Iteratively Calibratable Network for Reliable EEG-Based Robotic Arm Control

Byeong-Hoo Lee, Jeong-Hyun Cho, Byung-Hee Kwon, Minji Lee, *Member, IEEE*, and Seong-Whan Lee, *Fellow, IEEE*

***Abstract*— Robotic arms are increasingly being utilized in shared workspaces, which necessitates the accurate interpretation of human intentions for both efficiency and safety. Electroencephalogram (EEG) signals, commonly employed to measure brain activity, offer a direct communication channel between humans and robotic arms. However, the ambiguous and unstable characteristics of EEG signals, coupled with their widespread distribution, make it challenging to collect sufficient data and hinder the calibration performance for new signals, thereby reducing the reliability of EEG-based applications. To address these issues, this study proposes an iteratively calibratable network aimed at enhancing the reliability and efficiency of EEG-based robotic arm control systems. The proposed method integrates feature inputs with network expansion techniques. This integration allows a network trained on an extensive initial dataset to adapt effectively to new users during calibration. Additionally, our approach combines motor imagery and speech imagery datasets to increase not only its intuitiveness but also the number of command classes. The evaluation is conducted in a pseudo-online manner, with a robotic arm operating in real-time to collect data, which is then analyzed offline. The evaluation results demonstrated that the proposed method outperformed the comparison group in 10 sessions and demonstrated competitive results when the two paradigms were combined. Therefore, it was confirmed that the network can be calibrated and personalized using only the new data from new users.**

***Index Terms*— Brain–machine interface, electroencephalogram, robotic arm, deep learning, network calibration.**

Manuscript received 23 October 2023; revised 17 January 2024, 24 April 2024, and 25 June 2024; accepted 23 July 2024. Date of publication 29 July 2024; date of current version 6 August 2024. This work was supported in part by the National Research Foundation of Korea (NRF) Grant funded by Ministry of Science and ICT (MSIT) (MetaSkin: Developing Next-Generation Neurohaptic Interface Technology that Enables Communication and Control in Metaverse by Skin Touch) under Grant 2022-2-00975; and in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant, funded by the Korea Government (MSIT) (Artificial Intelligence Graduate School Program, Korea University), under Grant 2019-0-00079. *(Corresponding author: Seong-Whan Lee.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Korea University under Approval No. 1040548-KU-IRB-17-172-A-2.

Byeong-Hoo Lee, Jeong-Hyun Cho, and Byung-Hee Kwon are with the Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea (e-mail: bh_lee@korea.ac.kr; jh_cho@korea.ac.kr; bh_kwon@korea.ac.kr).

Minji Lee is with the Department of Biomedical Software Engineering, The Catholic University of Korea, Bucheon 03083, South Korea (e-mail: minjilee@catholic.ac.kr).

Seong-Whan Lee is with the Department of Artificial Intelligence, Korea University, Seoul 02841, South Korea (e-mail: sw.lee@korea.ac.kr).

This article has supplementary downloadable material available at https://doi.org/10.1109//TNSRE.2024.3434983, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2024.3434983

I. Introduction

ADVANCES in robot intelligence enable a robotic arm to interact with humans physically [1], [2]. For safe and precise control, the robotic arm should interpret the control intentions of humans and execute accordingly. One of the intuitive ways to detect human intentions is by decoding brain signals such as electroencephalogram (EEG) signals [3]. EEG signals can be obtained non-invasively and are thus practical for external device control [4], [5], [6]. Brain-machine interface (BMI) commonly uses EEG signals as control signals [7], and therefore this study applies BMI settings in the context of robotic arm control. Two major endogenous paradigms are introduced to consistently generate EEG signals: motor imagery (MI) [8] and speech imagery (SI) [9]. MI involves imagining muscle movements without actual physical motion [8]. Meanwhile, SI—also known as imagined or silent speech—involves imagining words and sentences without actually speaking them out loud. SI is known as an intuitive paradigm because it simply requires a speaking imagination [10]. Through intuitive paradigms, BMI allows users to intuitively control external devices.

Although BMI has shown its potential, several obstacles hinder its practical use: (1) Due to the complex and unstable characteristics of EEG signals [11], the decoding algorithm needs to be calibrated to maintain its performance. However, these characteristics cause the data to form a large distribution, which interferes not only with the calibration of decoding algorithm for new data but also with data sharing among users. This leads to a degradation in performance, thereby undermining the reliability of BMI-based applications. (2) Factors such as user fatigue and concentration level can influence the generation of these inconsistent EEG signals [12]. As a result, obtaining ample amounts of data from individual users is challenging, which limits the number of classes available for BMI application. (3) The presence of background noise in EEG signals [13] and low signal-to-noise ratio leads to inconsistent data quality. This makes it difficult to apply advanced

machine learning and even data augmentation techniques [14], further increasing data distribution and complicating inter-user data combination—thereby posing challenges in acquiring sufficient data.

To address the aforementioned issues, we propose an iteratively calibratable decoding algorithm for EEG-based robotic arm control. Under the assumption of daily usage conditions, a session is defined as a sequence of steps where users wear an EEG cap, control a robotic arm based on their EEG signals and then remove the cap. The aim of this study is to enhance classification performance, enabling it to be calibrated for unseen users in each session. Moreover, as data from new users accumulate, the study also involves an experiment on whether the decoding algorithm can be personalized. To this end, a pipeline is proposed that increases the capacity of the neural network through feature input and network expansion strategies. However, to prevent indiscriminate network expansion, selective initialization and training are applied prior to expanding the network. Our baseline evaluation was performed using publicly available datasets. Subsequently, we conducted a total of 10 sessions with novice users for data collection; this collected data was then divided into calibration and test sets. The calibration set was collected under identical circumstances as the public datasets and was employed to adjust the model parameters. In contrast, the test set was collected through pseudo-online test in an environment where the robotic arm was controlled, and the collected data was evaluated and analyzed offline, as detailed in references [15], [16]. Using these datasets, the proposed method was calibrated to ensure classification performance on unseen data. Essentially, our goal is to ascertain if the proposed method can effectively adapt to EEG data that significantly deviates from the initial training data. Therefore, experiments were conducted to test whether the proposed system exhibits resilience against variations in EEG patterns as sessions progress.

The main contributions are as follows: (1) The proposed method consistently improved classification performance across 10 experimental sessions. The results indicated that the suggested approach is robust against variations in EEG patterns and exhibits proficiency in personalizing decoding algorithm. (2) To the best of our knowledge, this is the first attempt at combining two intuitive endogenous BMI paradigms with the aim to increase the number of control signals for intuitive robotic arm control. (3) We introduced a network capacity expansion strategy to prevent the decrease in decoding algorithm calibration performance due to the large distribution of EEG data during the calibration stage, thereby obtaining a decoding algorithm robust against the variability of EEG and advantageous for real-life use. (4) Through ablation studies, it is confirmed that the proposed method could be considered an efficient approach towards practical EEG-based robotic arm control using advanced machine learning methods.

II. Related Work

In recent years, numerous studies have contributed to improving endogenous BMI classification performance based on machine learning techniques. Schirrmeister et al. [17] proposed a deep and shallow convolutional neural networks (CNNs) to classify EEG signals and provided an understanding of the network design and training through EEG features extracted by CNNs. Particularly, the shallow ConvNet pipeline was inspired by the filter bank common spatial pattern (FBCSP) [18], which specifically deals with band power features that are advantageous for MI classification. Lawhern et al. [19] introduced depth-wise and separable convolutions to summarize individual features over time, by considering more channel information. They applied separable convolution to the BMI and demonstrated that CNNs can be trained with a small number of parameters. Amin et al. [20] proposed the use of multiple CNN models to extract different levels of relevant features from the raw EEG signals. They explored raw data using CNN models of different depths to extract abundant features. Xie et al. [21] introduces Transformer-based models designed for the classification of MI EEG signals. By utilizing the attention mechanism inherent in Transformers, these models are capable of extracting features from long-sequence data and providing visualization. The spatial-temporal dependencies found within EEG signals serve as vital information for precise classification. Song et al. [22] proposed a model called EEG Conformer that employs a convolution module for learning low-level local features and a self-attention module to extract global correlations from these features. The model also uses a simple classifier based on fully-connected layers for EEG signal categorization. Additionally, the paper introduces a visualization strategy that projects class activation mapping onto brain topography to enhance interpretability.

Some studies have conducted investigations on multi-session and multi-paradigm approaches. Tam et al. [23] carried out a multi-session investigation to determine a minimal set of electrodes for individual stroke patients, which were utilized in MI tasks to control assistive devices through functional electrical stimulation across 20 sessions. Lee et al. [24] presented a dataset for brain-computer interface (BCI) systems, incorporating three key paradigms: MI, event-related potential and steady-state visually evoked potential. The data, collected from numerous subjects across multiple sessions, includes psychological and physiological user details. The study evaluates decoding accuracies per paradigm and investigates performance differences across subjects and sessions. Thomas et al. [25] examined evaluation metrics for increasingly complex MI BCI, featuring adaptive classification, error detection and correction, signal fusion and shared control. They encompassed simulated and experimental data, also surveying recent literature to understand BCI evaluations, particularly focusing on the correlation between data usage and the BCI subcomponent under scrutiny.

However, the aforementioned studies still raise questions about their reliability. This is because variations in EEG patterns limit the number of classes and contribute to performance degradation, thereby hindering further progress in research. Therefore, it is imperative that the decoding network be adequately calibrated to unfamiliar EEG patterns, necessitating a substantial calibration process. However, applying personalized network weights to each individual user proves challenging due to significant differences in EEG signal

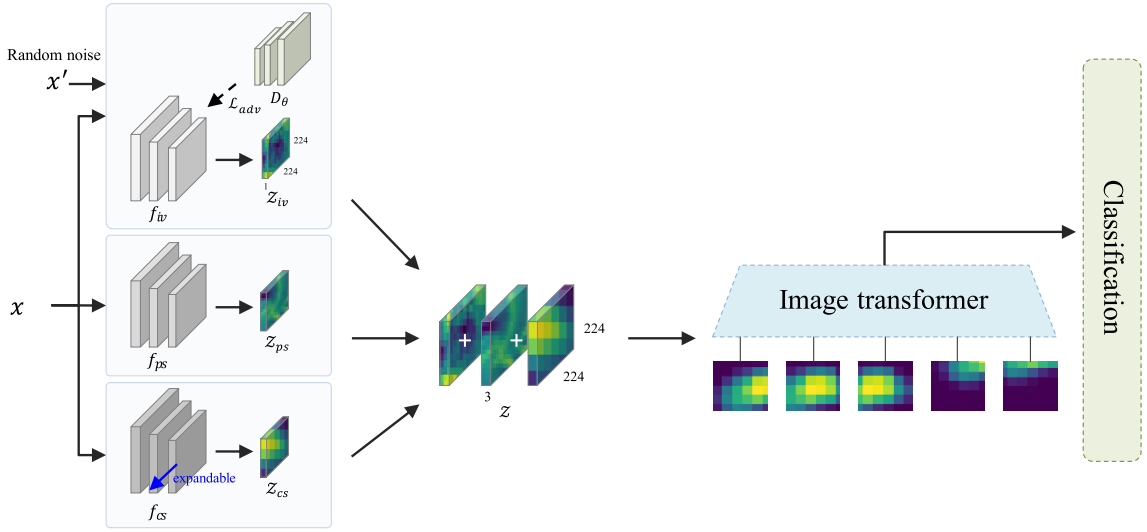

Fig. 1. Overall architecture. Three types of features are obtained from the feature extractors. The adversarial loss $\mathcal{L}_{adv}$ is introduced to extract EEG-invariant features z_{iv} and consequently, fixed random noise x' is provided as an anchor for the discriminator D_θ. Paradigm-specific z_{ps} and class-specific features z_{cs} are obtained by their corresponding feature extractors f_{ps} and f_{cs}. These three features are concatenated, yielding the same shape as the image and are provided as feature input z to the image transformer. The image transformer then conducts the classification.

patterns among users. This presents a complication as it becomes difficult to predict various users' intentions with only one set of network weights, which restricts its reliability. Therefore, applications based on BMI must undergo reliability assessments across multiple sessions.

A. Prerequisites

We posit that the major obstacles to reliable BMI are the lack of classification performance due to large variations in EEG patterns and the diversity of EEG signals between different users [14], [26]. We hypothesize that a network with sufficient capacity can robustly handle diverse EEG patterns. Therefore, our proposed method primarily focuses on ensuring efficient network capacity to mitigate these aforementioned problems.

Sufficient capacity allows the network to learn from new data in order to calibrate its parameters, even if this data greatly differs from the initial training dataset. One way to increase network capacity is through network expansion. For efficient expansion of the network, training is required to identify redundant nodes before initializing and retraining them; new nodes should be added as necessary. However, despite the efficient growth of the network, expansion leads to increased inference time and computational costs. Therefore, it's essential for an expanded network to be compressed in order to maintain computational costs. Thus our proposed method incorporates both network expansion and compression steps while maintaining manageable computational costs.

III. Methods

To achieve robust performance amidst variations in EEG patterns of users, this section describes novel strategies including feature inputs, image transformer, network expansion, and compression for maintaining constant computational complexity. The architecture of the proposed method is depicted in Fig. 1.

A. Dataset Description

1) Baseline Dataset: We investigate whether the network can improve the classification performance of new users, even though it was initially trained using a public dataset which may contain significant distribution differences. We used BCI competition 2020 Track #3 and 4 [27] which are SI and MI datasets to construct a baseline dataset $D_{base} = \{D_{tr}, D_{te}\}$. The datasets were combined in accordance with the subjects (e.g. subjects # of Tracks #3 and #4 were considered as the same subject). The baseline training was conducted in a subject-independent manner [26] such that D_{te} was one of the subjects of D_{tr} according to leave-one-subject-out validation. Since Track #4 is composed of three MI classes ('cylindrical', 'spherical', and 'lumbrical'), we selected 'help me', 'thank you', and 'yes' from Track #3 to form a total of 6 classes, as these are suitable for robotic arm control. To avoid the data imbalance problem, fifty trials were randomly selected from 'help me', 'thank you' and 'yes' trials. Thus, combined dataset D_{tr} and D_{te} contained 50 trials per class and total of 4200 and 300 trials, respectively. Each trial $x \in \mathbb{R}^{C\times T}$ was downsampled at 250 Hz yielding 1000 time points T, hence an imagery period lasting for four seconds was applied for each of the overlapping 58 channels C. Selected channels are described in the supplementary document Section I.

2) Calibration Dataset: We collected a calibration dataset from ten naive subjects aged 24 to 30 years (5 males and 5 females, all right-handed). Subjects were presented with class labels for a duration of one second via the monitor display, followed by a four-second imagination period accompanied by a sound cue. This procedure was conducted in accordance with the same environment and protocol as the recording environment of D_{base}. Once the imagination period is over, a two-second rest period is provided. Prior to the EEG recording, a 20-minute practice session is assigned. Once the recording begins, the subjects are instructed to minimize movements during the imagination period and to

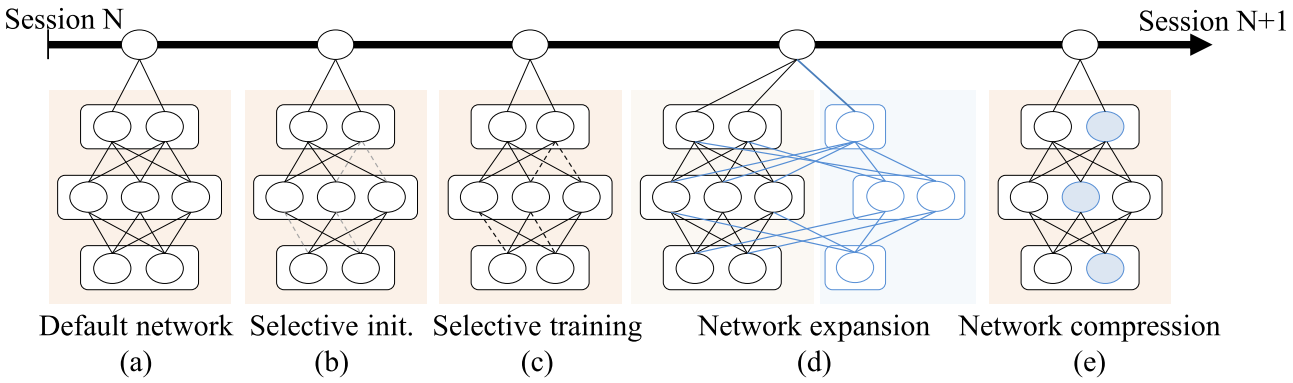

Fig. 2. Network expansion pipeline of class-specific feature extractor. (a) The default network is a fixed size of the network. (b) Selective initialization and (c) training are performed to improve calibration performance without expansion. (d) If a larger network capacity is required, nodes are added to the network to increase feature size. Consequently, the size of the network continues to grow. (e) The expanded network shrinks to the default network size while maintaining the knowledge learned from calibration data through knowledge distillation.

conduct imagination according to the corresponding label. The collected dataset $\hat{D^j} = \{\hat{D^j_{tr}}, \hat{D^j_{val}}\}$ consists of 6-class with the same labels as D_{base} and contains 20 training trials and 10 validation trials per class, yielding 120 training trials and 60 validation trials per subject, where j ranges from 1 to 10, representing the session number.

3) Test Dataset: The test dataset $\hat{D^j_{te}}$ is associated with a robotic arm control scenario. The scenario involved the robotic arm picking up an object using various grasping methods and handing it to the subject. In contrast to the calibration dataset, the monitor display is not included in the recording. Instead, only the sound cue and robotic arm are provided to the subject, as introduced in [28]. Prior to the experiment, the subjects were informed of the robotic arm control scenario and were given sufficient preparation time for the imagery period before recording $\hat{D^j_{te}}$. It is composed of 6-class and 10 recording sessions yielding $\hat{D^j} = \{\hat{D^j_{tr}}, \hat{D^j_{val}}, \hat{D^j_{te}}\}$. It consists of 12 trials per subject, with each subject generating three trials each of the 'help me', 'yes' and 'thank you' classes and one trial each of the 'cylindrical', 'spherical' and 'lumbrical' classes. A detailed description of the scenario can be found in Section II of the supplementary document. The operation of the robotic arm to collect $\hat{D^j_{te}}$ was real-time but was independent of model predictions. Subjects generated EEG signals in response to the robotic arm operation according to predefined scenarios. The model predictions never controlled the robotic arm; however, this information was not disclosed to the subjects in order to make them believe that they were controlling the robotic arm themselves, thereby inducing a sense of agency to improve EEG quality [5].

$\hat{D^j}$ is originally recorded at 2000 Hz and downsampled to 250 Hz. It is known that the *mu* band (8-13 Hz) and beta band (13-31 Hz) include MI features [17], [19] and also one of the efficient bandwidths for SI classification is 30-125 Hz [29]. Therefore, a band-pass filter (8-125 Hz) and a 60 Hz notch filter are applied to D^j. This not only includes all these ranges but also excludes the lower bandwidth where eye movement artifacts are prominently present [30], [31], [32]. The impact of an imprecise filtering choice on network performance is challenging to predict [33], [34]. Moreover, losing information through filtering from the training data contradicts our training strategy, which aims to maximize the latent space of the training data and pursue data augmentation of feature inputs induced by modulation. We provide experimental results to support this assumption in Table V of the supplementary document. We did not apply any additional filtering methods, such as eye movement or noise removal filtering, to maintain standard performances. Fig. 2 in the supplementary document illustrates the environment used for calibration dataset collection. Furthermore, we represent the experimental results of applying four types of filtering methods in Table V of the supplementary document.

To prevent the data from being biased toward a particular class or becoming too familiar, the classes were randomly presented to the subjects as they constructed $\hat{D^j}$. The experimental setup was created in MATLAB 2019a (MathWorks Inc., USA), and EEG signals were recorded using the BBCI toolbox [35]. This study obtained consent from all subjects for data collection and duration. The protocols and environments were reviewed and approved by the Institutional Review Board of Korea University [1040548-KU-IRB-17-172-A-2].

B. Feature Inputs and Network Expansion

It has been established that spectral amplitude features are effective for EEG classification [17], [19]. As such, convolution layers are employed for feature extraction. Three convolution-based feature extractors were introduced to obtain three types of features: EEG-invariant z_{iv}, paradigm-specific z_{ps} and class-specific z_{cs} features from $D_{tr} = (x_i, y_i)$ where $x_i \in X$ represents tuple inputs and $y_i \in Y$ denotes the corresponding class label.

To obtain $z_{iv} \in Z_{iv}$, adversarial learning was applied without a generator [36]. The invariant feature extractor ($f_{iv} : X \rightarrow Z_{iv}$) generates features to fool the discriminator ($D_\theta : Z_{iv} \rightarrow K$) using x and random noise ($x' \in \mathbb{R}^{C\times T}$) where $K \in \{0, 1\}$ denotes true or fake (binary class). Thus, the loss function is defined as follows:

$$\mathcal{L}_{adv} = \min_{f_{iv}} \max_{D_\theta} \sum_{i=0}^{N} k_i log(D_\theta(f_{iv}(x_i))), \quad (1)$$

where N is the number of trials in D_{tr}^{j}. In this training, i starts from 0 because x_i includes x'. Note that x' is fixed random noise. The goal is to train f_{iv} to induce D_θ to confuse z_{iv} and fake features z'_{iv} from x'. Therefore, z'_{iv} explicitly serves as an anchor for f_{iv} to extract "z'_{iv}-like" z_{iv}. As training progresses, it is conjectured that z_{iv} should resemble z'_{iv} (i.e., they should be similar to "fake" features), thereby containing EEG-invariant information. Although z_{iv} is not directly associated with classification, it is designed to enhance classifier training performance by expanding the input latent space.

To obtain z_{ps} from x_i and its paradigm label $p_i \in P$, paradigm-specific extractor ($f_{ps} : X \rightarrow P$) is trained to extract features that can distinguish between MI and SI. Therefore, surrogate binary labels are provided for the training instead of Y. The cross-entropy loss is formulated as:

$$\mathcal{L}_{ps} = -\frac{1}{N}\sum_{i=1}^{N} \hat{p}_i log(f_{ps}(x_i)). \tag{2}$$

Therefore, f_{ps} is trained to classify paradigms. Similar to z_{iv}, these features are not directly associated with classification but are provided to expand the input latent space of the classifier.

In contrast, z_{cs} is more closely associated with classification because its objective aligns with that of the classifier. Class-specific extractor f_{cs} is designed for 6-class classification. However, it is designed as an expandable network to increase its capacity in response to variations in EEG patterns, a concept inspired by [37]. Firstly, we sparsely train the network and then sequentially expand its capacity to avoid excessive network expansion. To this end, elementwise L_1-norm is added to the loss function to obtain a sparse network by penalizing weights. This can be formulated as follows:

$$\mathcal{L}_{cs} = -\frac{1}{N}\sum_{i=1}^{N} \hat{y}_i log(f_{cs}(x_i)) + \lambda \sum_{l=1}^{L} |W_{base}^{l}|, \tag{3}$$

where W_{base}^{l} denotes weights of l^{th} layer of the network trained using D_{base} and λ and L denote regularization parameter and the number of layers, respectively.

Once $\hat{D}$ is provided, the calibration begins with W_{base}. At a certain point T_1 during calibration, the network searches the nodes that have been trained in associated with the $\hat{D}$. The W_{base} is partially updated with L_1 regularization from the topmost hidden layer. The connections between L are searched by solving the equation obtaining w^{part}, defined as:

$$\underset{W^{part}}{min}\,(\mathcal{L}(W^{part}; W_{base}^{1:L-1}, \hat{D}) + \lambda_1 |W^{part}|), \tag{4}$$

where $W_{base}^{1:L-1}$ and λ_1 denote the weights excluding W^{part} and L_1 is the regularization term. Since all non-zero connections in W^{part} are associated with $\hat{D}$, the breadth-first search is conducted on all corresponding nodes G to identify all connections between every node. It includes all nodes related to $\hat{D}$; in other words, the other node group $\bar{G}$ comprises nodes that are not associated with $\hat{D}$. If $\bar{G}$ is calibrated, it would enhance calibration performance. As such, the weights of unrelated nodes $W^{\bar{G}}$ are initialized and the calibration continues until T_2. This process is selective initialization and is carried out only once during the calibration.

At T_2, if loss remains above the threshold τ, selective training commences. This process assumes that W^G has been adequately calibrated and thus freezes W^G, calibrating only $W^{\bar{G}}$ until T_3. Selective training allows gradients, which typically have a negligible contribution (and thus could be ignored), to influence the training process by solving the following equation:

$$\underset{W^{\bar{G}}}{min}(\mathcal{L}(W^{\bar{G}}, \hat{D}) + \lambda_2 ||W^{\bar{G}}||), \tag{5}$$

where $\lambda_2||W^{\bar{G}}||$ denotes L_2 regularization term to avoid increasing the complexity of $W^{\bar{G}}$.

If the loss value remains higher than τ even after T_3, then the network's capacity needs to be increased. In response, an arbitrary number of o nodes are added to the layers, consequently expanding the network weights $W = [W; W^E]$, where W^E denotes the weights of added nodes (the expanded parts). Since an arbitrary number of o nodes can be added, objective function includes calculating the optimal number of nodes for each layer in order to minimize weight complexity, which is defined as follows:

$$\underset{W}{min}(\mathcal{L}(W; W^E, \hat{D}) + \lambda |W| + \lambda_g \sum_{e} ||W^e||), \tag{6}$$

where λ_g and e denote the regularization parameter and a set of activated connections of added nodes, respectively. The term $\lambda_g \sum_e ||W^e||$ is the filter-wise group LASSO [38] to remove unnecessary nodes to obtain the optimal number of nodes for all layers. Essentially, this process results in a temporary increase in the number of convolution channels. As a result, network is expanded, allowing it to extract more critical features for classification. Algorithm 1 provides an overview of f_{cs} training.

C. Network Compression

We propose a network that temporarily expands f_{cs} to calibrate weights when a large deviation is observed between D_{base} and $\hat{D}^j$. This is achieved by increasing the network capacity. However, due to limitations in computational complexity and inference time, it's crucial to maintain the size of the network constant. To address this issue, we introduce a network compression step before the next calibration. We employ basic knowledge distillation as outlined by Hinton et al., 2015 [39], which allows us to preserve the size of our model and transfer knowledge from the expanded network using a defined distillation loss.

$$\mathcal{L}_d = \sum_{(x,y)\in\hat{D}_j} \mathcal{L}_{KD}(f_{cs}(x, W_{j-1}), f_{cs}(x, W_j)) + \gamma \mathcal{L}_{CE}(f_{cs}(x, W_{j-1}), y), \tag{7}$$

where $\mathcal{L}_{KD}$ and $\mathcal{L}_{CE}$ denote knowledge distillation and cross-entropy loss, respectively. Initially (j=1), W_0 is W_{base}. Fig. 2 depicts the pipeline of network expansion and compression. Note that baseline training is performed once.

Algorithm 1 Network Expansion

Input: Calibration dataset $\hat{D}$, training epochs T_1, T_2, T_3, T_4, threshold τ
Output: Calibrated network weights W
Procedures:
Initial calibration
Solve eq. (4) until T_1
Selective initialization
Obtain $\bar{G}$
Initialize $W^{\bar{G}}$
Solve eq. (4) until T_2
if $\mathcal{L} > \tau$ **then**
  Selective training
  Solve eq. (5) to calibrate $W^{\bar{G}}$ until T_3
end
if $\mathcal{L} > \tau$ **then**
  Network expansion
  Solve eq. (6) to calibrate $W = [W; W^E]$ until T_4
end
Obtain W

D. Image Transformer With Data Augmentation

DeiT, an image transformer [40], offers various sizes of distilled models. However, it's uncertain whether DeiT would be efficient in BMI applications where the amount of data is limited. To address this issue, we applied simple data augmentation techniques in the image domain. Gaussian noise added D'_{tr}, data flipped over time dimension $\vec{D}_{tr}$ and kernel dimension $\tilde{D}_{tr}$ datasets were added to baseline and collected dataset. Through data augmentation techniques, six times more trials than the original dataset are obtained, yielding $D_{tr} = \{D_{tr}, D'_{tr}, \vec{D}_{tr}, \vec{D}'_{tr}, \tilde{D}_{tr}, \tilde{D}'_{tr}\}$ (for simplicity, $\hat{D}^{j}_{tr}$ is omitted).

The underlying assumption of this approach is that even if these techniques may compromise the essential information of x, DeiT would enhance classification performance if it takes z as input. In other words, the classifier can avoid underfitting by indirectly learning from the augmented data while minimizing the loss of essential information. While all f are trained directly with raw augmented data, making them susceptible to performance drops, any such decrease is deemed insignificant if classifier f_θ exhibits a performance enhancement. In this context, we consider any performance drop as inducing modulations of feature inputs z. We selected small and tiny versions of DeiT, which are applicable to BMI, as f_θ. To implement it, we accumulated z_{iv}, z_{ps} and z_{cs} along a axis to form three channels corresponding to RGB channels of an image. For multi-class classification tasks, we define the cross-entropy loss as follows:

$$\mathcal{L}_\theta = -\frac{1}{N}\sum_{i=1}^{N} \hat{y}_i log(f_\theta(z)), \tag{8}$$

where Y is same class labels used when obtaining z_{cs}. Initially, f_θ employs the Adam optimizer for baseline training but switches to stochastic gradient descent (SGD) during calibration [41].

TABLE I
PARAMETERS SELECTED FOR THE EXPERIMENTS. PARAMETERS WERE HEURISTICALLY DETERMINED

T_1	T_2	T_3	T_4	τ	o	λ	λ_1	λ_2	λ_g	γ
200	300	400	500	0.02	25	0.5	0.00001	0.0001	0.001	0.5

IV. EXPERIMENTS

A. Evaluation Protocols

The evaluations are primarily segmented into three components: (1) Baseline evaluation, which assesses the classification performance on the baseline dataset; (2) Calibration evaluation, which assesses the model performance on the calibration dataset and is designed to quantify calibration efficiency; and (3) Test evaluation, intended to gauge the classification efficacy of networks within a robotic arm control scenario. Decoding involves the process of matching model predictions to labels based on a predefined size of model input. Thus, continuous EEG signals are segmented into uniform sizes, with each segment considered a single trial. The model predictions are discrete, and classification accuracy is defined by calculating the extent to which the labels of the corresponding trials match the model predictions. Note that the labels for classifying between SI and MI are only applied when training the paradigm-specific feature extractor and for all evaluations, including test evaluations, accuracy is assessed based on 6-class classification.

All parameters used in the experiments are listed in Table I. The evaluations were executed on a system equipped with an Intel Core i9 12900K CPU operating at 3.20 GHz, 128 GB of DDR4 RAM, twelve NVIDIA TITAN V GPUs (each functioning at 1200 MHz) and Python version 3.9 complemented by PyTorch version 1.12 and CUDA 11.3.

B. Comparison Group and the Proposed Methods

Publicly available methods were employed for evaluation: Common spatial pattern (CSP)-based methods [42], FBCSP [18], EEGNet [19], Deep and Shallow ConvNet [17], ERA-CNN [43], MCNN [20], EEG-Transformer [44] and EEG Conformer [22] which are described in Section II. To ensure fairness in evaluation, the hyperparameters of the comparison group were meticulously adjusted to accommodate the size of the dataset. The data cropping method designed by [17] was adopted for CNN-based methods utilizing a sliding time window with a stride of 100 ms. The mean value of all cropped data served as the final prediction [17]. CSP projects the signal into a novel space and is designed to maximize the variance for one class while simultaneously minimizing it for another. Random forest (RF), support vector machine (SVM) and linear discriminant analysis (LDA) were selected as classifier for CSP [18], [45]. FBCSP operates by extracting features from EEG signals that have been processed through multiple bandpass filters based on CSP to identify and select the most discriminative features from each filter bank. For experiments, LDA was selected as classifier [18]. For evaluating the proposed methods, we selected VGG16 [46], tiny and small DeiT, which respectively have 13.8 M, 5 M and 22 M trainable parameters. We opted for VGG16 as it is widely recognized as a baseline classifier. We conducted training until

TABLE II
RESULTS OF BASELINE EVALUATION. PARENTHESES"()" ENCOMPASS STANDARD DEVIATION. SUPERSCRIPTS ^{v}, ^{t} AND ^{s} DENOTE VGG16, TINY AND SMALL DEIT. MI AND SI DENOTE INDIVIDUAL DATASETS. $|_{v}$ DENOTES THE p-VALUE OF VGG16 IS GREATER THAN 0.05. THE HIGHEST PERFORMANCES ARE HIGHLIGHTED IN BOLD. CHANCE LEVELS OF MI AND SI ARE 33.33 AND COMBINED IS 16.67

Methods	MI	SI	Combined	p-value
CSP+RF [45]	35.36 (5.51)	37.84 (3.75)	19.88 (5.48)	< 0.05
CSP+SVM [18]	36.35 (2.23)	39.51 (4.65)	21.61 (6.89)	< 0.05
CSP+LDA [18]	36.01 (3.85)	38.06 (3.67)	22.98 (6.12)	< 0.05
FBCSP [18]	43.06 (4.75)	46.61 (5.54)	25.41 (9.10)	< 0.05
EEGNet [19]	64.97 (8.19)	**70.28 (8.02)**	52.56 (4.30)	< 0.05
Deep ConvNet [17]	62.58 (6.76)	67.99 (7.69)	50.52 (4.02)	< 0.05
Shallow ConvNet [17]	63.39 (6.91)	68.67 (8.41)	53.46 (4.72)	< 0.05
ERA-CNN [43]	57.34 (4.41)	62.80 (5.03)	54.83 (3.76)	$< 0.05\|_{v}$
MCNN [20]	64.59 (5.12)	69.09 (7.49)	49.32 (3.81)	$< 0.05\|_{v}$
EEG-Transformer [44]	64.32 (6.59)	67.85 (6.90)	51.58 (5.26)	< 0.05
EEG Conformer [22]	**67.44 (6.16)**	63.52 (5.11)	53.37 (8.94)	$< 0.05\|_{v}$
Proposed Method^{v}			57.99 (5.12)	
Proposed Method^{t}	64.11 (3.25)	67.65 (4.43)	**67.74** (3.84)	
Proposed Method^{s}			67.05 (5.37)	

500 epochs and selected the network weights that yielded minimal loss post 200 epochs [19]. Here, an 'epoch' refers to a complete pass of the dataset through the algorithm. For baseline evaluation we utilized Adam optimizer [47] with a learning rate set at 0.001, the mini-batch size at 128 and weight decay at 0.01. During calibration, we performed training using an SGD optimizer along with early stopping [48]. Additionally, we set 500 epochs for knowledge distillation with a default temperature (=1).

C. Calibration and Test Evaluation Conditions

Given that the baseline dataset comprises data originating from 15 individuals, it yields 15 separate network weights. This procedure is exclusively conducted during session 1, and only one network weight is selected per test. The initially selected network weights are then calibrated and applied to tests over the subsequent 10 sessions. Therefore, the network weights used in the final session are the final weights that have been calibrated to be personalized for each individual subject.

V. RESULTS AND DISCUSSION

A. Baseline Evaluation

Table II shows the results of the baseline dataset and individual paradigms. In single paradigm classification, the class-specific feature extractor conducts evaluation. Given ERA-CNN's two-branch architecture, we selected the shared layer for individual paradigm classification. Transformer-based methods showed around 65% classification accuracy. EEG Conformer [22] outperformed other methods in MI classification (67.44%). CSP-based methods showed a slightly higher accuracies compared to the chance level (33.33) and demonstrated relatively lower performance compared to other techniques. EEGNet [19], as a CNN architecture, surpassed in SI classifications. The results indicate that SI classification performance outstripped that of MI classification. The proposed methods showed comparable results with other CNN-based approaches; however, they notably outperformed others when applied to a combined dataset. Compared to individual paradigm classifications, the performance of comparison group declined by up to 20%, while the proposed methods maintained consistent results with the exception of VGG16. ERA-CNN [43] achieved top-ranking performance among the comparison group albeit by a small margin but was still 13% less effective than the proposed methods. The independent samples t-test was conducted on the combined dataset and VGG16 only showed a p-value greater than 0.05 against ERA-CNN, MCNN and EEG Conformer, while the rest of the proposed methods showed statistically significant performance differences.

B. Calibration Evaluation

Table III presents the results on calibration dataset. Remarkably, the proposed methods not only achieved superior performance on the calibration dataset but also demonstrated consistent performance improvement as sessions progressed. This suggests that the proposed methods continued to accumulate knowledge from each subject throughout the session progression. They exhibited performance enhancements that appeared to be personalized to individual subjects. Conversely, other techniques displayed inconsistent performances irrespective of session progression. CSP-based methods showed approximately 19%, revealing insufficient capacity. Among CNN-based methods, MCNN [20] recorded the highest performance and least degradation; it was designed specifically to expand input latent space. In comparison with baseline evaluations, ERA-CNN's calibration performance was found to be less efficient and inferior to other CNN-based methods - potentially due to underfitting considering its three CNN modules and limited calibration data. In this evaluation, EEG-Transformer and Conformer [44] achieved marginally superior performance among comparison group. The proposed methods reported statistically significant differences compared to the other methods (p-values are less than 0.05).

C. Test Evaluation

The results are depicted in Fig. 3. The test was conducted before knowledge distillation of the proposed methods. Calibrated weights of each method were used for test. Notably, our proposed methods exhibited superior performance across most sessions and demonstrated a consistent upward trend in performance as sessions progressed. In particular, the proposed method^{t} significantly outperformed comparative methods. Other methods showed fluctuating performances depending on the session at hand. Excluding CSP-based methods, other techniques recorded comparable performances at the initial session with MCNN marginally outperforming others. As data accumulated over subsequent sessions, an improvement in performance was observed for the proposed methods aligning with an upward trend; however, such a pattern was absent within the comparison group irrespective of data accumulation.

The scenario for the test evaluation consisted of picking up an object and bringing it to the user by performing the 6-class classification. The results demonstrate the possibility of controlling a robotic arm in real life by combining MI and SI. This can be utilized for the rehabilitation of patients, such as those with stroke, because these paradigms are based on the principle of activating corresponding brain areas by having the user imagine specific actions or speech. This can be used for

TABLE III
RESULTS OF CALIBRATION BETWEEN THE PROPOSED AND COMPARISONS METHODS. PARENTHESES "()" DENOTES STANDARD DEVIATION. SUPERSCRIPTS ^{v}, ^{t} AND ^{s} DENOTE VGG16, TINY AND SMALL DEIT. THE HIGHEST PERFORMANCES ARE DENOTED IN BOLD. CHANCE LEVEL IS 16.67

Session #	1	2	3	4	5	6	7	8	9	10	Average	p-value
CSP+RF [45]	21.61	18.86	20.71	21.42	20.25	19.99	17.74	17.36	18.08	19.96	19.60	< 0.05
CSP+SVM [18]	17.74	17.14	18.97	18.06	20.90	19.47	17.59	20.29	19.64	20.11	18.99	< 0.05
CSP+LDA [18]	17.56	17.23	22.07	18.37	18.01	20.03	19.00	19.63	18.24	17.07	18.72	< 0.05
FBCSP [18]	22.42	14.10	21.02	16.68	17.59	21.92	17.80	22.12	17.81	19.35	19.08	< 0.05
EEGNet [19]	41.61	40.26	38.26	42.88	39.70	41.80	45.72	49.09	46.85	38.33	42.45	< 0.05
Deep ConvNet [17]	43.79	36.15	39.96	37.97	40.49	39.81	43.35	47.82	47.73	39.44	41.65	< 0.05
Shallow ConvNet [17]	39.91	44.49	39.85	30.69	34.84	42.18	49.54	47.90	41.78	46.73	41.79	< 0.05
ERA-CNN [43]	40.83	47.50	42.64	37.11	39.08	39.53	42.66	47.16	38.08	35.68	40.63	< 0.05
MCNN [20]	46.49	36.33	49.17	44.52	46.42	35.66	43.34	47.38	46.54	49.15	44.50	< 0.05
EEG-Transformer [44]	46.36	50.04	47.43	42.01	47.93	46.97	43.95	35.50	39.13	47.53	44.68	< 0.05
EEG Conformer [22]	48.94	46.73	50.90	50.11	52.73	44.01	38.77	46.16	45.21	44.32	46.79	< 0.05
Proposed Method^{v}	49.75	50.63	51.27	52.43	54.89	56.52	58.38	60.97	61.01	63.18	55.90	-
Proposed Method^{t}	**52.85**	**53.08**	**54.21**	**58.93**	**59.60**	**60.57**	**62.51**	**64.71**	**66.82**	**67.15**	**60.04**	-
Proposed Method^{s}	50.15	51.33	51.42	53.03	56.77	58.72	61.42	62.84	64.63	66.26	57.66	-

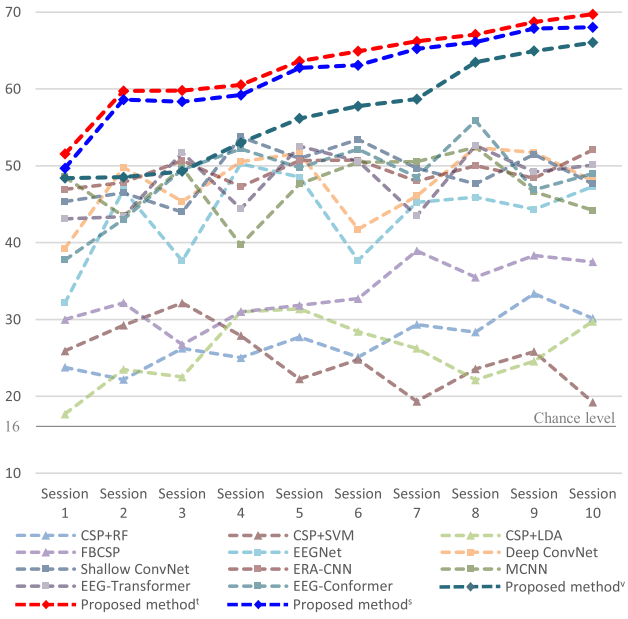

Fig. 3. Results of test evaluation. Superscripts ^{v}, ^{t} and ^{s} denote VGG16, tiny and small DeiT. The proposed methods outperformed the comparative methods in all sessions, with the exception of the initial sessions involving the VGG16.

training the brain activity patterns and characteristics of stroke patients and it can help recover lost functions by activating related areas of the brain through imagination alone, even if the patient cannot actually move or speak [49]. For example, a patient who has lost the ability to move a certain part of the body or the ability to speak due to a stroke can activate the corresponding neural circuits in the brain by imagining movement or speech, which can help regain motor skills. Thus, imagery-based BCIs can be very effective tools in the rehabilitation of stroke patients [50], [51], [52], [53].

D. Ablation Study

Several studies were conducted to explain the effect of individual design choices and to offer a more comprehensive understanding through in-depth analysis.

TABLE IV
RESULTS OF WITH AND WITHOUT EACH FEATURE EXTRACTOR

	Only f_{cs}	W/o f_{cs}	W/o f_{iv}	W/o f_{ps}	Report.
VGG16	49.85	19.26	51.04	52.34	**57.99**
Tiny	51.42	21.13	54.88	56.02	**67.74**
Small	52.62	20.64	56.6	59.23	**67.05**

TABLE V
PERFORMANCE COMPARISON BETWEEN WITH AND WITHOUT DATA AUGMENTATION TECHNIQUES

Methods	With augment.	W/o augment.	Diff.
VGG16	57.99	53.85	4.14
Tiny	67.74	58.12	9.62
Small	67.05	56.48	10.57
ERA-CNN	52.56	52.07	0.49

1) Effect of Feature Extractors: We hypothesize that a classifier, when supplied with a more expansive input latent space, can be trained more effectively for superior classification performance. To validate this, we conducted ablation studies on the baseline dataset by selectively excluding certain feature extractors while preserving the size of the feature input through zero-padding features where necessary. As demonstrated in Table IV, the lowest accuracy was observed when f_{cs} was omitted; however, when other feature extractors were added to f_{cs}, the classification performance improved compared to using f_{cs} alone. In other words, while f_{cs} contributes the most to performance, it yields the best results when used in conjunction with other feature extractors. This confirms that expanding the input latent space contributes to improving classification performance.

Fig. 4 visualizes features by reducing them to a lower dimension using t-distributed stochastic neighbor embedding (t-SNE) [54]. Interestingly, even though z_{ps} formed more distinct clusters per class than z_{iv} as depicted in Fig. 4, it was determined that z_{iv} had a greater contribution towards enhancing performance according to Table IV. In addition, while it seemed that z_{ps} formed clearer clusters than z_{cs},

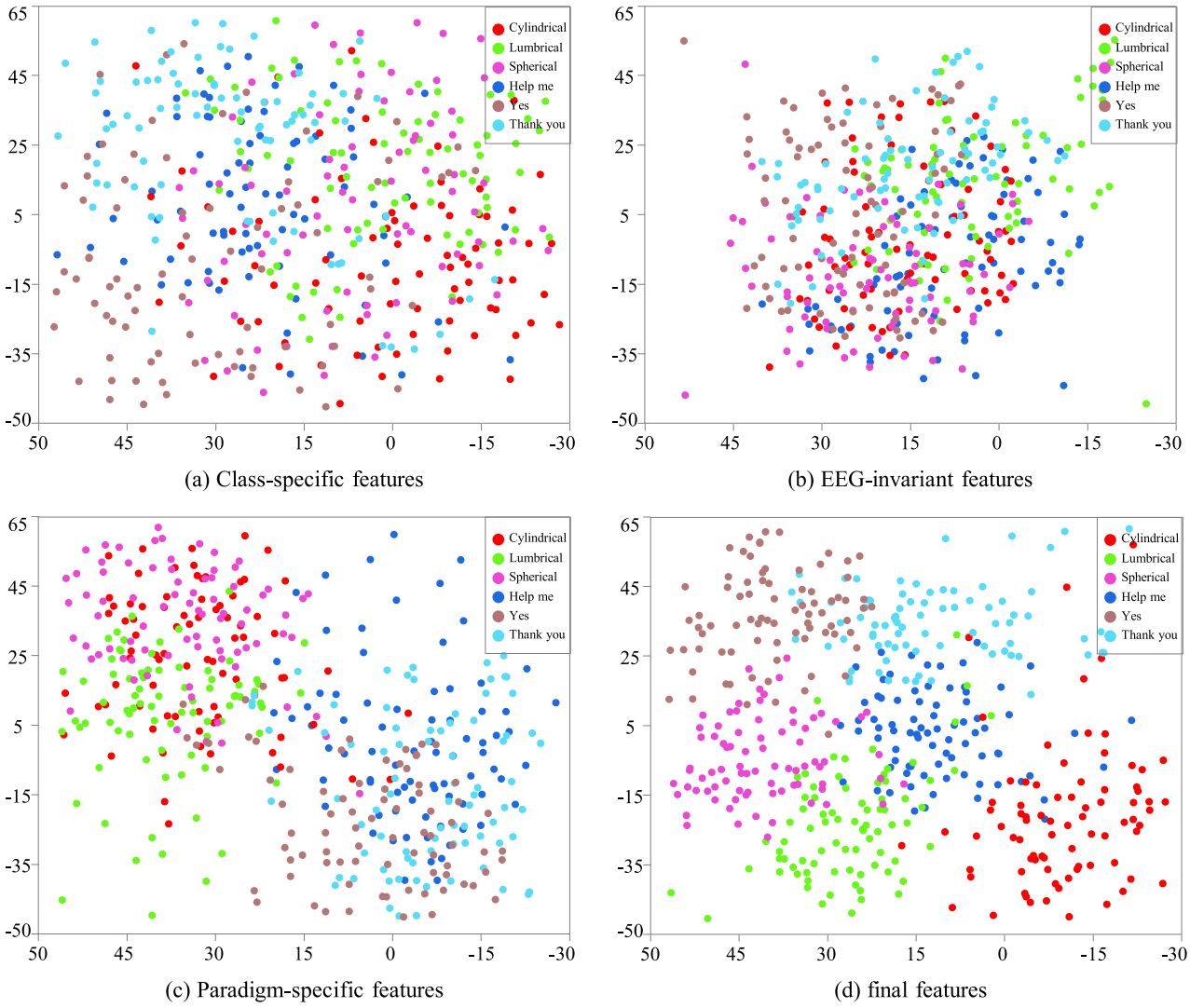

Fig. 4. t-SNE visualization of features. (a) class-specific z_{cs}, (b) EEG-invariant z_{iv}, (c) paradigm-specific z_{ps} and (d) final features of baseline training through t-SNE. Compared to (a), (b) seems to form a single cluster rather than class-wise clusters (i.e. shared or invariant features of all EEG signals). On the other hand, two paradigms are clearly distinguished in (c). Final features of DeiT form distinct class-wise clusters more pronounced than those in (a).

this did not directly affect classification outcomes. It can be deduced that since the objective function of z_{ps} is oriented towards paradigm classification rather than class-specific distinction, it has less relevance to classification tasks. However, integrating all three features, z_{cs}, z_{iv} and z_{ps}, generated more vital information compared to using only z_{cs}, as depicted in Fig. 4(d).

Moreover, we observed that while simple data augmentation techniques can decrease each feature extractor's performance, they improve overall classifier performance as shown in Table VI. Despite the increased loss of f_{iv} and performance decrease in the other two feature extractors with augmented data, the performance of f_θ improved. These results suggest that utilizing feature input could serve as an approach to mitigate constraints imposed by unclear ground truth of EEG signals.

2) Effect of Data Augmentation: We combined data from different paradigm datasets and their included subjects, despite the potential for large distribution differences among the data. Additionally, we introduced data augmentation techniques to ensure that state-of-the-art machine learning methods could be sufficiently trained. For this purpose, we introduced feature input. One advantage of using feature input is that it allows EEG signals to be treated like images. We employed data augmentation techniques such as adding noise and flipping over to dataset. Given the characteristics of EEG signals, using raw augmented data for training would likely be inefficient [55]. However, it's crucial to verify the impact of data augmentation techniques on feature inputs. An ablation study was conducted comparing performance with and without the application of noise addition and flipping over techniques. ERA-CNN was chosen to evaluate the effect of raw augmented data because it exhibited superior performance among comparison methods on the baseline dataset. The results are presented in Table V. Data augmentation techniques resulted in performance improvements of 4.14%, 9.62% and 10.57% with

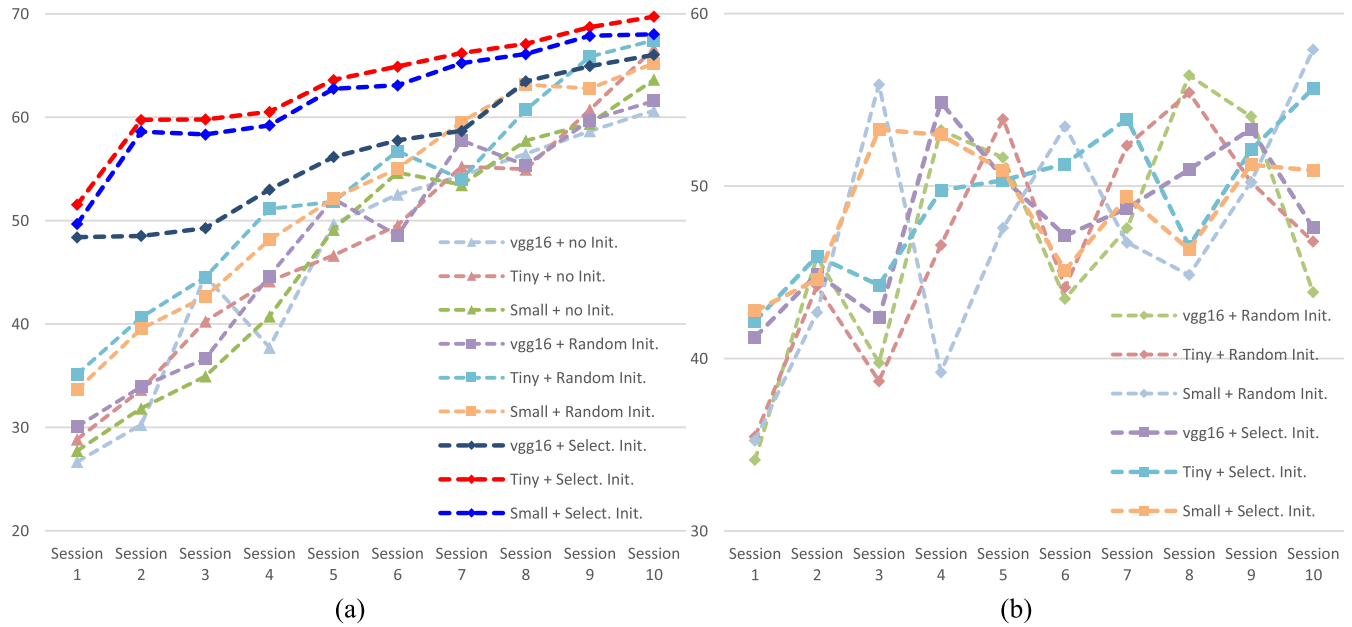

Fig. 5. Performance comparison between with and without network expansion and selective initialization. (a) network expansion with random and selective initialization and expansion only, (b) random and selective initialization without network expansion. Selective initialization results in superior performance in the initial session compared to other methods in both (a) and (b). Random initialization exhibits a more pronounced performance increase than no initialization. In the absence of network expansion, performance is inconsistent, mirroring trends observed in other comparison groups.

TABLE VI
EFFECTS OF DATA AUGMENTATION TECHNIQUES ON INDIVIDUAL FEATURE EXTRACTORS WHEN USED WITH TINY DEIT

Methods	With augment.	W/o augment.	Diff.
f_{iv}	3.34	1.10	2.24
f_{ps}	86.57	90.21	-3.64
f_{cs}	48.59	51.73	-3.14
f_{θ}	67.74	58.12	9.62

the proposed methods. However, ERA-CNN only showed a slight performance improvement (0.49%). Despite an increase in data volume, this negligible difference in ERA-CNN's classification performance suggests that directly using simply modulated EEG signals for training may not be effective. These results demonstrate that if data augmentation techniques are used in conjunction with feature input, state-of-the-art machine learning methods can be effectively utilized even under conditions where the data distribution is large and the quality is inconsistent and when there is a small amount of data.

3) Network Expansion and Selective Initialization: Even with meticulous calibration steps, the calibration performance of the network can be degraded due to large data distributions or intricate characteristics as shown in Table III. To mitigate this, a network capacity expansion technique was introduced. We designed a network that only expands f_{cs} to increase its capacity while maintaining low computational complexity. Experimental results demonstrate that network expansion leads to the best and progressively improved performance in both calibration and test. Therefore, this demonstrates that our proposed method enables reliable EEG-based robotic arm control by facilitating session-by-session calibration. Table VII presents the calibration performance of the proposed methods without network expansion. They show relatively consistent performance but no gradual improvement. It is confirmed that selective initialization prevents severe performance degradation and network expansion contributes to the improvement of classification performance by increasing the network capacity during the calibration. Finally, this demonstrates that network expansion can effectively calibrate data with varying distributions across sessions. Despite potentially higher computational costs, it suggests that network expansion could be a feasible strategy for performance improvement with appropriate cost-reduction measures.

Additional experiments were conducted on the calibration dataset to assess the contribution of both network expansion and selective initialization. Fig. 5 illustrates the classification performance of network with different settings of expansion and initialization. As per Fig 5(a), gradual performance improvement was observed with network expansion. Selective initialization yielded higher initial session performance compared to other settings. While random and no initialization also led to gradual improvements, recording a steep rise over selective initialization; however, they displayed inconsistent performance trends with several sessions experiencing drops in efficiency. In contrast, selective initialization showed steadily improving performances. Without employing network expansion, the proposed methods exhibited similar trends as comparison group as shown in Fig 5(b). Nevertheless, selective initialization results in smaller fluctuations in performance compared to random initialization. Selective initialization recorded higher initial session performances than random initialization did. Experimental results affirm that both selective initiation and network expansion are critical considerations for effective network calibration.

TABLE VII
CALIBRATION PERFORMANCE WITHOUT NETWORK EXPANSION. TINY AND SMALL DENOTE THE SIZE OF DEIT

Session #	1	2	3	4	5	6	7	8	9	10	Average	**Reported**
VGG16	43.47	45.41	44.19	46.23	48.48	50.64	51.73	53.06	52.29	55.97	49.15	**55.90**
Tiny	46.84	47.01	46.70	47.66	49.14	52.38	52.83	53.97	54.52	55.20	50.63	**60.04**
Small	45.32	48.22	49.33	49.42	51.29	50.26	52.39	53.22	53.12	50.84	50.34	**57.66**

TABLE VIII
THE BEST ACCURACY ACCORDING TO THE TRAINING EPOCH ON THE BASELINE DATASET

Methods	∼50	∼100	∼200	∼300	∼400	∼500	∼600
Tiny	48.83	57.11	61.95	64.73	66.36	67.74	65.15
Small	46.41	54.93	58.12	60.59	63.18	67.05	66.81

4) Performance Difference Between Tiny and Small DeiT: In all experiments, tiny DeiT performed better than small DeiT, especially in test, achieving a more robust performance than other methods. Both networks have the same number of layers, but small DeiT is a larger network in terms of the number of heads and embedding dimension. It is known that large networks are more easily optimized without decreasing generalization performance [56]. Given this, it was expected that the small DeiT would converge faster [57], [58], but this was not the case according to Table VIII. Despite applying data augmentation techniques, there is a possibility that underfitting occurred because the small DeiT has four times more parameters than the tiny DeiT. While the tiny DeiT showed convergence from the 400 epoch, the small DeiT only demonstrated its best performance upon reaching 500 epochs. By the point of 600 epochs, both models started to show signs of performance degradation; thus even if training were to continue further, there is little room for improvement in small DeiT's performance. Contrary to our expectations, this suggests that the amount of data was not sufficient to train small DeiT effectively. Despite these findings, further investigation into this phenomenon is necessary.

VI. CONCLUSION

Our study presents a robust method that consistently improves classification performance across multiple sessions, pioneers the combination of two intuitive endogenous BMI paradigms for intuitive robotic arm control and confirms through ablation studies the efficiency of this approach for practical EEG-based robotic arm control using advanced machine learning methods. To achieve this, we propose the utilization of feature input and network expansion techniques to obtain well-optimized network weights capable of accommodating variations in EEG patterns and new users. The feature input expands the latent space of classifiers, effectively maximizing the benefits of data augmentation techniques to overcome the inherent instability of EEG signals. Furthermore, network expansion enables upward performance trends with faster convergence. In all conducted experiments, the proposed methods consistently outperform comparison group, particularly during the test, demonstrating consistent performance across multiple sessions. However, it is crucial to investigate the underlying reasons for observed performance differences based on network size. Considering the amount of available calibration data, it remains uncertain whether the small DeiT model has been adequately trained. Therefore, one aspect of our future work will involve exploring the effects of network size and overparameterization in order to develop an iterative calibration pipeline that ensures stable performance. Additionally, EEG signals contain valuable high-level cognitive information pertaining to control speed, force and trajectory associated with robotic arm movements. However, current technologies predominantly focus on classifying EEG signals and subsequently controlling robotic arms based on those classifications. Consequently, one area we intend to explore in future research is interpreting higher-level cognitive information derived from EEG signals.

ACKNOWLEDGMENT

The authors thanks to M.-S. Oh and D.-K. Han for their help with the dataset construction and discussion of the data analysis.

REFERENCES

[1] A. Akce, M. Johnson, O. Dantsker, and T. Bretl, "A brain–machine iInterface to navigate a mobile robot in a planar workspace: Enabling humans to fly simulated aircraft with EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 2, pp. 306–318, Sep. 2012.

[2] E. Formaggio, S. Masiero, A. Bosco, F. Izzi, F. Piccione, and A. D. Felice, "Quantitative EEG evaluation during robot-assisted foot movement," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 9, pp. 1633–1640, Sep. 2017.

[3] A. Ravi, J. Lu, S. Pearce, and N. Jiang, "Enhanced system robustness of asynchronous BCI in augmented reality using steady-state motion visual evoked potential," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 85–95, 2022.

[4] X. Mao, W. Li, C. Lei, J. Jin, F. Duan, and S. Chen, "A brain–robot interaction system by fusing human and machine intelligence," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 533–542, Mar. 2019.

[5] C. I. Penaloza and S. Nishio, "BMI control of a third arm for multitasking," *Sci. Robot.*, vol. 3, no. 20, Oct. 2018, Art. no. eaat1228.

[6] S.-H. Hsu, T. R. Mullen, T.-P. Jung, and G. Cauwenberghs, "Real-time adaptive EEG source separation using online recursive independent component analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 3, pp. 309–319, Mar. 2016.

[7] M. Vaidya et al., "Hemicraniectomy in traumatic brain injury: A noninvasive platform to investigate high gamma activity for brain machine interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 7, pp. 1467–1472, Jul. 2019.

[8] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw, "Spatial filter selection for EEG-based communication," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 3, pp. 386–394, Sep. 1997.

[9] C. S. DaSalla, H. Kambara, M. Sato, and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns," *Neural Netw.*, vol. 22, no. 9, pp. 1334–1339, 2009.

[10] B. Min et al., "Vowel imagery decoding toward silent speech BCI using extreme learning machine with electroencephalogram," *Biomed Res. Int.*, vol. 2016, 2016.

[11] P. Gaur, K. McCreadie, R. B. Pachori, H. Wang, and G. Prasad, "Tangent space features-based transfer learning classification model for two-class motor imagery brain–computer interface," *Int. J. Neural Syst.*, vol. 29, no. 10, 2019, Art. no. 1950025.

[12] N. Görnitz et al., "When brain and behavior disagree: A novel ML approach for handling systematic label noise in EEG data," in *Proc. Mach. Learn. Interpret. Neuroimaging Workshop*, 2013, pp. 1–23.

[13] 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, pp. 1–15, 2020.

[14] S. Fazli, F. Popescu, M. Danóczy, B. Blankertz, K.-R. Müller, and C. Grozea, "Subject-independent mental state classification in single trials," *Neural Netw.*, vol. 22, no. 9, pp. 1305–1312, 2009.

[15] M. Rodríguez-Ugarte, E. Iáñez, M. Ortíz, and J. M. Azorín, "Personalized offline and Pseudo-online BCI models to detect pedaling intent," *Frontiers Neuroinform.*, vol. 11, p. 45, Aug. 2017.

[16] S. M. S. Hasan, M. R. Siddiquee, and O. Bai, "Asynchronous prediction of human gait intention in a pseudo online paradigm using wavelet transform," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 7, pp. 1623–1635, Jul. 2020.

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

[18] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain–computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jun. 2008, pp. 2390–2397.

[19] 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, 2018, Art. no. 056013.

[20] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Bencherif, and M. S. Hossain, "Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification," *IEEE Access*, vol. 7, pp. 18940–18950, 2019.

[21] J. Xie et al., "A transformer-based approach combining deep learning network and spatial–temporal information for raw EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2126–2136, 2022.

[22] Y. Song, Q. Zheng, B. Liu, and X. Gao, "EEG conformer: Convolutional transformer for EEG decoding and visualization," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 710–719, 2023.

[23] W.-K. Tam, K.-Y. Tong, F. Meng, and S. Gao, "A minimal set of electrodes for motor imagery BCI to control an assistive device in chronic stroke subjects: A multi-session study," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 6, pp. 617–627, Dec. 2011.

[24] M.-H. Lee et al., "EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy," *GigaScience*, vol. 8, no. 5, 2019, Art. no. giz002.

[25] E. Thomas, M. Dyson, and M. Clerc, "An analysis of performance evaluation for motor-imagery based BCI," *J. Neural Eng.*, vol. 10, no. 3, 2013, Art. no. 031001.

[26] S. Fazli, C. Grozea, M. Danóczy, B. Blankertz, F. Popescu, and K.-R. Müller, "Subject independent EEG-based BCI decoding," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 22, 2009, pp. 1–25.

[27] J.-H. Jeong et al., "International brain–computer interface competition: A review," *Frontiers Hum. Neurosci*, vol. 16, no. 1, 2020, Art. no. 898300.

[28] J.-H. Jeong, J.-H. Cho, B.-H. Lee, and S.-W. Lee, "Real-time deep neurolinguistic learning enhances noninvasive neural language decoding for brain–machine interaction," *IEEE Trans. Cybern.*, vol. 1, no. 1, pp. 1–14, Oct. 2022.

[29] S.-H. Lee, M. Lee, and S.-W. Lee, "Neural decoding of imagined speech and visual imagery as intuitive paradigms for BCI communication," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 2647–2659, Oct. 2020.

[30] T. Gasser, P. Ziegler, and W. F. Gattaz, "The deleterious effect of ocular artefacts on the quantitative EEG, and a remedy," *Eur. Arch. Psychiatry Clin. Neurosci.*, vol. 241, pp. 352–356, Dec. 1992.

[31] S. Romero, M. A. Mañanas, and M. J. Barbanoj, "A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: A simulation case," *Comput. Biol. Med.*, vol. 38, no. 3, pp. 348–360, 2008.

[32] J. A. Urigüen and B. Garcia-Zapirain, "EEG artifact removal—State-of-the-art and guidelines," *J. Neural Eng.*, vol. 12, no. 3, 2015, Art. no. 031001.

[33] A. C. Grant et al., "EEG interpretation reliability and interpreter confidence: A large single-center study," *Epilepsy Behav.*, vol. 32, pp. 102–107, Nov. 2014.

[34] Y.-E. Lee, N.-S. Kwak, and S.-W. Lee, "A real-time movement artifact removal method for ambulatory brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2660–2670, Dec. 2020.

[35] B. Blankertz, "The Berlin brain–computer interface: Non-medical uses of BCI technology," *Frontiers Neurosci.*, vol. 4, p. 198, Sep. 2010.

[36] B.-H. Lee, J.-H. Cho, B.-H. Kwon, and S.-W. Lee, "Factorization approach for sparse spatio-temporal brain–computer interface," in *Proc. 26th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2022, pp. 1090–1097.

[37] J. Yoon, E. Yang, J. Lee, and S. J. Hwang, "Lifelong learning with dynamically expandable networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–18.

[38] W. Wen, C. Wu, Y. Wang, Y. Chen, and H. Li, "Learning structured sparsity in deep neural networks," in *Proc. Adv. Neural Inf. Process Syst.*, vol. 29, 2016, pp. 1–36.

[39] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," in *Proc. Adv. Neural Inf. Process Syst.*, 2014, pp. 1–4.

[40] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2021, pp. 10347–10357.

[41] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–24.

[42] Z. J. Koles, M. S. Lazar, and S. Z. Zhou, "Spatial patterns underlying population differences in the background EEG," *Brain Topogr.*, vol. 2, pp. 275–284, Aug. 1990.

[43] B.-H. Lee, J.-H. Jeong, K.-H. Shim, and S.-W. Lee, "Classification of high-dimensional motor imagery tasks based on an end-to-end role assigned convolutional neural network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1359–1363.

[44] Y. Song, X. Jia, L. Yang, and L. Xie, "Transformer-based spatial–temporal feature learning for EEG decoding," 2021, *arXiv:2106.11170*.

[45] Y. Qi, *Random Forest for Bioinformatics*. Boston, MA, USA: Springer, 2012, pp. 307–323.

[46] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–20.

[47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–12.

[48] Y. Yao, L. Rosasco, and A. Caponnetto, "On early stopping in gradient descent learning," *Constr. Approx.*, vol. 26, no. 2, pp. 289–315, 2007.

[49] N. Sharma, V. M. Pomeroy, and J.-C. Baron, "Motor imagery: A backdoor to the motor system after stroke?" *Stroke*, vol. 37, no. 7, pp. 1941–1952, 2006.

[50] S. J. Page, P. Levine, and A. Leonard, "Mental practice in chronic stroke: Results of a randomized, placebo-controlled trial," *Stroke*, vol. 38, no. 4, pp. 1293–1297, 2007.

[51] M. Jeannerod, *Motor Cognition: What Actions Tell the Self*, vol. 42. New York, NY, USA: Oxford, 2006.

[52] K. J. Miller, G. Schalk, E. E. Fetz, M. D. Nijs, J. G. Ojemann, and R. P. Rao, "Cortical activity during motor execution, motor imagery, and imagery-based online feedback," *PNAS*, vol. 107, no. 9, pp. 4430–4435, 2010.

[53] C. Neuper, A. Schlögl, and G. Pfurtscheller, "Enhancement of left-right sensorimotor EEG differences during feedback-regulated motor imagery," *J. Clin. Neurophysiol.*, vol. 16, no. 4, pp. 373–382, 1999.

[54] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–32, 2008.

[55] F. Fahimi, S. Dosen, K. K. Ang, N. Mrachacz-Kersting, and C. Guan, "Generative adversarial networks-based data augmentation for brain–computer interface," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 9, pp. 4039–4051, Sep. 2021.

[56] R.-D. Buhai, Y. Halpern, Y. Kim, A. Risteski, and D. Sontag, "Empirical study of the benefits of overparameterization in learning latent variable models," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1211–1219.

[57] Z. Allen-Zhu, Y. Li, and Y. Liang, "Learning and generalization in overparameterized neural networks, going beyond two layers," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–26.

[58] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," *Commun. ACM*, vol. 64, no. 3, pp. 107–115, 2021.