentation method for automatic speech recognition, ArXiv., https://arxiv.org/abs/1904.08779 (2019).
[25] X. Song, Z. Wu, Y. Huang, D. Su, H. Meng, SpecSwap: a simple data augmentation method for end-to-end speech recognition, Proceedings of INTERSPEECH 2020, Shanghai, China, 2020, pp. 581–585.
[26] A. Vergara, S. Vembu, T. Ayhan, M.A. Ryan, M.L. Homer, R. Huerta, Gas sensor arrays in open sampling settings [Dataset]. UCI Machine Learning Repository, University of California, Irvine (2013). Available at: https://doi.org/10.24432/C5JP5N.
[27] E.F. Montesuma, F.M.N. Mboula, A. Souloumiac, Recent advances in optimal transport for machine learning, IEEE Trans. Pattern Anal. Mach. Intell. 47 (2024) 1161–1180.
[28] J.H. Zar, Spearman rank correlation, Encycl. Biostat. 7 (2005) 4990–4998.