Spatial Variation Generation Algorithm for Motor Imagery Data Augmentation: Increasing the Density of Sample Vicinity

Chengxuan Qin, Rui Yang, *Member, IEEE*, Mengjie Huang, *Member, IEEE*, Weibo Liu, *Member, IEEE*, and Zidong Wang, *Fellow, IEEE*

***Abstract*— The imbalanced development between deep learning-based model design and motor imagery (MI) data acquisition raises concerns about the potential overfitting issue—models can identify training data well but fail to generalize test data. In this study, a Spatial Variation Generation (SVG) algorithm for MI data augmentation is proposed to alleviate the overfitting issue. In essence, SVG generates MI data using variations of electrode placement and brain spatial pattern, ultimately elevating the density of the raw sample vicinity. The proposed SVG prevents models from memorizing the training data by replacing the raw samples with the proper vicinal distribution. Moreover, SVG generates a uniform distribution and stabilizes the training process of models. In comparison studies involving five deep learning-based models across eight datasets, the proposed SVG algorithm exhibited a notable improvement of 0.021 in the area under the receiver operating characteristic curve (AUC). The improvement achieved by SVG outperforms other data augmentation algorithms. Further results from the ablation study verify the effectiveness of each component of SVG. Finally, the studies in the control group with varying numbers of samples show that the SVG algorithm consistently improves the AUC, with improvements ranging from approximately 0.02 to 0.15.**

***Index Terms*— Brain–computer interfaces, data augmentation, deep learning, electroencephalogram.**

Manuscript received 12 May 2023; revised 1 September 2023; accepted 8 September 2023. Date of publication 12 September 2023; date of current version 20 September 2023. This work was supported in part by the Jiangsu Provincial Qinglan Project, in part by the Suzhou Science and Technology Program under Grant SYG202106, in part by the Research Development Fund of Xi'an Jiaotong-Liverpool University (XJTLU) under Grant RDF-18-02-30 and Grant RDF-20-01-18, and in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant 20KJB520034. *(Corresponding authors: Rui Yang; Mengjie Huang.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University Ethics Committee of Xi'an Jiaotong–Liverpool University under Proposal No. EXT20-01-07, March 31, 2020.

Chengxuan Qin is with the School of Advanced Technology, Xi'an Jiaotong–Liverpool University, Suzhou 215123, China, and also with the School of Electrical Engineering, Electronics and Computer Science, University of Liverpool, L69 3BX Liverpool, U.K. (e-mail: Chengxuan.Qin21@student.xjtlu.edu.cn).

Rui Yang is with the School of Advanced Technology, Xi'an Jiaotong–Liverpool University, Suzhou 215123, China (e-mail: R.Yang@xjtlu.edu.cn).

Mengjie Huang is with the Design School, Xi'an Jiaotong–Liverpool University, Suzhou 215123, China (e-mail: Mengjie.Huang@xjtlu.edu.cn).

Weibo Liu and Zidong Wang are with the Department of Computer Science, Brunel University London, Uxbridge, UB8 3PH Middlesex, U.K. (e-mail: Weibo.Liu2@brunel.ac.uk; Zidong.Wang@brunel.ac.uk).

Digital Object Identifier 10.1109/TNSRE.2023.3314679

I. Introduction

ELECTROENCEPHALOGRAM (EEG) is a kind of physiological signal collected by brain-computer interfaces (BCIs) having the ability to infer various brain activities in several paradigms. Motor imagery (MI), recorded from populations of human posterior parietal cortex neurons, is one of the paradigms representing the rehearsal of movements [1]. With the proper classification of MI signals activated from different limbs (such as left or right hand [2], [3], [4], [5], [6], [7]), MI-based BCI can be developed as a promising technique to assist people with disabilities to control external devices (such as a robotic limb or computer cursor [1]).

Recently studies on the classification of MI signals using deep learning have gained attention due to their impressive generalization abilities and accurate results. Deep learning-based models typically possess a large number of parameters, prompting researchers to focus on balancing efficiency with data requirements to prevent overfitting, whereby the model can identify the training data well but cannot generalize the test data. From the perspective of model efficiency. the principal strategies to mitigate the overfitting issue can be categorized into three categories: 1) optimization of parameter numbers, such as incorporating depthwise or separable convolutions [4], [8], pooling layers [3], [9], and progressively reducing the number of layers [10]; 2) optimization of feature extraction efficiency, such as employing a multi-branch structure [11] and an inception block [10]); 3) dynamic modification of models, such as adding dropout regularization [3], [5], [10], [12]. Empirical evidence suggests that such a well-designed architecture can partially alleviate the issue of overfitting. Besides, in terms of the amount of data required, a substantial dataset is typically required for effective training and accurate generalization [13], [14], [15], [16].

Data scarcity remains a pervasive issue in the field of EEG. One of the most substantial challenges affecting accurate classification and good generalization is termed inter-device incompatibility in this paper. The inter-device incompatibility suggests that conventional deep learning-based models are

prone to domain shift problems from changes in tasks or scenarios and cannot employ the EEG data gathered under diverse conditions [17], [18]. For instance, it would be infeasible to apply a model trained on the Neuroscan SynAmps RT acquisition system to classify Biosemi ActiveTwo acquisition system data [19]. The inter-device incompatibility arises from various factors, including but not limited to onboard filter characteristics, electrode types, contact methods, electrode locations, or online reference schemes [20]. Therefore, the number of samples available in an EEG dataset is insufficient compared to the fields where deep learning-based models are commonly used, such as natural language processing or computer vision. Overall, the limitations caused by inter-device incompatibility contribute to critical data scarcity, ultimately resulting in the diminished performance of deep learning-based models in MI data classification.

To address the challenge of data scarcity, researchers have focused on data augmentation as a potential solution, aiming to generate augmented data (satisfying the vicinal distribution) via the modification or recombination of existing data (satisfying the empirical distribution) according to specific inductive bias [21], [22], [23]. In the field of EEG, the data augmentation methods can be classified based on the information used for modification or recombination, including temporal [3], [5], [24], [25], [26], spatial [3], [27], spectral [26], [28], and component information [29], [30]. However, the existing work on data augmentation with EEG data is still limited. In particular, there has been inadequate consideration given to the variations in electrode placement and brain spatial pattern, as described as follows:

A. Variation of Electrode Placement

Each channel's signal comes from its electrode, making electrode placement a crucial factor in measurement accuracy. Among the commonly used commercial EEG caps, the electrodes are generally positioned following a pre-defined montage or placement scheme (such as international 10/5, 10/10, 10/20 systems [31]). However, the actual electrode positions cannot be precisely captured by the above two placement schemes, due to measurement errors (mainly caused by personal error and dynamic electrode-skin contact) [32] and individual differences (mainly caused by incompatibility between montages/schemes and subjects) [33], [34].

B. Variation of Brain Spatial Pattern

As the neocortex of the human brain is divided into specialized cortical areas for different functions [35], the corresponding spatial pattern of brain activity is distinct to different behaviors. However, physiological studies have shown that there exist slight variations in the brain network. Brain imaging techniques suggest that the brain network and connectivity exhibit temporal variations over a period of time [36], [37], [38], [39], especially between the two hemispheres (dominant/non-dominant) [4], [5], [40], [41], [42]. From the histological perspective, the gyri and sulci of the cerebral cortex show inter-subject variability [43], therefore the dynamic distribution of neurons along gyri and sulci can lead to differences in the spatial patterns of EEG.

Based on the above discussions, the variations of electrode placement and brain spatial pattern are common and unavoidable in experiments. One of the main problems caused by these variations is the relative spatial movement of activated brain regions and electrode positions. To make samples representative and increase model robustness to such variations, a spatial variation generation (SVG) algorithm for MI data augmentation is proposed in this paper. This study poses several difficulties that need to be addressed: 1) how to design a robust model that can translate variations of electrode placement and brain spatial pattern into spatial transformation applicable to MI data; 2) how to define meaningful and effective mathematical expressions for perturbations caused by variations; 3) how to control the time complexity of the algorithm in model training. Therefore, this paper contributes to the field in three main ways:

1) A thorough investigation of the variations in electrode placement and brain spatial patterns is presented, providing novel insights to the data augmentation with EEG signal;

2) The proposal of a novel approach to mathematically express how the variations of electrode placement and brain spatial pattern hold spatial influence over EEG signal;

3) The feasibility that MI data augmentation using variations of electrode placement and brain spatial pattern is experimentally proven.

The remaining sections of this paper are organized as follows: Section II describes the proposed algorithm, Section III presents the experimental setup and provides a detailed analysis of the results, and the conclusions and discussions on relevant future work are presented in Section IV.

II. Methodology

This section details the proposed algorithm for MI data augmentation. The first part introduces preliminary knowledge related to empirical risk minimization (ERM) and vicinal risk minimization (VRM) to establish the foundational concepts for the algorithm. The second part describes the perturbations arising from variations associated with electrode placement and brain spatial pattern. The third part details the mathematical modeling of the above-mentioned variations. Lastly, the proposed algorithm is presented step by step with pseudocode provided. The list of symbols in this section is shown in Table I.

A. Preliminary Knowledge

Data augmentation has been interpreted in many ways. Inspired by the works related to VRM [21], [22], the working principle of data augmentation is explained from the perspective of VRM.

1) Empirical Risk Minimization: Let f represents the mapping model, $X = \{x_i\}_{i=1}^{n}$ represents the sampled data and $Y = \{y_i\}_{i=1}^{n}$ represents the labels of sampled data. To search a model f describing the real joint distribution of the data $(x, y) \sim P(x, y)$, the expected risk $R(f)$ (loss function $\mathcal{L}$

TABLE I
LIST OF SYMBOLS

Symbol	Description
f	Mapping model
f^*	Optimal mapping model
(x_i, y_i)	Data pair
$(\widetilde{x}, \widetilde{y})$	Augmented data pair
$X = \{x_i\}_{i=1}^n$	Data set
$Y = \{y_i\}_{i=1}^n$	Label set
$R(f)$	Risk function
$R_{emp}(f)$	Empirical risk function
$\mathcal{L}$	Loss function
$P(x, y)$	Probability
$P_{emp}(x, y)$	Empirical probability distribution
δ	Dirac function
$\mathcal{T}_x(x)$	Transformation function
c	Channel count
S	Source signal
M	Measurement signal
Θ	Latitude
Λ	Longitude
Θ_e^r, Λ_e^r	Raw electrode positions
Θ_e^a, Λ_e^a	Augmented electrode positions
Θ_s^r, Λ_s^r	Raw signal positions
Θ_s^a, Λ_s^a	Augmented signal positions
ϵ^1	1-dimensional error term
ϵ^c	c-dimensional error term
$N^1(\mu, \Sigma)$	1-D Gaussian distribution
$N^c(\mu, \Sigma)$	c-D Gaussian distribution
μ	Mean
Σ	Covariance
$\{d_{ij}^r\}_{i,j=1}^c$	Distances for raw positions
$\{d_{ij}^a\}_{i,j=1}^c$	Distances for augmented positions
$F_\phi(D)$	Special operator
Φ^r	Raw weight matrix
Φ^a	Augmented weight matrix
$\mathcal{V}_e$	Variability of placement
$\mathcal{V}_s$	Brain spatial pattern
$\mathcal{M}$	Transformation model

over the joint distribution) is created by:

$$R(f) = \int \mathcal{L}(f(x), y) dP(x, y)$$

As the continuous and real distribution $P(x, y)$ is unknown, it is common to estimate the true distribution based on sampled data $X = \{x_i\}_{i=1}^n$ with Dirac mass function δ as the window for each sampled data. Then, the empirical distribution $P_{emp}(x, y)$ is defined by:

$$P_{emp}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \delta(x = x_i, y = y_i)$$

Then the empirical risk $R_{emp}(f)$ is defined by:

$$R_{emp}(f) = \int \mathcal{L}(f(x), y) dP_{emp}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(f(x_i), y_i)$$

The optimal model f^* can be obtained via the minimization of empirical risk $R_{emp}(f)$:

$$f^* = \arg\min_f R_{emp}(f) = \arg\min_f \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(f(x_i), y_i)$$

2) Vicinal Risk Minimization: The imbalanced development of deep learning-based model design and data acquisition raises the possibility of overfitting problems. To address this, the concept of ERM has been modified to VRM, with each initial example replaced with a distribution representing the desired invariances [22]. Specifically, the estimated distribution is adjusted to vicinal distribution $P_{vic}(\widetilde{x}, \widetilde{y})$ denoted by:

$$P_{vic}(\widetilde{x}, \widetilde{y}) = \frac{1}{n} \sum_{i=1}^{n} v(\widetilde{x}, \widetilde{y} | x_i, y_i)$$

where $v(\widetilde{x}, \widetilde{y} | x_i, y_i)$ represents the estimated density in the vicinity of (x_i, y_i), as shown below:

$$v(\widetilde{x}, \widetilde{y} | x_i, y_i) = \mathbb{E}_k[\delta(\widetilde{x}_k = \mathcal{T}_x^k(x_i), \widetilde{y}_k = \mathcal{T}_y^k(y_i))] \quad (1)$$

where $(\widetilde{x}, \widetilde{y})$ represents the augmented data generated according to specific expert knowledge.

B. Specific Perturbations From Variations

To effectively utilize variations in data augmentation, it is essential to establish a clear understanding of the perturbation exerted by the variations. Supposing there exists a raw state of the brain model with EEG cap (as shown in Fig. 1(a)), the variations may cause four kinds of perturbations, including flipping, scaling, rotation and distortion (as shown in Fig. 1(b) to 1(e) respectively). The detailed explanations of each type of perturbation are provided as follows:

- *Flipping*: representing the swapping of EEG signals between hemispheres. In practical MI applications, there exists both symmetry and laterality in MI data [4], [5], [40], [41], [42]. Brain symmetry reveals the symmetric marginal distribution of certain components in MI data, and laterality (such as the significance in the dominant hemisphere [40], [41], [42]) causes variations in brain activities across hemispheres among subjects and sessions.
- *Scaling*: representing the shrinking or expanding of electrode positions relative to a reference location of the EEG cap. In practical MI applications, due to the variations in scalp shape (e.g., head circumferences fluctuate dramatically among subjects [44], [45], [46]), even though most EEG caps are stretchable, there is no guarantee that the cap can be stretched/scaled to the proper electrode positions.
- *Rotation*: representing the rotational offset of all electrodes around a specific axis. In practical MI-related applications, due to the measurement error at the calibration stage, the angle of the EEG cap may vary among subjects and sessions, eventually leading to the rotation of electrode positions [27].
- *Distortion*: representing random shifts of certain electrode positions [47]. In practical MI-related applications, due to the irregular surface of the cap and measurement error (e.g., unevenness of the hair and scalp [34], [43], [48] and dynamic skin-electrode contact [32]), the electrodes or attached structures may be shifted disorderly, leading to electrode position distortions.

C. Mathematical Modeling of Variations Associated With Electrode Placement and Brain Spatial Pattern

When considering using variation as expert knowledge for data augmentation, it is necessary to construct a mathematical model describing these variations. In practical MI applications,

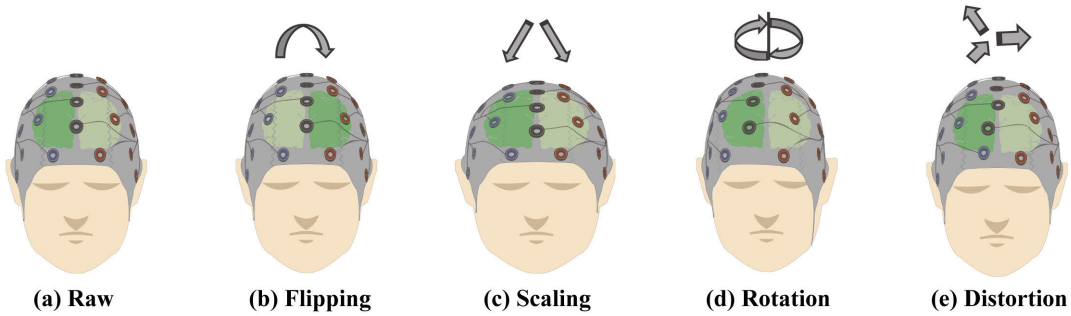

Fig. 1. Perturbations caused by variations associated with electrode placement and brain spatial pattern.

the sampled data can be considered as a matrix. Therefore, the spatial transformation $\mathcal{T}_x(x)$ can be estimated as a linear transformation matrix. To compute such a transformation matrix, the proposed SVG is constructed using variations of electrode placement and brain spatial pattern. The computation for estimation of variations of electrode placement and brain spatial pattern can be divided into three parts:

1) Definition of Electrode and Source Signal Positions: Let S and M represent source and measurement signals with specific descriptions: 1) source signals $S = [s_1, s_2, \cdots, s_c]^T$ refer to the signals generated by activated neurons; 2) measurement signals $M = [m_1, m_2, \cdots, m_c]^T$ refer to the signals received by EEG cap with certain electrode positions (c denotes the number of channels). Within this framework, the variations of the electrode positions and source signal positions (represented by a geographic coordinate system) are used to simulate the variations of electrode placement and brain spatial pattern, respectively.

2) Specification of Electrode and Source Signal Positions: For the expression of variations, at least two states associated with electrode and source signal positions need to be involved in computation: raw state (given by a preset montage) and augmented state (modified based on perturbations). Based on the types of positions and states mentioned above, the positions in the mathematical model can be divided into four different categories:

- raw electrode positions: $\{\Theta_e^r, \Lambda_e^r\}$;
- raw source signal positions: $\{\Theta_s^r, \Lambda_s^r\}$;
- augmented electrode positions: $\{\Theta_e^a, \Lambda_e^a\}$;
- augmented source signal positions: $\{\Theta_s^a, \Lambda_s^a\}$.

where Θ and Λ are c-dimensional vectors, representing the latitude and longitude coordinates of the electrodes, respectively (e.g., $\Theta_e^r = [\theta_{e_1}^r, \theta_{e_2}^r, \cdots, \theta_{e_c}^r]^T$, $\Lambda_e^r = [\lambda_{e_1}^r, \lambda_{e_2}^r, \cdots, \lambda_{e_c}^r]^T$).

3) Modification Rules: The augmented electrode and source signal positions are obtained by modification rules related to certain perturbations. As shown in Fig. 2, there are subtle differences between the augmented electrode positions (represented by blue solid circles) and source signal positions (represented by blue dashed circles) as compared to the raw electrode positions (represented by orange solid circles) and original signal positions (represented by orange solid circles). The subtle variations change the distances between the electrode positions and source signal positions, eventually altering the components of the signals captured by each electrode. In this way, the implicit expression of variations can be

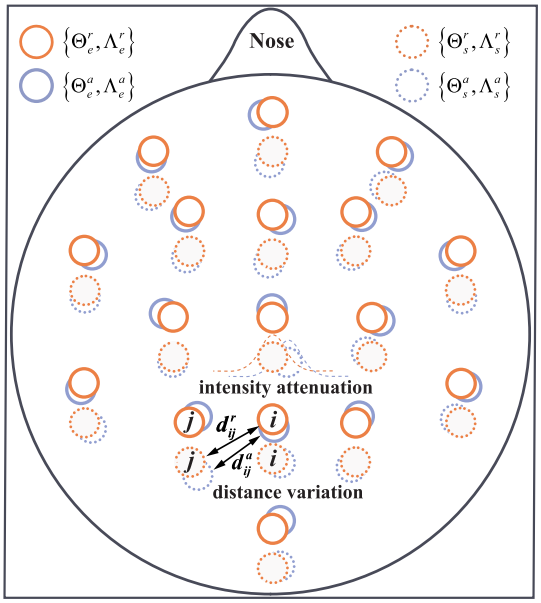

Fig. 2. Illustration of the mathematical modeling of variations.

established by rules taking $\{\Theta_e^r, \Lambda_e^r\}$, $\{\Theta_s^r, \Lambda_s^r\}$, $\{\Theta_e^a, \Lambda_e^a\}$, $\{\Theta_s^a, \Lambda_s^a\}$ as input. Fig. 3(a) shows the raw electrode and source signal positions, while Figs. 3(b)-(e) depict the augmented electrode and source signal positions after flipping, scaling, rotation, and distortion, respectively. Supposing the modification values follow a Gaussian distribution, the modification rules of flipping, scaling, rotation, and distortion are expressed by (2), (3), (4), and (5) respectively.

$$\begin{aligned} Flipping: \quad & \Theta_s^a = -\Theta_s^r \\ & \Lambda_s^a = \Lambda_s^r \\ & \Theta_e^a = \Theta_e^r \\ & \Lambda_e^a = \Lambda_e^r \end{aligned} \tag{2}$$

$$\begin{aligned} Scaling: \quad & \Theta_s^a = \Theta_s^r \\ & \Lambda_s^a = \Lambda_s^r \\ & \Theta_e^a = \Theta_e^r \\ & \Lambda_e^a = \epsilon_1(\Lambda_e^r - \frac{\pi}{2}) + \frac{\pi}{2}, \epsilon^1 \sim N^1(1, \Sigma_{sca}) \end{aligned} \tag{3}$$

$$\begin{aligned} Rotation: \quad & \Theta_s^a = \Theta_s^r \\ & \Lambda_s^a = \Lambda_s^r \end{aligned}$$

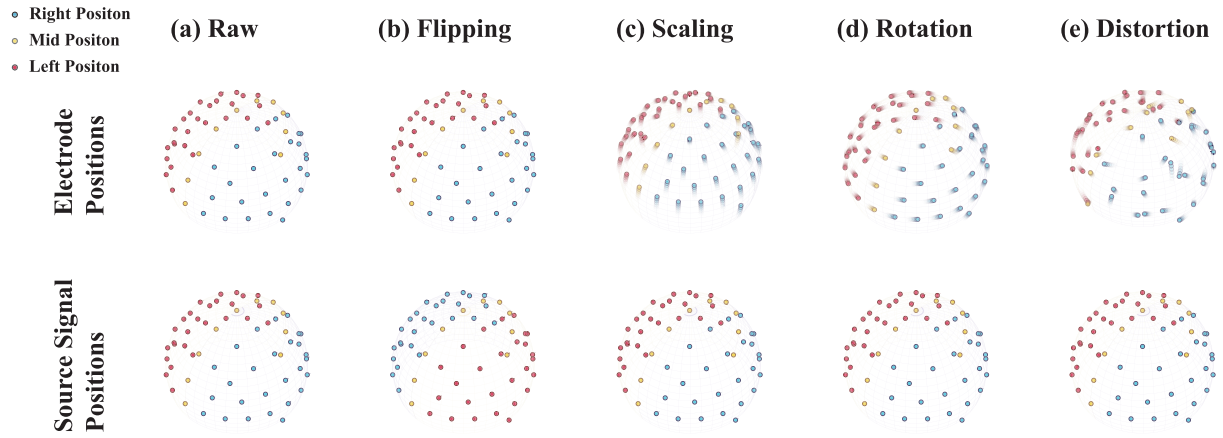

Fig. 3. Illustration of modification rules for corresponding perturbations.

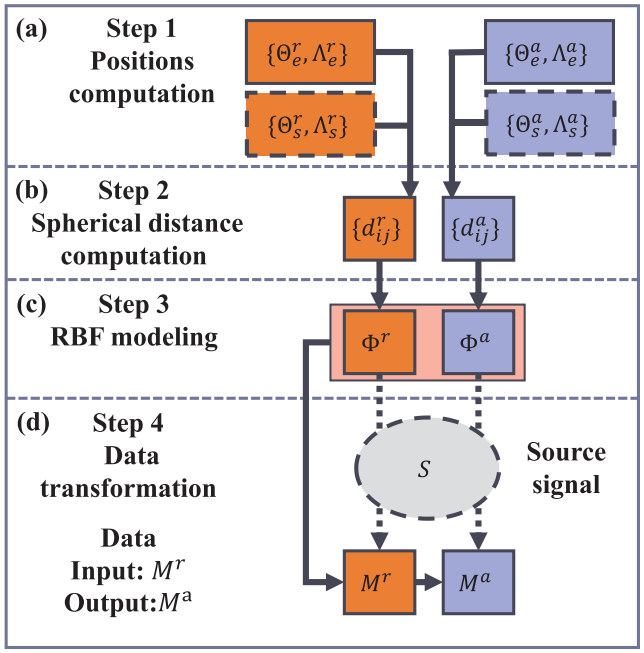

Fig. 4. Flowchart of proposed SVG algorithm.

$$
\begin{aligned}
&\Theta_e^a = \Theta_e^r + \epsilon^1, \epsilon^1 \sim N^1(0, \Sigma_{rot}) \\
&\Lambda_e^a = \Lambda_e^r \\
Distortion: \quad &\Theta_s^a = \Theta_s^r \\
&\Lambda_s^a = \Lambda_s^r \\
&\Theta_e^a = \Theta_e^r + \epsilon^c, \epsilon^c \sim N^c(0, \Sigma_{dis\theta}) \\
&\Lambda_e^a = \Lambda_e^r + \epsilon^c, \epsilon^c \sim N^c(0, \Sigma_{dis\lambda})
\end{aligned}
\tag{4}
\tag{5}
$$

where ϵ^1 and ϵ^c denote the 1-dimensional and c-dimensional perturbations respectively, and $N^1(\mu, \Sigma)$ and $N^c(\mu, \Sigma)$ denote the 1-dimensional and c-dimensional Gaussian distribution with mean μ and variance Σ.

D. Spatial Variation Generation Algorithm

The proposed SVG algorithm (with flowchart shown in Fig. 4) consists of the following four steps:

1) Step 1: Positions Computation: The raw electrode positions $\{\Theta_e^r, \Lambda_e^r\}$ and raw source signal positions $\{\Theta_s^r, \Lambda_s^r\}$ are set according to the pre-defined coordinates in EEGLAB (a leading tool for processing EEG signals [49], [50]). The augmented electrode/source signal positions $\{\Theta_e^a, \Lambda_e^a\}/\{\Theta_s^a, \Lambda_s^a\}$ are designed for data augmentation according to the modification rules of SVG, including flipping, scaling, rotation, and distortion. The selection of modification rules follows certain probability distributions.

2) Step 2: Spherical Distance Computation: The spherical distance between the electrode and source signal positions (with sphere radius set to 1) can be computed by the Great-Circle Distance [51]:

$$d_{ij}^r = \cos^{-1}(\sin\theta_{e_i}^r \sin\theta_{s_j}^r + \cos\theta_{e_i}^r \cos\theta_{s_j}^r \cos(\lambda_{e_i}^r - \lambda_{s_j}^r)) \tag{6}$$

$$d_{ij}^a = \cos^{-1}(\sin\theta_{e_i}^a \sin\theta_{s_j}^a + \cos\theta_{e_i}^a \cos\theta_{s_j}^a \cos(\lambda_{e_i}^a - \lambda_{s_j}^a)) \tag{7}$$

where i and j denote the index of electrode and source signal positions, respectively; $\{d_{ij}^r\}_{i,j=1}^c$ denotes the spherical distance between raw electrode positions and source signal positions; $\{d_{ij}^a\}_{i,j=1}^c$ denotes the distance between augmented electrode positions and source signal positions.

3) Step 3: RBF Modeling: The intensity attenuation of the source signal is expressed by a radial basis function (RBF) [52], [53] capable of fitting arbitrary surfaces. The measurement signal of a single electrode is then obtained as the weighted sum of source signals, with the weight of the source signal computed by RBF taking distance as input, as shown in (8) and (9).

The measurement signal from a single electrode is obtained as the weighted sum of source signals, with RBF computing the weight of each source signal based on distance, as shown in (8) and (9).

$$m_i^r = \sum_j^c s_j \phi_j(d_{ij}^r) \tag{8}$$

$$m_i^a = \sum_j^c s_j \phi_j(d_{ij}^a) \tag{9}$$

where ϕ_j denotes the radial basis function with the Gaussian kernel of signal source j; m_i^r and m_i^a denote the raw and augmented source signal of channel i respectively.

The weights of all source signals are written in the form of matrices through equation vectorization of (6) and (7) for clear description and fast computation, as shown in Fig. 4(c). The mathematical expressions are shown in (10) and (11):

$$\begin{aligned}\Phi^r &= \begin{bmatrix} \phi_1(d_{11}^r) & \phi_2(d_{12}^r) & \cdots & \phi_c(d_{1c}^r) \\ \phi_1(d_{21}^r) & \phi_2(d_{22}^r) & \cdots & \phi_c(d_{2c}^r) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(d_{c1}^r) & \phi_2(d_{c2}^r) & \cdots & \phi_c(d_{cc}^r) \end{bmatrix} \\ &= F_\phi(\cos^{-1}(\sin(\Theta_e^r)\sin(\Theta_s^r)^T \\ &\quad + \cos(\Theta_e^r)\cos(\Theta_s^r)^T \odot \cos(\Lambda_e^r \ominus (\Lambda_s^r)^T))) \end{aligned} \tag{10}$$

$$\begin{aligned}\Phi^a &= \begin{bmatrix} \phi_1(d_{11}^a) & \phi_2(d_{12}^a) & \cdots & \phi_c(d_{1c}^a) \\ \phi_1(d_{21}^a) & \phi_2(d_{22}^a) & \cdots & \phi_c(d_{2c}^a) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(d_{c1}^a) & \phi_2(d_{c2}^a) & \cdots & \phi_c(d_{cc}^a) \end{bmatrix} \\ &= F_\phi(\cos^{-1}(\sin(\Theta_e^a)\sin(\Theta_s^a)^T \\ &\quad + \cos(\Theta_e^a)\cos(\Theta_s^a)^T \odot \cos(\Lambda_e^a \ominus (\Lambda_s^a)^T))) \end{aligned} \tag{11}$$

where the $F_\phi(D)$ denotes that an RBF kernel represented by ϕ_i is applied to the ith column of the square matrix D, written as $F_\phi(D) = [\phi_1(D_1), \phi_2(D_2), \cdots, \phi_c(D_c)]$; $\odot$ denotes the Hadamard product; $A \ominus B^T$ denotes the broadcasting subtraction of matrices $A = [a_1, a_2, \cdots, a_c]^T$ and $B = [b_1, b_2, \cdots, b_c]^T$, as shown below:

$$A \ominus B^T = \begin{bmatrix} a_1 - b_1 & a_1 - b_2 & \cdots & a_1 - b_i \\ a_2 - b_1 & a_2 - b_2 & \cdots & a_2 - b_i \\ \vdots & \vdots & \ddots & \vdots \\ a_i - b_1 & a_i - b_2 & \cdots & a_i - b_i \end{bmatrix}$$

4) Step 4: Data Transformation: As shown in Fig. 4(d), the raw measurement signal M^r and the augmented measurement signal M^a can be computed by (12) and (13), respectively.

$$M^r = \Phi^r S \tag{12}$$

$$M^a = \Phi^a S \tag{13}$$

Based on (12) and (13), the raw measurement signal can be transformed to the augmented measurement signal, with transformation matrix $\Phi^a(\Phi^r)^{-1}$:

$$M^a = \Phi^a(\Phi^r)^{-1} M^r \tag{14}$$

Therefore, the algorithm has successfully computed the transformation matrix by leveraging the variations and following the prescribed steps 1-4, with the principle of data augmentation summarized below:

$$\begin{aligned}\widetilde{x} &= \mathcal{T}_x(x) \\ &= \mathcal{M}(\mathcal{V}_e, \mathcal{V}_s)x \\ &= \mathcal{M}(\{\Theta_e^r, \Lambda_e^r\}, \{\Theta_e^a, \Lambda_e^a\}, \{\Theta_s^r, \Lambda_s^r\}, \{\Theta_s^a, \Lambda_s^a\})x \\ &= \Phi^a(\Phi^r)^{-1}x \end{aligned}$$

where $\mathcal{V}_e$ and $\mathcal{V}_s$ represent the variations of electrode placement and brain spatial pattern respectively, and $\mathcal{M}$ represents the transformation model. Notably, $\mathcal{V}_e$ and $\mathcal{V}_s$ are implicitly expressed by the equations associated with positions in the mathematical model. The pseudocode of the proposed SVG is shown in Algorithm 1.

Algorithm 1 SVG for MI Data Augmentation

Input:
raw electrode positions $\{\Theta_e^r, \Lambda_e^r\}$,
raw source signal positions $\{\Theta_s^r, \Lambda_s^r\}$,
probability distribution of the selection of modification rules $\{t_i\}_{i=1}^n$,
raw measurement signal (sampled data) M^r
Output:
augmented measurement signal (augmented data) M^a

```
compute the distances {d_ij^r}_{i,j=1}^c between the raw electrode positions and source signal positions by (6);
compute the weight matrix of raw measurement signal by (10);
for each perturbation i do
    v is a random variable that follows continuous uniform distribution from 0 to 1;
    if v > t_i then
        compute augmented electrode positions {Θ_e^a, Λ_e^a} and source signal positions {Θ_s^a, Λ_s^a} by the certain modification rules: (2), (3), (4), (5);
        compute the distances {d_ij^a}_{i,j=1}^c between the augmented electrode positions and the source signal positions by (7);
        compute weights of augmented measurement signal Φ^a by (11);
        transform the data by (14);
    end
end
return augmented data M^a
```

III. Experiment and Result Analysis

This section details the dataset and model description, experimental setting, and result analysis. This research has been approved by the University Ethics Committee of Xi'an Jiaotong-Liverpool University with proposal number EXT20-01-07 on March 31, 2020.

A. Dataset and Model Description

As overfitting is influenced by both models and data, experiments should be conducted on multiple datasets and models of varying scales to ensure a comprehensive verification of the proposed algorithm's effectiveness. It is worth noting that the labels of augmented data after flipping may need to be adjusted if MI data is lateralized (e.g., the labels of the left and right fists should be exchanged after flipping). Therefore, experimental datasets comprising both lateralized and non-lateralized MI data are necessary for conclusive results. The following section describes the models and datasets separately.

The utilized datasets comprise of the following: PhysioNet motor movement (P-MM), PhysioNet motor imagery (P-MI),

TABLE II
DESCRIPTION OF DATASETS USED IN THE STUDY

Dataset	Subject	Channel	Trial	Sampling rate (Hz)	Class
P-MI-RL [54]	109	64	4748	160	right/left fist
P-MI-FF [54]	109	64	4531	160	both fists/feet
P-MM-RL [54]	109	64	4748	160	right/left fist
P-MM-FF [54]	109	64	4513	160	both fists/feet
M-E1&2 [19]	30	62	10045	1000	right/left cursor
M-E3 [19]	8	64	4923	1024	right/left cursor
BCI2a-RL [55]	9	22	1296	256	right/left hand
BCI2a-FT [55]	9	22	1296	256	both feet/tongue

TABLE III
DESCRIPTION OF MODELS USED IN THE STUDY

Model	Number of parameters	Flops	Year
ShallowConvNet [5]	54402	38015280	2017
DeepConvNet [5]	283927	17934300	2017
EEGNet [4]	41482	9088304	2018
Inception-EEG [10]	27840	19004546	2020
EEGSym [3]	222820	967225656	2022

Meng2019 experiments 1/2 (M-E1&2), Meng2019 experiment 3 (M-E3) (with differences between M-E1&2 and M-E3 in the devices and subjects involved), and BCI Competition IV 2a (BCI2a). A summarized version of the datasets' details is available in Table II. To homogenize the datasets, the data are downsampled to 160Hz and segmented using a 3-second window. The segmented data are processed by eliminating baseline variances and normalizing all channels separately. A total of five models with varying scales are utilized in this study, including ShallowConvNet, DeepConvNet, EEGNet, Inception-EEG, and EEGSym. Table III provides the number of parameters, Flops (floating-point operations per second), and publish year of each aforementioned model.

B. Experimental Setting

In order to achieve a comprehensive evaluation of the proposed algorithm, two experiments are conducted: a comparison study and an ablation study. These studies are implemented on a server consisting of four NVIDIA RTX 2080Ti GPUs, an Intel (R) Xeon (R) Gold 5218 CPU, and implementing CUDA 11.0, cuDNN 8.1.0, and PyTorch 1.7.

1) Comparison Study: A comparison study is designed to evaluate the performance of the proposed SVG algorithm against six previous data augmentation algorithms: Hemisphere Perturbation, Patch Perturbation, and Random Shift, Mixup, Frequency Shift, FT surrogate [3], [21], [26]. The following sections detail each of these techniques:

- Hemisphere Perturbation randomly alters channel positions or replaces data with standard Gaussian noise, assuming that differences between classes can be distinguished regardless of hemisphere [3];
- Patch Perturbation randomly replaces several channels with either zeros or standard Gaussian noise, based on the hypothesis that the difference between classes can be distinguished from partially occluded objects [3], [56];
- Random Shift technique shifts the onset of the trial forward by up to half a second and replaces blank section with either zeros or Gaussian noise, based on the hypothesis that a robust classification model should be able to account for varied reaction times of users [3], [5].
- Mixup applies linear interpolations between two randomly selected pairs of data and labels, based on the hypothesis that the linear interpolations can broaden the support of the training distribution, eventually increasing model robustness and generalizability [21], [26].
- Frequency Shift technique alters the frequencies of signals by an offset sampled uniformly from a range, based on the hypothesis that the frequency peaks across individuals exhibit slight shifts [26].
- FT surrogate replaces the Fourier coefficients with random numbers, based on the hypothesis that EEG signals comprise stationary linear random processes, thereby ensuring distribution consistency between generated data and original data [26].

TABLE IV
THE HYPERPARAMETERS OF THE SVG USED IN THE STUDY

Modification Rules	Probability	Σ
Flipping	0.5	Nil
Scaling	0.3	0.05
Rotation	0.3	0.314
Distortion	0.3	0.05
None	0.1	Nil

2) Ablation Study: In order to verify the effectiveness of the proposed algorithm, an ablation study is conducted in which each modification rule is implemented separately. To increase data diversity, the probabilities of scaling, rotation, and distortion are set to 1, and the probability of flipping is set to 0.5. Other experimental settings (optimizer, datasets, and models) are consistent with the comparison study.

3) Control Group Study: To assess the potential impact of the SVG algorithm on MI research, a control group study is conducted taking into account comprehensive control variables. Specifically, the investigation focuses on the classification accuracy of the model under different sample sizes, compared with the conventional processes with training enhanced by the SVG algorithm. The P-MI-RL dataset is chosen for this study, while the experimental settings, including optimizer and models, remain consistent with those of the comparison and ablation study.

4) Training Schedule: To train the model in both the comparison and ablation studies, the AdamW optimizer is employed with the learning rate initialized at 0.001 and decayed by a multiplicative factor of 0.5 every 50 epochs. Additionally, the batch size and max epoch are set to 64 and 200, respectively. The hyperparameters of the proposed algorithm used in this study are presented in Table IV, where the second and third columns represent the probability distribution of selection and variance of modification rules, respectively. The values are chosen from a range recommended by experimental results on head circumference.

C. Results and Discussion

Due to the inconsistency of MI classification, performance improvement that results from data augmentation may not be

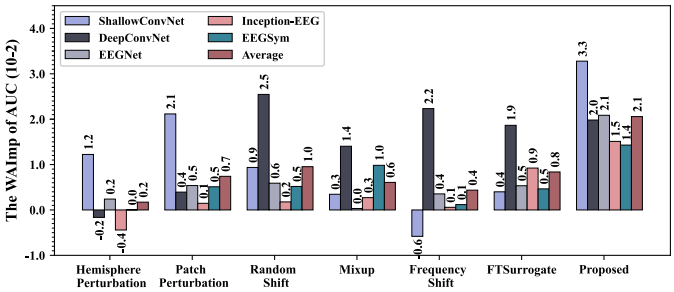

Fig. 5. The weighted average improvement of AUC on five models with different data augmentation algorithms.

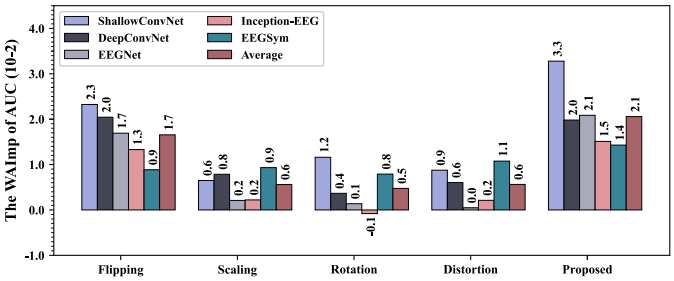

Fig. 6. The weighted average improvement of AUC of the ablation study.

apparent. To obtain stable and reliable results, the following criteria are adopted: 1) utilizing the validation set results with high measurement consistency; 2) evaluating models comprehensively across different classification thresholds based on the area under the receiver operating characteristic curve (AUC); 3) conducting cross-subject classification studies using a high volume of data; 4) performing cross-validation leveraging the entire dataset to compute the results.

1) Comparison Study: The results of the comparison study are presented in Table V, where the AUCs are presented in 'mean ± standard variance' format, and boldface and underlined texts indicate the best and second-best results for each model and dataset, respectively. The first column presents: 1) time (train), which represents the total time consumed during training relative to the process without data augmentation, averaged across five models and eight datasets; 2) time (method), which indicates the time consumed specifically by each data augmentation algorithm, averaged over 1,000 iterations. According to Table V, the proposed algorithm generally outperforms other algorithms, with statistical significance between the AUC of SVG and others determined using the t-test (P-value$<$0.05).

The relationship between data augmentation algorithms and network compatibility can be seen in Fig. 5:

- The magnitude of improvement in AUC gradually diminishes as deep learning-based models progress. For example, ShallowConvNet or DeepConvNet exhibit considerable variability across different data augmentation methods. The results from the seven algorithms applied to five networks indicate that earlier networks achieved greater improvement, possibly due to suitable parameters, structures, and regularization methods compensating for the data augmentation. This highlights the maturing trend of deep learning network architectures.
- Different deep learning networks exhibit distinct compatibility with various types of perturbations. For instance, the ShallowConvNet demonstrates significant improvement after experiencing spatial perturbations (such as Hemisphere Perturbation, Patch Perturbation, and SVG), while DeepConvNet displays substantial improvement after temporal-frequency perturbations (such as Random Shift, Frequency Shift, and FTSurrogate). These insights can guide the optimization of relevant models.
- The networks with limited trainable parameters exhibit minor improvement. Theoretically, an increase in the number of parameters can significantly increase the risk of overfitting, thus bringing a large improvement in the data augmentation method. For instance, the models with few parameters, such as EEGNet and Inception-EEG, do not experience significant improvement from data augmentation. This observation implies that models with large parameters have a high potential for performance improvement via data augmentation.

2) Ablation Study: Another grouped bar chart illustrates the WAImp of AUC of the ablation study, as shown in Fig. 6. The first bar cluster indicates significant improvement by applying the flipping rule (WAImp of AUC increased by approximately 0.017) while scaling, rotation, and distortion rules provide smaller gains in improvements of about 0.006, 0.005, and 0.006 respectively. Notably, the efficacy of these rules varies considerably with models; for instance, the EEGNet and Inception-EEG are found to be less impacted by rotation, scaling, and distortion compared with other models, indicating they may not be sufficiently robust to these modification rules. The results of the ablation study confirm that the use of appropriate data augmentation methods can lead to improved model generalization, underlining the importance of considering multiple techniques when evaluating the impact on a given model.

3) Control Group Study: The results of the control group study are presented in Fig. 8, where the x-axis stands for the number of samples on a logarithmic scale and the y-axis stands for the AUC. The following observations can be made from Fig. 8:

- Firstly, there is an exponential relationship between the sample size and the AUC when the sample quantity is minimal, particularly less than 32. In such circumstances, the application of the SVG algorithm can significantly improve the performance of the model, resulting in an AUC as high as 0.15.
- Secondly, there is a logarithmic relationship between the sample size and the AUC when the sample quantity is considerable, particularly more than 64. In such circumstances, the application of the SVG algorithm enables classification to achieve nearly equivalent performance while utilizing only a quarter of the original sample size.
- Lastly, regardless of the sample quantity, the training process that incorporates the SVG algorithm (represented by the red line) consistently outperforms the conventional process (represented by the blue line).

TABLE V
THE RESULTS OF COMPARISON STUDY

Methods	Models	P-MM-RL	P-MM-FF	P-MI-RL	P-MI-FF	Me-E1&2	Me-E3	B-FT	B-RL
None	**ShallowConvNet**	0.83±0.04	0.75±0.03	0.83±0.03	0.74±0.02	0.92±0.03	0.72±0.07	0.78±0.04	0.76±0.07
time(train): 100%	**DeepConvNet**	0.82±0.03	0.75±0.02	0.81±0.03	0.70±0.04	0.92±0.03	0.79±0.07	0.76±0.05	0.52±0.03
time(method): N.A.	**EEGNet**	0.83±0.03	0.77±0.02	0.83±0.02	0.76±0.01	0.93±0.03	0.80±0.06	0.79±0.03	0.70±0.02
	Inception-EEG	0.84±0.03	0.77±0.02	0.83±0.03	0.76±0.04	0.93±0.03	0.80±0.06	0.80±0.04	0.71±0.07
	EEGSym	0.84±0.03	0.78±0.02	0.84±0.03	0.76±0.02	0.94±0.03	0.80±0.07	0.80±0.04	0.78±0.09
Hemisphere Perturbation	**ShallowConvNet**	0.82±0.03	0.76±0.03	0.84±0.03	0.75±0.03	0.93±0.03	0.78±0.05	0.77±0.04	0.78±0.06
time(train): 101.02%	**DeepConvNet**	0.79±0.04	0.72±0.04	0.81±0.03	0.72±0.03	0.92±0.03	0.79±0.06	0.76±0.05	0.61±0.05
time(method): 0.5384 ms	**EEGNet**	0.82±0.03	0.76±0.03	0.83±0.02	0.77±0.03	0.94±0.03	0.82±0.06	0.79±0.03	0.70±0.04
	Inception-EEG	0.84±0.03	0.75±0.03	0.83±0.03	0.76±0.04	0.93±0.03	0.81±0.06	0.79±0.03	0.63±0.06
	EEGSym	0.84±0.03	0.76±0.02	0.85±0.02	0.76±0.03	**0.95±0.02**	0.81±0.05	0.78±0.05	0.73±0.05
Patch Perturbation	**ShallowConvNet**	0.85±0.03	0.79±0.03	0.86±0.02	0.77±0.02	0.93±0.03	0.77±0.07	0.76±0.04	0.75±0.06
time(train): 105.82%	**DeepConvNet**	0.81±0.04	0.73±0.03	0.82±0.02	0.71±0.03	0.93±0.03	0.79±0.06	0.74±0.05	0.61±0.04
time(method): 2.0417 ms	**EEGNet**	0.84±0.03	0.78±0.02	**0.85±0.02**	0.77±0.03	0.93±0.03	0.81±0.05	0.80±0.03	0.63±0.04
	Inception-EEG	0.85±0.02	0.77±0.02	0.83±0.03	0.76±0.04	0.94±0.03	0.81±0.06	0.80±0.03	0.64±0.06
	EEGSym	0.85±0.03	0.78±0.02	0.85±0.03	0.77±0.03	0.95±0.03	**0.82±0.05**	0.78±0.04	0.71±0.06
Random Shift	**ShallowConvNet**	0.84±0.03	0.78±0.03	0.85±0.03	0.76±0.03	0.92±0.03	0.72±0.07	0.78±0.04	0.80±0.05
time(train): 100.18%	**DeepConvNet**	0.85±0.04	**0.78±0.03**	**0.85±0.03**	0.75±0.04	0.92±0.04	0.79±0.07	0.79±0.04	**0.73±0.07**
time(method): 0.1348 ms	**EEGNet**	0.83±0.03	0.78±0.03	0.84±0.02	**0.77±0.02**	0.93±0.03	0.82±0.06	**0.81±0.04**	0.71±0.04
	Inception-EEG	0.85±0.03	0.77±0.02	0.84±0.03	0.76±0.04	0.92±0.03	0.80±0.06	0.81±0.04	0.72±0.05
	EEGSym	0.86±0.04	0.78±0.01	**0.86±0.03**	0.76±0.03	0.94±0.03	0.80±0.07	0.81±0.04	0.80±0.05
Mixup	**ShallowConvNet**	0.82±0.03	0.75±0.03	0.83±0.02	0.75±0.02	0.92±0.04	0.74±0.06	0.78±0.03	0.76±0.05
time(train): 102.67%	**DeepConvNet**	0.82±0.03	0.75±0.02	0.82±0.02	0.74±0.04	0.92±0.04	0.78±0.07	**0.81±0.04**	0.72±0.07
time(method): 0.2306 ms	**EEGNet**	0.84±0.01	0.77±0.03	0.83±0.02	0.77±0.03	0.93±0.03	0.80±0.07	0.80±0.04	0.67±0.05
	Inception-EEG	0.84±0.03	0.78±0.02	0.83±0.02	0.76±0.02	0.94±0.03	0.80±0.07	0.81±0.04	0.68±0.04
	EEGSym	0.86±0.02	**0.79±0.03**	0.85±0.02	0.77±0.03	0.95±0.03	0.80±0.08	**0.83±0.05**	0.81±0.04
Frequency Shift	**ShallowConvNet**	0.82±0.04	0.74±0.02	0.83±0.03	0.73±0.03	0.92±0.04	0.71±0.07	0.78±0.05	0.78±0.08
time(train): 112.92%	**DeepConvNet**	0.84±0.04	0.76±0.03	0.84±0.03	0.74±0.04	0.93±0.03	0.79±0.08	0.79±0.04	0.72±0.06
time(method): 23.6901 ms	**EEGNet**	0.84±0.03	0.78±0.02	0.83±0.03	0.77±0.02	0.93±0.03	0.80±0.06	0.80±0.03	0.73±0.05
	Inception-EEG	0.84±0.03	0.78±0.03	0.83±0.03	0.76±0.03	0.93±0.03	0.80±0.05	0.80±0.05	0.71±0.05
	EEGSym	0.86±0.03	0.78±0.01	0.85±0.03	0.76±0.04	0.94±0.03	0.79±0.06	0.81±0.03	0.81±0.03
FT Surrogates	**ShallowConvNet**	0.85±0.03	**0.80±0.03**	0.84±0.02	0.76±0.03	0.90±0.04	0.71±0.05	0.73±0.03	0.81±0.05
time(train): 102.56%	**DeepConvNet**	**0.85±0.04**	0.77±0.03	0.84±0.02	**0.75±0.02**	0.92±0.04	0.78±0.07	0.76±0.02	0.68±0.08
time(method): 22.6991 ms	**EEGNet**	0.85±0.03	**0.78±0.01**	0.82±0.02	0.77±0.02	0.93±0.04	0.80±0.04	0.78±0.02	0.78±0.03
	Inception-EEG	**0.87±0.03**	**0.79±0.03**	0.84±0.02	**0.78±0.04**	0.93±0.04	0.80±0.06	0.77±0.02	**0.79±0.03**
	EEGSym	0.86±0.03	0.79±0.03	0.84±0.03	0.76±0.03	0.94±0.03	0.80±0.06	0.79±0.04	**0.82±0.04**
Proposed	**ShallowConvNet**	**0.86±0.03**	0.79±0.02	**0.87±0.02**	**0.77±0.03**	**0.93±0.03**	**0.79±0.06**	**0.80±0.04**	**0.82±0.05**
time(train): 103.83%	**DeepConvNet**	0.84±0.04	0.75±0.03	0.83±0.04	0.72±0.04	**0.93±0.03**	**0.81±0.06**	0.78±0.04	0.72±0.05
time(method): 11.6442 ms	**EEGNet**	**0.86±0.03**	0.78±0.03	0.85±0.03	0.77±0.03	**0.95±0.03**	**0.83±0.05**	0.81±0.03	**0.79±0.06**
	Inception-EEG	0.86±0.03	0.77±0.03	**0.85±0.03**	0.76±0.03	**0.94±0.03**	**0.82±0.05**	**0.82±0.04**	0.78±0.05
	EEGSym	**0.87±0.03**	0.79±0.03	0.85±0.03	**0.78±0.03**	0.95±0.02	0.82±0.06	0.81±0.03	0.81±0.04

4) Feature Distribution Visualization: A feature distribution visualization study is conducted to illustrate the characteristics of the proposed algorithm by analyzing feature distributions. With a particular focus on understanding how data augmentations impact model overfitting, the features in the hidden layer of a pre-trained overfitting model are visualized using the t-SNE algorithm [57], [58]. Specifically, the following steps are performed in turn: 1) training a slightly overfitting model using the P-MI-RL dataset; 2) sampling 100 raw samples uniformly from the training set; 3) generating 2000 augmented samples using different data augmentation methods; 4) extracting features by feeding the raw and augmented samples to the overfitting model; 5) visualizing all the features (14100) by the same transformation based on t-SNE algorithm. The visualization results of the feature distribution are presented in Fig. 7.

As illustrated in Fig. 7, all features are projected onto two dimensions, with the x-axis indicating the direction of the most significant model decision changes (hereafter called x-direction). For better illustration, a probability distribution of the features in the x-direction is also provided. Both the two-dimensional data distribution diagram and the x-direction probability distribution graph show that the seven data augmentation algorithms significantly modify the original data distribution, resulting in a more even distribution. These differences can be interpreted from two perspectives: differences between raw samples and augmented samples, as well as differences among the augmented samples:

- Differences between raw and augmented samples: It can be inferred that all data augmentation techniques (Fig. 7(b)-(h)) make the feature distribution dense and uniform. The distribution of augmented samples suggests an increase in the density of the sample vicinity due to the minor perturbation applied by each method, ensuring the samples remain distinguishable. Concurrently, these augmentation techniques modify the shape of the data distribution in the x-direction to varying degrees. The probability density of the raw data in the x-direction is truncated at the intersection, as shown in Fig. 7(a). However, such truncation does not occur in the augmented data, suggesting a reduction in overfitting.
- Differences across augmentation algorithms: Each data augmentation algorithm produces varying effects on the shape of the data distribution in the x-direction. Notably,

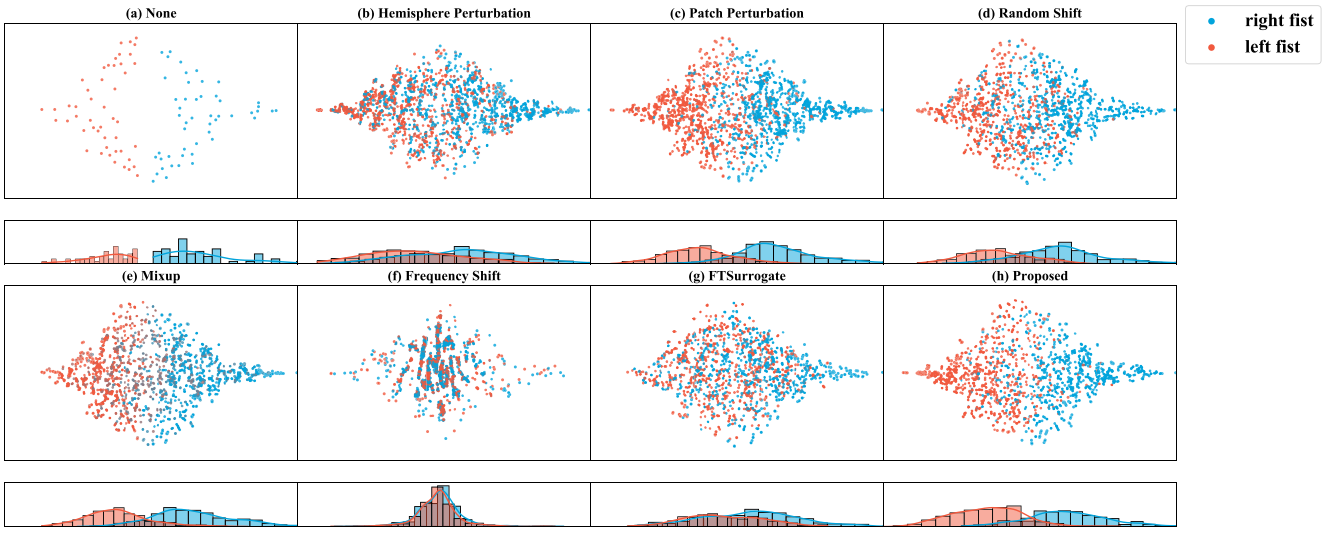

Fig. 7. Feature distribution visualization results by different data augmentation techniques.

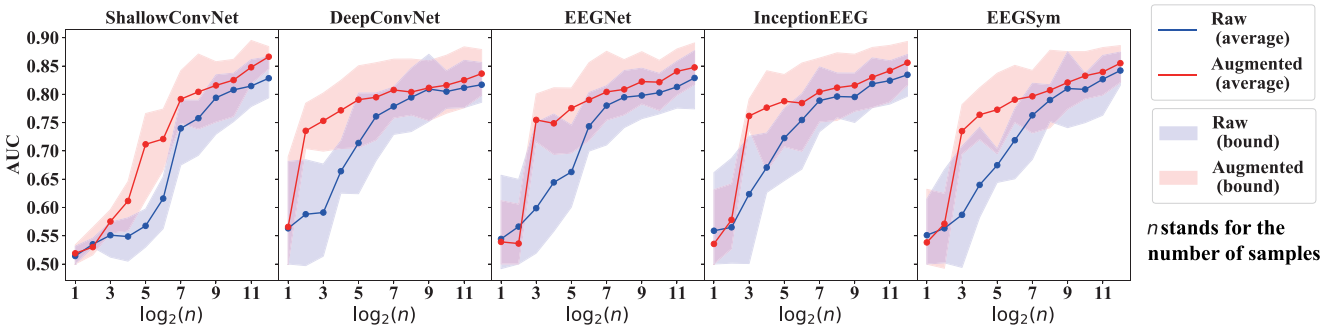

Fig. 8. The result of control group study in five deep learning-based models.

Frequency Shift, FTSurrogate, and Hemisphere Perturbation result in greater intersections between classes; in contrast, Random Shift, Patch Perturbation, SVG, and Mixup yield less intersection. The comparison results demonstrate that the more powerful the reshaping ability of the distribution shape, the more unstable the effect of data augmentation becomes. For instance, Frequency Shift and Hemisphere Perturbation have adverse effects on some models. Conversely, Patch Perturbation, SVG, and Mixup consistently yield positive effects, suggesting the safety and efficacy of these data augmentations.

IV. Conclusion

In recent years, considerable efforts have been devoted to addressing the issue of data scarcity in motor imagery. This paper proposes a novel data augmentation algorithm, called Spatial Variation Generation, to generate augmented MI data based on variations in electrode placement and brain spatial pattern. Notably, the proposed SVG performs four modification rules (including flipping, scaling, rotation, and distortion) based on different variations, establishing a bridge between variations and MI data transformation. The bridge is constructed by the equations related to positions in the mathematical model, implicitly expressing how the variation holds an influence on MI data. To evaluate the performance of the proposed SVG algorithm, the comparison study and ablation study of each modification rule is conducted on eight datasets by five deep learning-based models. Experimental results reveal that the proposed SVG outperforms the other data augmentation techniques, with the ablation study demonstrating the effectiveness of each modification rule in performance enhancement. Despite the effectiveness of the current SVG algorithm in enhancing MI classification performance, there exist opportunities for further improvement in representation, efficiency, and feasibility. Future research endeavors could include the comprehensive exploration of additional variabilities in EEG for representative data. Furthermore, the development of fast and efficient transformation models is needed to enhance the speed of the transformation process. Finally, it is crucial to verify the feasibility of applying these variabilities to other EEG experimental paradigms to ensure the generalizability of the findings.

References

[1] T. Aflalo et al., "Decoding motor imagery from the posterior parietal cortex of a tetraplegic human," *Science*, vol. 348, no. 6237, pp. 906–910, May 2015.

[2] J. Ma, B. Yang, W. Qiu, Y. Li, S. Gao, and X. Xia, "A large EEG dataset for studying cross-session variability in motor imagery brain–computer interface," *Sci. Data*, vol. 9, no. 1, p. 531, Sep. 2022.

[3] S. Pérez-Velasco, E. Santamaría-Vázquez, V. Martínez-Cagigal, D. Marcos-Martínez, and R. Hornero, "EEGSym: Overcoming inter-subject variability in motor imagery based BCIs with deep learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1766–1775, 2022.

[4] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.

[5] R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.

[6] Y. Chen, R. Yang, M. Huang, Z. Wang, and X. Liu, "Single-source to single-target cross-subject motor imagery classification based on multisubdomain adaptation network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1992–2002, 2022.

[7] N. Yu, R. Yang, and M. Huang, "Deep common spatial pattern based motor imagery classification with improved objective function," *Int. J. Netw. Dyn. Intell.*, vol. 1, pp. 73–84, Dec. 2022.

[8] A. Dong, A. Starr, and Y. Zhao, "Neural network-based parametric system identification: A review," *Int. J. Syst. Sci.*, vol. 54, no. 13, pp. 2676–2688, Aug. 2023.

[9] H. Alshaya and M. Hussain, "EEG-based classification of epileptic seizure types using deep network model," *Mathematics*, vol. 11, no. 10, p. 2286, May 2023.

[10] E. Santamaría-Vázquez, V. Martínez-Cagigal, F. Vaquerizo-Villar, and R. Hornero, "EEG-inception: A novel deep convolutional neural network for assistive ERP-based brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2773–2782, Dec. 2020.

[11] X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, and L. Sun, "A multi-branch 3D convolutional neural network for EEG-based motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2164–2177, Oct. 2019.

[12] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[13] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," in *Proc. 6th Int. Conf. Learn. Represent. (ICLR)*, 2017, pp. 1–14.

[14] G. Zhao, Y. Li, and Q. Xu, "From emotion AI to cognitive AI," *Int. J. Netw. Dyn. Intell.*, vol. 1, no. 1, pp. 5–72, 2022.

[15] D. Ji, C. Wang, J. Li, and H. Dong, "A review: Data driven-based fault diagnosis and RUL prediction of petroleum machinery and equipment," *Syst. Sci. Control Eng.*, vol. 9, no. 1, pp. 724–747, 2021.

[16] C. Wang, Z. Wang, W. Liu, Y. Shen, and H. Dong, "A novel deep offline-to-online transfer learning framework for pipeline leakage detection with small samples," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, 2023.

[17] M. Nakanishi, Y.-T. Wang, C.-S. Wei, K.-J. Chiang, and T.-P. Jung, "Facilitating calibration in high-speed BCI spellers via leveraging cross-device shared latent responses," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 4, pp. 1105–1113, Apr. 2020.

[18] D. Wu, Y. Xu, and B.-L. Lu, "Transfer learning for EEG-based brain–computer interfaces: A review of progress made since 2016," *IEEE Trans. Cognit. Develop. Syst.*, vol. 14, no. 1, pp. 4–19, Mar. 2022.

[19] J. Meng and B. He, "Exploring training effect in 42 human subjects using a non-invasive sensorimotor rhythm based online BCI," *Frontiers Human Neurosci.*, vol. 13, p. 128, Apr. 2019.

[20] D. Wu, V. J. Lawhern, W. D. Hairston, and B. J. Lance, "Switching EEG headsets made easy: Reducing offline calibration effort using active weighted adaptation regularization," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 11, pp. 1125–1137, Nov. 2016.

[21] H. Zhang, M. Cissé, Y. N. Dauphin, and D. Lopez-Paz, "*mixup*: Beyond empirical risk minimization," in *Proc. 6th Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–13.

[22] O. Chapelle, J. Weston, L. Bottou, and V. Vapnik, "Vicinal risk minimization," in *Advances in Neural Information Processing Systems*, vol. 13, T. K. Leen, T. G. Dietterich, and V. Tresp, Eds. Denver, CO, USA: MIT Press, 2000, pp. 416–422.

[23] P. Lu, B. Song, and L. Xu, "Human face recognition based on convolutional neural network and augmented dataset," *Syst. Sci. Control Eng.*, vol. 9, no. sup2, pp. 29–37, May 2021.

[24] P. Durongbhan et al., "A dementia classification framework using frequency and time-frequency features based on EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 826–835, May 2019.

[25] L. Yang, Y. Song, K. Ma, and L. Xie, "Motor imagery EEG decoding method based on a discriminative feature learning strategy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 368–379, 2021.

[26] C. Rommel, T. Moreau, J. Paillard, and A. Gramfort, "CADDA: Class-wise automatic differentiable data augmentation for EEG signals," in *Proc. 10th Int. Conf. Learn. Represent. (ICLR)*, 2022, pp. 1–24.

[27] M. M. Krell and S. K. Kim, "Rotational data augmentation for electroencephalographic data," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 471–474.

[28] Y. Li, X.-R. Zhang, B. Zhang, M.-Y. Lei, W.-G. Cui, and Y.-Z. Guo, "A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1170–1180, Jun. 2019.

[29] B. Li et al., "Component-mixing strategy: A decomposition-based data augmentation algorithm for motor imagery signals," *Neurocomputing*, vol. 465, pp. 325–335, Nov. 2021.

[30] S. N. A. Seha and D. Hatzinakos, "A new training approach for deep learning in EEG biometrics using triplet loss and EMG-driven additive data augmentation," *Neurocomputing*, vol. 488, pp. 194–211, Jun. 2022.

[31] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clin. Neurophysiol.*, vol. 112, no. 4, pp. 713–719, Apr. 2001.

[32] M. Shahbakhti et al., "SWT-kurtosis based algorithm for elimination of electrical shift and linear trend from EEG signals," *Biomed. Signal Process. Control*, vol. 65, Mar. 2021, Art. no. 102373.

[33] C. L. Scrivener and A. T. Reader, "Variability of EEG electrode positions and their underlying brain regions: Visualizing gel artifacts from a simultaneous EEG-fMRI dataset," *Brain Behav.*, vol. 12, no. 2, p. e2476, 2022.

[34] H. Kim and C.-H. Im, "Influence of the number of channels and classification algorithm on the performance robustness to electrode shift in steady-state visual evoked potential-based brain–computer interfaces," *Frontiers Neuroinform.*, vol. 15, Oct. 2021, Art. no. 750839.

[35] A. Bhaduri et al., "An atlas of cortical arealization identifies dynamic molecular signatures," *Nature*, vol. 598, no. 7879, pp. 200–204, Oct. 2021.

[36] A. Sohrabpour, Z. Cai, S. Ye, B. Brinkmann, G. Worrell, and B. He, "Noninvasive electromagnetic source imaging of spatiotemporally distributed epileptogenic brain sources," *Nature Commun.*, vol. 11, no. 1, p. 1946, 2020.

[37] G. Deco, J. Cruzat, and M. L. Kringelbach, "Brain songs framework used for discovering the relevant timescale of the human brain," *Nature Commun.*, vol. 10, no. 1, p. 583, Feb. 2019.

[38] T. M. Pearce and D. W. Moran, "Strategy-dependent encoding of planned arm movements in the dorsal premotor cortex," *Science*, vol. 337, no. 6097, pp. 984–988, Aug. 2012.

[39] S. Saha, K. I. U. Ahmed, R. Mostafa, L. Hadjileontiadis, and A. Khandoker, "Evidence of variabilities in EEG dynamics during motor imagery-based multiclass brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 371–382, Feb. 2018.

[40] J. Gao et al., "Hemispheric difference of regional brain function exists in patients with acute stroke in different cerebral hemispheres: A resting-state fMRI study," *Frontiers Aging Neurosci.*, vol. 13, Jul. 2021, Art. no. 691518.

[41] L. Zago et al., "Predicting hemispheric dominance for language production in healthy individuals using support vector machine," *Hum. Brain Mapping*, vol. 38, no. 12, pp. 5871–5889, Dec. 2017.

[42] B. J. Poole, M. Mather, E. J. Livesey, I. M. Harris, and J. A. Harris, "Motor-evoked potentials reveal functional differences between dominant and non-dominant motor cortices during response preparation," *Cortex*, vol. 103, pp. 1–12, Jun. 2018.

[43] K. Amunts, H. Mohlberg, S. Bludau, and K. Zilles, "Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture," *Science*, vol. 369, no. 6506, pp. 988–992, Aug. 2020.

[44] T. T. Hshieh et al., "Head circumference as a useful surrogate for intracranial volume in older adults," *Int. Psychogeriatrics*, vol. 28, no. 1, pp. 157–162, Jan. 2016.

[45] S. Treit et al., "Relationships between head circumference, brain volume and cognition in children with prenatal alcohol exposure," *PLoS One*, vol. 11, no. 2, Feb. 2016, Art. no. e0150370.

[46] L. Yong-Ping et al., "Impact of maternal smoking associated lysophosphatidylcholine 20:3 on offspring brain development," *J. Steroid Biochem. Mol. Biol.*, vol. 199, May 2020, Art. no. 105591.

[47] T. Song, W. Zheng, S. Liu, Y. Zong, Z. Cui, and Y. Li, "Graph-embedded convolutional neural network for image-based EEG emotion recognition," *IEEE Trans. Emerg. Topics Comput.*, vol. 10, no. 3, pp. 1399–1413, Jul. 2022.

[48] H. J. Baek, H. J. Lee, Y. G. Lim, and K. S. Park, "Conductive polymer foam surface improves the performance of a capacitive EEG electrode," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3422–3431, Dec. 2012.

[49] R. Martínez-Cancino et al., "The open EEGLAB portal interface: High-performance computing with EEGLAB," *NeuroImage*, vol. 224, Jan. 2021, Art. no. 116778.

[50] W. L. Lim, O. Sourina, and L. P. Wang, "STEW: Simultaneous task EEG workload data set," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 11, pp. 2106–2114, Nov. 2018.

[51] H. S. Courellis, J. R. Iversen, H. Poizner, and G. Cauwenberghs, "EEG channel interpolation using ellipsoid geodesic length," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2016, pp. 540–543.

[52] T. Henneron, A. Pierquin, and S. Clénet, "Mesh deformation based on radial basis function interpolation applied to low-frequency electromagnetic problem," *IEEE Trans. Magn.*, vol. 55, no. 6, pp. 1–4, Jun. 2019.

[53] P. Hashemzadeh and A. S. Fokas, "Helmholtz decomposition of the neuronal current for the ellipsoidal head model," *Inverse Problems*, vol. 35, no. 2, Feb. 2019, Art. no. 025002.

[54] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain–computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.

[55] M. Tangermann et al., "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, p. 55, Jan. 2012.

[56] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proc. 34th Conf. Artif. Intell. (AAAI)*, 2020, pp. 13001–13008.

[57] D. Yang, J. Lu, H. Dong, and Z. Hu, "Pipeline signal feature extraction method based on multi-feature entropy fusion and local linear embedding," *Syst. Sci. Control Eng.*, vol. 10, no. 1, pp. 407–416, Dec. 2022.

[58] H. Zhang, C. Zhang, J. Dong, and K. Peng, "A new key performance indicator oriented industrial process monitoring and operating performance assessment method based on improved Hessian locally linear embedding," *Int. J. Syst. Sci.*, vol. 53, no. 16, pp. 3538–3555, 2021.