



Data Augmentation for EEG Motor Imagery Classification Using Diffusion Model

Nutapol Soingern^{1(✉)}, Akraradet Sinsamersuk¹, Itthi Chatnuntaweche²,
and Chaklam Silpasuwanchai¹

¹ School of Engineering and Technology, Data Science and Artificial Intelligence,
Asian Institute of Technology, Pathum Thani, Thailand
nutapol1997@gmail.com, {st121413,chaklam}@ait.asia

² National Science and Technology Development Agency, National Nanotechnology
Center, Pathum Thani, Thailand
itthi.cha@nanotec.or.th

Abstract. Motor imagery classification using electroencephalogram (EEG) signals is an important research topic that has been extensively studied in the field of brain-computer interfaces (BCIs). However, due to the limited amount of available data, overfitting is a common problem, especially when using a deep-learning classifier. One way to address this is by performing data augmentation. In this paper, we investigate the efficacy of the diffusion model as a data augmentation method for motor imagery classification. We evaluated the diffusion method by comparing it with commonly-used EEG data augmentation techniques namely such as Noise Addition, Fourier Transform Surrogates, Frequency Shift, and SmoothTimeMask. The result shows that the diffusion method outperformed other methods in terms of classification accuracy by 17.49%. The Kullback-Leibler (KL) divergence is used for assessing the similarity between the training set (with and without augmentation) and validation set, thus showing the effectiveness of the diffusion approach compared to other techniques.

Keywords: Motor Imagery · EEG · Brain Computer Interface · Deep learning · Data Augmentation · Diffusion · KL divergence

1 Introduction

Brain-computer interfaces (BCI) establish a direct pathway between the human brain and a computer via signal processing and decoding techniques. One classic paradigm of EEG is motor imagery (MI), in which its physiological basis is based on body movements or imagined movements that can produce α (8–13 Hz) and β (13–30 Hz) event-related synchronization (ERS) and event-related desynchronization (ERD) rhythms in the motor-sensory areas of the brain [1]. Recently, deep learning (DL) model has been used for motor imagery classification. For

example, EEGNet [2] is a compact convolutional neural network designed for EEG-based brain-computer interfaces that effectively extracts spatial-temporal features from EEG signals. In any case, the paucity of data is a prevalent issue in the field of EEG classification, as it hinders the development and performance of DL models. Consequently, a common symptom is overfitting, which reduces the model’s accuracy and robustness on test set [3].

Data augmentation (DA) has been widely used to improve the robustness and accuracy of DL by artificially increasing the number of training data. Traditional EEG data augmentation methods include **Noise Addition** [4–6], fourier transform surrogates [7], **Frequency Shifting** [8,9] and **SmoothTimeMask** [10]. **Noise Addition** [5,6] adds random white noise to all channels. Fourier transform surrogates [7] randomizes the Fourier-transform (FT) phases of temporal-spatial data generates surrogates that approximate examples from the data-generating distribution. **Frequency Shifting** [8,9] randomly shifts the frequency spectrum on all channels. Last, **SmoothTimeMask** [10] randomly masks consecutive time steps of the EEG signal and replaces them with zeros. The length of the masked segment was also selected randomly. The motivation is to force the model to disregard minor irrelevant events.

Recently, diffusion model [11] was proposed which generates synthetic data based on Langevin dynamics. These models naturally admitted a progressive lossy decompression scheme that can be interpreted as an extension of autoregressive decoding. Diffusion mode has been used as a DA method to generate synthetic training data for skin disease classification [12], etc. Moreover, **WaveGrad** [13] model is a research-based approach that has been developed to generate audio waveforms of superior quality. The proposed approach involves utilizing score matching [14] and diffusion probabilistic models to estimate gradients of the data density within a conditional model. The research methodology involves initializing the model with a Gaussian white noise signal and subsequently improving the signal quality through an iterative process that utilizes a gradient-based sampler. The sampler is conditioned on the mel-spectrogram. The **WaveGrad** model presents a method to balance inference speed and sample quality through the manipulation of refinement steps. Additionally, it serves as a connection between non-autoregressive and autoregressive models in relation to audio quality. According to the research findings, **WaveGrad** has the ability to produce high-quality audio samples with only six iterations.

In this work, we demonstrated the use of the diffusion model based on **WaveGrad** [13] as a DA method for motor imagery classification. We evaluated the effect of the proposed method by performing DA on BCI Competition IV 2a [15] with various size of synthetic data with five standards EEG MI models (EEGNet [2], ATCNet [16], EEG-ITNet [17], Deep ConvNet [18] and ShallowFBCSPNet [18]). The proposed method improve the performance outperform other traditional EEG data augmentation methods.

2 Related Work

We reviewed commonly-used data augmentation for EEG MI such as **Noise Addition**, **Fourier Transform Surrogates**, **Frequency Shift** and **Smooth TimeMask**.

2.1 Noise Addition

Noise Addition has two main categories for adding noise to the EEG signals in purpose of DA [19]. A common technique in EEG research involved introducing Gaussian noise with zero mean and a standard deviation of 0.1 to the recorded data [5]. The simulation of EEG data variability was commonly utilized to replicate the effects of electrode noise or subject movement during experimental procedures. The introduction of noise to the training data could enhance the robustness of the model by compelling it to learn features that were less susceptible to minor fluctuations in the data. Previous research has demonstrated that the inclusion of Gaussian noise in EEG signals enhances the efficacy of the MI classification model when applied to BCI competition IV dataset 2b [20], resulting in a 10% improvement in performance.

2.2 Fourier Transform Surrogates

The Fourier transform surrogates (**FTSurrogate**) method utilized the phase data of frequency elements, which were subsequently rearranged in a random manner while maintaining their original magnitude spectrum [7]. The generation of synthetic data samples has been utilized as a means to address the underrepresentation of certain classes. This approach has been shown to improve the balance of class distribution and enhance the accuracy of classification. The method proposed in this study has the potential to enhance classification performance either as a standalone technique or in conjunction with other data augmentation methods [7]. The extent of enhancement varies based on the particular dataset and classification issue. The extent of enhancement differs based on the particular dataset and classification issue. The study aimed to enhance the mean F1-score of a convolutional neural network utilized for sleep stage classification by 7% through the implementation of surrogate-based augmentation on the CAPSLPDB sleep database [21].

2.3 Frequency Shift

In the **Frequency Shift** method, the frequency spectrum of an EEG signal was randomly shifted to a different frequency range while maintaining the amplitude spectrum [9]. The proposed technique involved generating novel EEG signals that exhibit identical spectral characteristics as the initial signal, albeit with altered frequencies. The study's findings indicated that the **Frequency Shift** method was successful in enhancing the classification accuracy of certain EEG datasets.

In a study on motor imagery datasets, the implementation of the **Frequency Shift** method resulted in a 2.5% increased in classification accuracy when compared to the baseline method. Moreover, Jaderberg et al. (2021) proposed generating augmented EEG signals by applying various transformations, including the **Frequency Shift** method, to the original signals. The study assessed the efficacy of a novel method on various EEG classification tasks and demonstrated its superiority over conventional data augmentation techniques, including random cropping and flipping, as well as other learned data augmentation methods. The study's findings indicated that the suggested approach yields optimal performance and required less time for training compared to gradient-based methods in the class-agnostic context. Additionally, it surpassed gradient-free methods in the class-wise context. The research paper lacked a specific numerical value for the quantity of effects or enhancements. The effectiveness of this method in enhancing classification performance was observed in certain datasets, such as the BCI Competition IV 2a dataset [8].

2.4 SmoothTimeMask

SmoothTimeMask was a research methodology that utilized time-domain augmentation to introduce smoothness into a signal. This is achieved by masking contiguous time intervals [10]. The **SmoothTimeMask** algorithm was a technique that used to apply a smooth mask to a segment of a time series signal. This method involved generating a mask by randomly selecting a starting point and masking a fixed length of contiguous samples. A common technique that used to create a smooth transition between masked and unmasked regions is the application of a convolution with a Gaussian kernel to the mask. The introduction of smoothness in the augmented signal has the potential to prevent overfitting and enhance the generalization of the model [10].

2.5 WaveGrad

WaveGrad is a generative model for waveform generation that uses score matching and denoising to improve the quality of generated waveforms. The basic idea behind the method is to estimate the probability density function of a dataset using a generative model, and then use this estimate to generate new data points that are similar to the original data.

To achieve this, **WaveGrad** uses an autoregressive architecture that predicts each sample of the waveform conditioned on the previous samples. Specifically, **WaveGrad** uses a modified version of the WaveNet architecture that replaces the dilated convolutions with a set of learned gates and skips connections, which reduces the computational cost of the model.

WaveGrad trains the generative model using score matching, which is a technique that involves matching the gradient of the log-density function of the model to the gradient of the true log-density function of the data. The idea behind score matching is that the gradient of the log-density function is easier

to estimate than the function itself, and that matching the gradients is sufficient to match the distributions.

Score matching is a technique used for estimating the probability density function (PDF) of a dataset by matching the score function of a model to the true score function of the PDF. The score function is the gradient of the log-density function, i.e., the vector of partial derivatives of the log-density function with respect to each input variable.

In score matching, the model is trained to minimize the difference between the score function of the model and the true score function of the PDF. This can be formulated as the following loss function:

$$L(\theta) = \sum_i |\nabla_x \log p(x_i; \theta) - \nabla_x \log \hat{p}(x_i)|^2 \quad (1)$$

where θ are the parameters of the model, x_i is a data point, $p(x_i; \theta)$ is the model's probability density function, and $\hat{p}(x_i)$ is the true probability density function of the dataset.

Denoising score matching is an extension of score matching that uses a denoising autoencoder to estimate the score function of the PDF. The denoising autoencoder is trained to remove noise from the input data, and the score function of the denoised data is used to estimate the true score function of the PDF. The loss function for denoising score matching is:

$$L(\theta) = \sum_i |\nabla_x \log \hat{p}(\tilde{x}_i) - \nabla_x \log \hat{p}(x_i)|^2 \quad (2)$$

where \tilde{x}_i is the denoised version of x_i .

Weighted denoising score matching is a further extension of denoising score matching that accounts for noisy labels. The idea is to assign higher weights to samples with less noise and lower weights to samples with more noise. This can be achieved by introducing a weighting function $w(x_i)$ into the loss function:

$$L(\theta) = \sum_i w(x_i) |\nabla_x \log \hat{p}(\tilde{x}_i) - \nabla_x \log \hat{p}(x_i)|^2 \quad (3)$$

where θ are the parameters of the model, x_i is a data point, $\hat{p}(x_i)$ is the true probability density function of the dataset, \tilde{x}_i is the denoised version of x_i , and $w(x_i)$ is a weighting function that assigns a weight to each sample based on the level of noise in its label.

WaveGrad further improves the optimization of the model by using a variant of stochastic gradient descent called Stochastic Gradient Hamiltonian Monte Carlo (SGHMC). SGHMC uses Hamiltonian dynamics to simulate the motion of particles in a potential energy landscape, which improves the exploration of the parameter space during optimization.

Overall, **WaveGrad** is able to generate high-quality waveforms that are comparable to or better than previous state-of-the-art methods. It achieves this by combining denoising and score matching with a modified version of the WaveNet architecture and SGHMC optimization.

3 Methodology

We compared the diffusion method against four DA methods and the baseline method without augmentation. Figure 1 shows how the training sets were obtained using different combinations of the DA method and sampling size. The sampling size was chosen at the ratio of 25, 50, 75, 100%. Five commonly-used models for MI classification was used. The models were trained using subject-dependent scheme and evaluated on their respective testing sets.

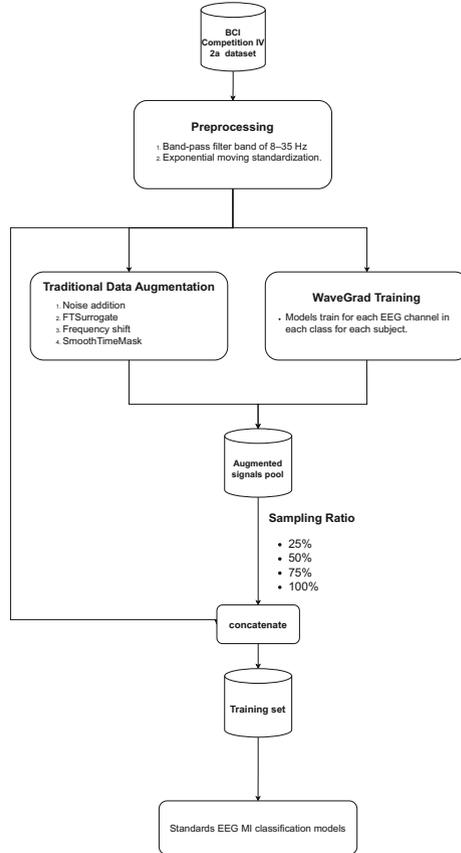


Fig. 1. Showed how to obtain training sets when using traditional DA and WaveGrad. WaveGrad was trained for each EEG channel in each class for each subject. Thus, we trained a total of 792 models from four classes, 22 channels and 9 subjects.

3.1 Datasets

BCI Competition IV 2a [15] was a collection of EEG data from 9 subjects who participated in a cue-based BCI paradigm involving four distinct motor imagery

tasks: imagining movement of the left hand (class 1), the right hand (class 2), both feet (class 3) and the tongue (class 4). Each subject completed the tasks in two distinct sessions on different days, with each session consisting of six runs separated by brief pauses, resulting in a total of 48 trials (12 for each of the four classes). The data were captured while the participants sat in a comfortable armchair in front of a computer screen, and a fixation cross appeared at the start of each trial. A cue consisting of an arrow pointing to the left, right, down, or up was used to prompt the subjects to perform the desired motor imagery task. The subjects were instructed to perform the motor imagery task until the fixation cross disappeared from the screen without receiving any feedback. Signals were sampled at 250 Hz and filtered between 0.5 Hz and 100 Hz using a 50 Hz notch filter in order to reduce line noise. We used one section for training sets and another for test sets.

3.2 Data Preprocessing

EEG signal was filtered with a band pass filter of 8–35 Hz followed by exponential moving standardization x_t . For exponential moving standardization, compute the exponential moving mean m_t at time t and as show in Eqs. (4). Then, compute the exponential moving variance v_t at time t as show in Eqs. (5). We set *factornew* is 0.001 and *esp* is 0.0001. Finally, standardize the data point x_t at time t as show in Eqs. (6).

$$m_t = \text{factornew} \cdot \text{mean}(x_t) + (1 - \text{factornew}) \cdot m_{t-1}, \quad (4)$$

$$v_t = \text{factornew} \cdot (m_t - x_t)^2 + (1 - \text{factornew}) \cdot v_{t-1}, \quad (5)$$

$$x'_t = (x_t - m_t) / \max(\sqrt{-}, v_t, \text{esp}), \quad (6)$$

3.3 Data Augmentation

Our objective was to assess the performance of each DA method with different ratios of augmented/synthetic data. Five DA methods (**Noise Addition**, **FTSurrogate**, **Frequency Shift**, **SmoothTimeMask**, **WaveGrad**) and four ratios (25%, 50%, 75%, 100%) = 20 combinations (5 methods \times 4 ratios) of the training set are to be created. Here, ratios refer to the amount of augmented/synthetic data used. For example, given 25% ratio, a 25% augmented data is randomly sampled and add to the original training set. By comparing different ratios, we can better understand the impact of number of augmentation to accuracy improvement.

Similar to previous works [8, 10, 20, 21], we used the subject-dependent scheme which augments data in subject level (i.e., each subject is treated separately) and channel level (i.e., each channel is treated independently).

Noise Addition: **Noise Addition** entails the inclusion of diverse forms of noise, such as Gaussian, Poisson, and others, that possess varying parameters to

the original EEG signal. The raw EEG signal was subjected to additive noise by incorporating a Gaussian distribution with a standard deviation of 0.1.

Fourier Transform Surrogates: Fourier transform surrogates are a type of data generated by randomizing the phases of temporal-spatial data.

Frequency Shift: The technique of Frequency Shift is characterized by the alteration of the frequency of the EEG signal by a specific amount. We random shift the frequency by ± 2 Hz.

SmoothTimeMask: SmoothTimeMask involves applying a smooth window function to mask a continuous segment of the signal, and optimizing it using gradient-based methods. The signal was randomly masked with a range of 100 sample points.

WaveGrad: It is first important that WaveGrad is originally a generative model, not a formal augmentation technique. Thus, in contrary to other DA methods, WaveGrad has to be trained before it can be used to generate a synthetic EEG signal. The dataset consists of 9 subjects and four MI classes. The EEG recording has 22 channels. Thus, the total number of WaveGrad models was $(9 \text{ subjects} \times 22 \text{ channels} \times 4 \text{ classes}) = 792$ models. The training procedure and parameters were the same across all WaveGrad models. The learning rate was set to 0.0001 and the diffusion steps to 1000.

3.4 Evaluation

First, it is important to evaluate the quality of the augmented/synthetic data. A common way is through dimensionality reduction KL divergence. The success of the diffusion method in comparison to other methods is demonstrated by the Kullback-Leibler (KL) divergence, which is used to measure the similarity between the training set (with and without augmentation) and validation set. We expect that high-quality augmented or synthetic data should exhibit similarity between the training set (with and without augmentation) and the validation set.

Second, once the quality of the augmented/synthetic data are quantified, we are now ready to quantify the usefulness of data augmentation techniques on actual EEG tasks. We first selected five commonly-used motor imagery classification models (EEGNet, ATCNet, EEG-ITNet, Deep ConvNet and ShallowF-BCSPNet) which would allow us to understand whether how complexity of the model relates with data augmentation. Here, note that we simply define the complexity based on the model’s number of parameters. Accuracy was then measured across all 21 combinations (5 DA methods \times 4 sampling ratios + 1 baseline method without augmentation). The details of each model were as follows:

EEGNet. EEGNet is a single CNN architecture that can accurately classify EEG signals from different BCI paradigms while being as compact as possible. The authors introduce the use of depthwise and separable convolutions to construct an EEG-specific model that encapsulates well-known EEG feature extraction concepts for BCI. They compare EEGNet to current state-of-the-art approaches across four BCI paradigms and show that EEGNet generalizes across paradigms better than, and achieves comparably high performance to, the reference algorithms when only limited training data is available across all tested paradigms.

ATCNet. ATCNet is predicting the onset of epileptic seizures using electroencephalogram (EEG) signals. The ATCNet consists of two blocks: an attention-based temporal convolutional (ATC) block and a transformer-based classification (TC) block. The ATC block is used to extract relevant features from the EEG signals, while the TC block is used to classify the extracted features into seizure and non-seizure classes. The proposed model is evaluated using the BCI Competition IV-2a (BCI-2a) dataset. The obtained accuracy ranges from 60.5% to 89.5%.

EEG-ITNet. EEG-ITNet uses inception modules and causal convolutions with dilation to extract rich spectral, spatial, and temporal information from multi-channel EEG signals with less complexity than other existing end-to-end architectures. The paper also provides a methodology for achieving intuitive visualisation structures such as topographic maps. The proposed EEG-ITNet model shows up to 5.9% improvement in classification accuracy compared to its competitors in different scenarios. The paper also explains and supports the validity of network illustration from a neuroscientific perspective.

Deep ConvNet. The deep ConvNet had four convolution-max-pooling blocks, with a special first block designed to handle EEG input, followed by three standard convolution-max-pooling blocks and a dense softmax classification layer. The authors found that recent advances in machine learning, including batch normalization and exponential linear units, together with a cropped training strategy, boosted the Deep ConvNets decoding performance, reaching at least as good performance as the widely used filter bank common spatial patterns (FBCSP) algorithm.

ShallowFBCSPNet. The shallow ConvNet are similar to the transformations of FBCSP. Concretely, the first two layers of the shallow ConvNet perform a temporal convolution and a spatial filter, as in the deep ConvNet. These steps are analogous to the bandpass and CSP spatial filter steps in FBCSP.

4 Results

4.1 Kullback-Leibler Divergence (KL Divergence)

To understand the quality of the generated data, we measured the similarity of signal by KL divergence process. We measure the KL divergence comparing of train set and test set, test set and train set with 25% of data augmentation, test set and train set with 50% of data augmentation, test set and train set with 75% of data augmentation, test set and train set with 100% of data augmentation, test set and data augmentation and train set and data augmentation. We randomize the augmentation data for this process 100 times and averaged the result. Overall, our method increases similarity as the ratios of augmented data are increased. When ratios are increased, the similarity for the SmoothTimeMask, Noise Addition, and FTSurrogate approaches that of the non-augmented data. On the other hand, as the augmented data from frequency shift increases, the similarity declines (Table 1).

Table 1. Average of KL divergence each data augmentation method. This result is based on 100 random iterations of each augmentation data process.

Data Augmentation	Test set vs Train set	Average of 25%	Average of 50%	Average of 75%	Average of 100%	Test set vs Augmentation	Train set vs Augmentation
WaveGrad	2421.27	2372.21	2336.35	2310.92	2294.04	2158.61	2090.67
Noise Addition	2316.79	2318.41	2324.93	2312.74	2316.05	2316.61	4.91
FTSurrogate	2316.79	2311.50	2305.57	2306.39	2306.27	2292.40	2226.75
SmoothTimeMask	2316.79	2320.10	2320.04	2322.51	2310.70	2316.79	0.00
Frequency shift.	2316.79	3513.38	4298.14	4888.52	5295.02	8287.84	8147.79

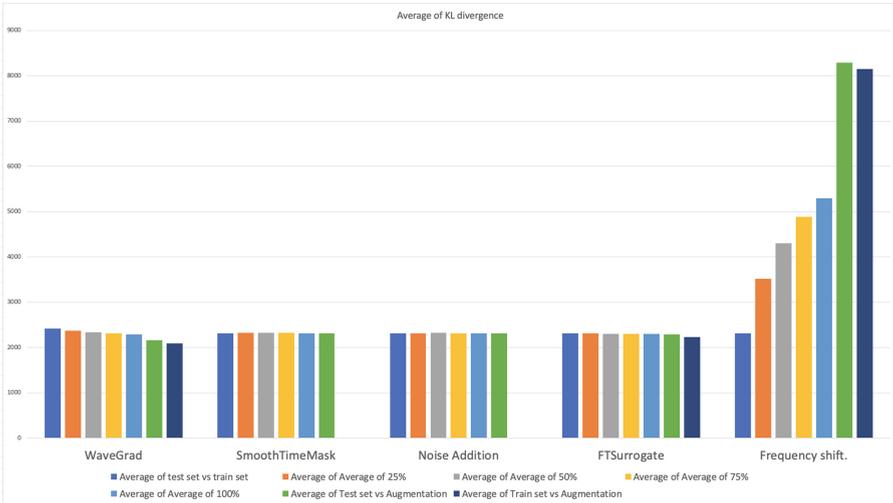


Fig. 2. Average of KL divergence each data augmentation method

4.2 Classification Performance

The baseline for the model was established by training it without any data augmentation, specifically without the use of synthetic data (i.e. 0% synthetic data). The average accuracy of standard EEG MI classification models were presented on the Table 2.

Table 2 showed the average accuracy of five standard EEG motor imagery (MI) classification models using different data augmentation techniques. The baseline accuracy of the models was 42.45%. The results showed that the highest accuracy was achieved using the **WaveGrad** technique, with an accuracy of 51.15%. This was an improvement of 8.7% over the baseline accuracy. As shown in Table 2, the highest accuracy values were obtained using **WaveGrad** (51.15%) and **Frequency Shift** (43.07%) techniques. The other two techniques, **Noise Addition** and Fourier transform surrogates, also improved the accuracy of the model to varying degrees. The **SmoothTimeMask** technique also resulted in higher accuracy values compared to the baseline, with the highest accuracy value of 38.84% obtained at a 25% size percentage.

Overall, our results demonstrated that DA techniques can be effective in improving the accuracy of EEG MI classification models. **WaveGrad** and **Frequency Shift** techniques were particularly promising, and were worth further investigation for future studies. Additionally, the size percentage of the DA techniques appeared to have an impact on model performance, suggesting that careful consideration of the amount of DA used was important for optimizing model accuracy.

Table 2. Average accuracy of five standards EEG MI classification models

Baseline Accuracy(%)	Size DA	WaveGrad (%)	σ	Noise Addition (%)	σ	Frequency Shift (%)	σ	Fourier transform surrogates(%)	σ	SmoothTimeMas (%)	σ
42.45	25%	51.15	17.61	35.83	13.95	43.07	18.28	41.25	14.31	38.84	12.38
	50%	50.56	17.38	33.46	11.10	44.49	17.82	37.03	10.85	36.27	10.84
	75%	51.64	17.24	33.77	11.23	43.58	18.25	37.01	10.14	34.95	10.73
	100%	50.78	17.72	31.39	8.84	42.61	18.80	36.20	11.03	34.31	9.51
		Average Accuracy (%)	51.03	17.49	33.86	11.53	43.44	18.04	37.12	11.08	36.09

4.3 Size of Augmentation

We present the impact of data augmentation on the similarity by KL-divergence in Fig. 2. This result show that **WaveGrad** are improve similarity when incurred number of ratio of data augmentation to raw signal. We present the impact on accuracy of ratio of data augmentation to raw signal in Table 2. This result show that the **WaveGrad** are not improve the accuracy.

5 Discussion

In this study, we demonstrated the implementation of the **WaveGrad**-based Diffusion Model as a DA for EEG MI datasets. From the result, our method

performed the accuracy based on the standard EEG MI classification models more than traditional DA. Thus, for the discussion section, we focused mainly on our method.

5.1 Key Improvement

The proposed methodology increased the variety of training sets, leading to an enhancement in the quality of the training set. As shown in Fig. 2, Our research indicated that our method enhanced the similarity of training sets with test set. The traditional DA (**Noise Addition**, **FTSurrogate**, and **SmoothTimeMask**) and training set without DA were the KL divergence are the most the same. Our method are improve the similarity of test set and train Fig. 2 our method more over another method are not improve the similarity. Spacial on **smooth time mask** and **noise add** the augmentation are not different the train set moreover the **Frequency shift** augment the seem make the train set and test set are more defence when decreased the similarity of train set and test set.

Our method are improve the similarity of test set and train set. Our method more over another method are not improve the similarity. Spacial on **SmoothTimeMask** and **Noise Addition** augmentation are not different the train set moreover the **Frequency shift** the seem make the train set and test set are more defence when decreased the similarity of train set and test set.

5.2 Subjects Level

In terms of research findings, it was observed that the data augmentation accuracy was highest for subjects who exhibited the greatest accuracy according to the standard EEG MI classification models prior to the implementation of data augmentation techniques. Subject 3's baseline accuracy was recorded to be 56.95% during the initial assessment. The research findings indicated that the accuracy for subject number 3 was improved by 15.46% using the implemented method. In contrast, the DA method employed in our study exhibited low efficacy when applied to the subject, as indicated by the subpar accuracy observed in relation to established EEG MI classification models. Subject 5's baseline accuracy was measured to be 28.08% in the study. The research findings indicate that a method has been implemented to enhance the accuracy of subject number 5, resulting in a 4.13% improvement. Table 3 displayed the mean increased in accuracy resulting from the implementation of our method.

5.3 Size of DA

The relationship between the size of DA and the performance of standard EEG classification was not a direct variation. Table 5 presented the average accuracy improvement achieved by our method on the DA size level. The limitations of the dataset that used to train the **WaveGrad** method have impacted its capacity to produce signals with greater valence. The observed limitation might account

Table 3. The Average accuracy improvement from our method on subject level

Subjects	Baseline Accuracy	Δ of WaveGrad
1	43.06	17.01
2	29.24	6.46
3	56.95	15.36
4	37.08	7.77
5	27.08	4.13
6	29.58	4.01
7	40.07	11.77
8	53.19	8.18
9	65.76	2.59

for the non-linear relationship between the size of the data augmentation and the corresponding increased in accuracy. Research suggested that increasing the number of samples may not necessarily lead to a significant enhancement in accuracy, particularly if the additional samples fail to encompass the complete spectrum of potential signal fluctuations.

However, despite this limitation, the WaveGrad method still outperformed the baseline accuracy on all levels of data augmentation, with the highest improvement achieved at a DA size of 75%. This finding suggested that even with a limited dataset, data augmentation techniques such as WaveGrad can still be effective in improving the accuracy of motor imagery classification using EEG signals.

It was worth noting that the results presented in the table were based on a single dataset, and the effectiveness of the WaveGrad method varies with other datasets or signal processing tasks. Future studies should investigate the generalizability of the WaveGrad method and explore its potential for improving accuracy in other EEG-based classification tasks.

The study concluded that although the WaveGrad method’s capacity to produce signals with higher valence is affected by the restricted dataset used for its training, the outcomes presented in the table indicated that the method can enhance the precision of the motor imagery classification task using EEG signals. Additional investigation was required to establish the applicability of the WaveGrad approach to alternative datasets and signal processing assignments (Table 4).

Table 4. The Average accuracy improvement from our method on DA size level

Baseline Accuracy	DA size	WaveGrad	σ	Δ
42.45	25%	51.15	17.61	8.71
	50%	50.56	17.38	8.11
	75%	51.64	17.24	9.19
	100%	50.78	17.72	8.33

Table 5. The Average accuracy improvement from our method.

Model	Baseline Accuracy	WaveGrad	σ	Δ
ATCNet	47.68	53.63	18.45	5.94
Deep ConvNet	32.60	52.93	18.49	20.33
EEGITNet	46.84	54.96	16.76	8.12
EEGNet	34.18	37.04	8.00	2.85
ShallowFBCSPNet	50.93	56.62	15.85	5.69

5.4 Complexity of Model

In a low-complexity model, our method improves accuracy more than a high-complexity model because our method improves the similarity of the train set and the test set Table 5. As a result, our method has an impact on low-complexity models. However, high-complexity models are at risk of overfitting on a training set. Hence, our method does not improve accuracy on high-complexity models. However, our method does not guarantee accuracy.

5.5 Limitation

First, the proposed data augmentation method was developed based on a limited dataset, specifically the BCI competition IV dataset 2a. While the method showed promising results on this dataset, its effectiveness might be limited by the size and quality of the dataset used for its development. Future work could involve evaluating the generalizability of the method on other EEG MI datasets to improve its confidence in wider usage scenarios. Second, the proposed data augmentation method was developed specifically for the EEG MI task. It might not necessarily generalize well to other EEG-based tasks. Future work could involve evaluating the method on other EEG signal tasks, such as P300 and SSEVP, to assess its robustness and adaptability across different EEG-based applications.

6 Conclusion

In this study, we investigated the effectiveness of data augmentation techniques on improving the accuracy of standard EEG MI classification models. We used five standard EEG MI classification models to classify EEG signals into left and right hand movements. We evaluated the performance of data augmentation techniques by comparing them with the baseline model, which were trained without any data augmentation. The data augmentation techniques, we used **WaveGrad**, **Noise Addition**, **Frequency Shift**, **Fourier transform surrogates**, and **SmoothTimeMas**.

Our results showed that data augmentation techniques improved the performance of standard EEG MI classification models, with **WaveGrad** being the most

effective technique. The accuracy of the baseline model was 42.45%, while the accuracy of the model trained with 75% synthetic data generated by **WaveGrad** was 51.64%. The other data augmentation techniques also improved the performance of the models, with **Noise Addition** and **Frequency Shift** being the least effective techniques.

References

1. Wolpaw, J.R.: Brain-computer interfaces. In: Handbook of Clinical Neurology, vol. 110, pp. 67–74. Elsevier (2013)
2. Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J.: EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *J. Neural Eng.* **15**(5), 056013 (2018)
3. Bilbao, I., Bilbao, J.: Overfitting problem and the over-training in the era of data: particularly for artificial neural networks. In: Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), pp. 173–177. IEEE (2017)
4. Wang, F., Zhong, S., Peng, J., Jiang, J., Liu, Y.: Data augmentation for EEG-based emotion recognition with deep convolutional neural networks. In: Schoeffmann, K., et al. (eds.) MMM 2018. LNCS, vol. 10705, pp. 82–93. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73600-6_8
5. Parvan, M., Ghiasi, A.R., Rezaii, T.Y., Farzamnia, A.: Transfer learning based motor imagery classification using convolutional neural networks. In: 2019 27th Iranian Conference on Electrical Engineering (ICEE), pp. 1825–1828. IEEE (2019)
6. Li, Y., Zhang, X.R., Zhang, B., Lei, M.Y., Cui, W.G., Guo, Y.Z.: A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding. *IEEE Trans. Neural Syst. Rehabil. Eng.* **27**(6), 1170–1180 (2019)
7. Schwabedal, J.T., Snyder, J.C., Cakmak, A., Nemati, S., Clifford, G.D.: Addressing class imbalance in classification problems of noisy signals by using Fourier transform surrogates. arXiv preprint [arXiv:1806.08675](https://arxiv.org/abs/1806.08675) (2018)
8. Rommel, C., Moreau, T., Paillard, J., Gramfort, A.: CADDA: class-wise automatic differentiable data augmentation for EEG signals. arXiv preprint [arXiv:2106.13695](https://arxiv.org/abs/2106.13695) (2021)
9. Rommel, C., Paillard, J., Moreau, T., Gramfort, A.: Data augmentation for learning predictive models on EEG: a systematic comparison. *J. Neural Eng.* **19**(6), 066020 (2022)
10. Mohsenvand, M.N., Izadi, M.R., Maes, P.: Contrastive representation learning for electroencephalogram classification. In: Machine Learning for Health, pp. 238–253. PMLR (2020)
11. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems, vol. 33, pp. 6840–6851 (2020)
12. Akrouf, M., et al.: Diffusion-based data augmentation for skin disease classification: impact across original medical datasets to fully synthetic images. arXiv preprint [arXiv:2301.04802](https://arxiv.org/abs/2301.04802) (2023)
13. Chen, N., Zhang, Y., Zen, H., Weiss, R.J., Norouzi, M., Chan, W.: WaveGrad: estimating gradients for waveform generation. arXiv preprint [arXiv:2009.00713](https://arxiv.org/abs/2009.00713) (2020)
14. Song, Y., Garg, S., Shi, J., Ermon, S.: Sliced score matching: a scalable approach to density and score estimation. In: Uncertainty in Artificial Intelligence, pp. 574–584. PMLR (2020)

15. Brunner, C., Leeb, R., Müller-Putz, G., Schlögl, A., Pfurtscheller, G.: BCI Competition 2008-Graz data set A. Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology, vol. 16, pp. 1–6 (2008)
16. Altaheri, H., Muhammad, G., Alsulaiman, M.: Physics-informed attention temporal convolutional network for EEG-based motor imagery classification. *IEEE Trans. Ind. Inf.* **19**(2), 2249–2258 (2022)
17. Salami, A., Andreu-Perez, J., Gillmeister, H.: EEG-ITNet: an explainable inception temporal convolutional network for motor imagery classification. *IEEE Access.* **10**, 36672–36685 (2022)
18. Schirrneister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggenesperger, K., Tangermann, M., et al.: Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain Mapp.* **38**(11), 5391–5420 (2017)
19. Lashgari, E., Liang, D., Maoz, U.: Data augmentation for deep-learning-based electroencephalography. *J. Neurosci. Methods* **346**, 108885 (2020)
20. Leeb, R., Brunner, C., Müller-Putz, G., Schlögl, A., Pfurtscheller, G.: BCI Competition 2008-Graz Data Set B, pp. 1–6. Graz University of Technology, Austria (2008)
21. Terzano, M.G., Parrino, L., Sherieri, A., Chervin, R., Chokroverty, S., Guilleminault, C., et al.: Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep. *Sleep Med.* **2**(6), 537–554 (2001)