This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TETCI.2025.3619564, IEEE TTransactions on Emerging Topics in Computational Intelligence

Multi-Scale Shapley Adaptation Pruning: Realizing Backdoor Defense in Brain-Computer Interface with Shapley-Value-based Neural Network Pruning

Fumin Li, Rui Yang, *Senior Member, IEEE*, Hanjing Cheng, Mengjie Huang, *Member, IEEE*, Fanglue Zhang, Fuad E. Alsaadi, Zidong Wang, *Fellow, IEEE*

***Abstract*—In the recent years, researchers made significant progress in electroencephalogram (EEG) classification tasks using deep neural networks, especially in brain-computer interface (BCI) systems. BCI systems rely on EEG signals for effective human-computer interaction, and deep neural networks have shown excellent performance in processing EEG signals. However, backdoor attack have a significant impact on the security of EEG-based BCI systems. In this paper, a novel multi-scale Shapley adaptation pruning (MSAP) method is proposed to solve the security problem caused by backdoor attack. In the proposed MSAP, the multi-scale Shapley segmented mapping method is used to accurately locate the backdoor weights. Subsequently, the cost function is utilized to adaptively prune the backdoor weights to ensure normal classification. Ultimately, the validity of the experiments is verified on the BCI competition public datasets (BCI-III-IVb, BCI-III-IVa, and BCI-IV-1a). The results show that the proposed MSAP method outperforms other pruning methods in defending EEG-based BCI systems against backdoor attack, maintaining a high baseline classification accuracy while reducing the attack success rate.**

This research has been approved by University Ethics Committee of Xi'an Jiaotong-Liverpool University. This project was funded by the National Natural Science Foundation of China (72401233), the Jiangsu Provincial Scientific Research Center of Applied Mathematics (BK20233002), the Jiangsu Provincial Qinglan Project, the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (23KJB520038), and the Research Enhancement Fund of XJTLU (REF-23-01-008). This project was also funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia, under grant no. (GPIP: 72-135-2024). The authors, therefore, acknowledge with thanks DSR for technical and financial support.

F. Li is with School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China, and also with Fudan Institute on Aging and Ministry of Education Laboratory for National Development and Intelligent Governance, Fudan University, Shanghai, 200433, China (e-mail: Fumin.Li22@alumni.xjtlu.edu.cn);

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

H. Cheng is with School of Electronic & Information Engineering, Suzhou University of Science and Technology, Suzhou, 215009, China (email: chj@mail.usts.edu.cn);

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

F. Zhang is with School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China, and also with School of Electrical Engineering, Electronics and Computer Science, University of Liverpool, Liverpool, L69 3BX, United Kingdom (e-mail: Fanglue.Zhang22@alumni.xjtlu.edu.cn);

Fuad E. Alsaadi is with the Communication Systems and Networks Research Group, Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: falsaadi@kau.edu.sa; tel: +966 6952183);

Z. Wang is with the Department of Computer Science, Brunel University London, Uxbridge, Middlesex, UB8 3PH, United Kingdom (e-mail: Zidong.Wang@brunel.ac.uk).

Corresponding authors: R. Yang.

***Index Terms*—Electroencephalogram, brain-computer interface, backdoor attack, Shapley value.**

I. Introduction

Brain-computer interface (BCI) systems establish an interaction bridge between users and external devices [6], [32], [43]. In recent years, several types of BCI systems have been invented, such as invasive, semi-invasive, and non-invasive [58]. Electroencephalogram (EEG), as a prevalent input method in BCI systems [57], efficiently extracts brain electrical signals [45], assisting researchers in comprehending the activities within distinct brain regions [1], [17]. To facilitate systematic research into brain region features, various deep neural networks (DNNs) have been developed for EEG classification in BCI systems [23], [56], [66]. Although the utilization of DNNs in BCI systems has favorable outcomes [29], [33], [62], [71], the black-box characteristic of DNNs poses a significant concern: security [65], [70].

Recent research suggests that EEG-based BCI systems are vulnerable and susceptible to malicious manipulation by attackers [60], [67]. A novel attack method known as a backdoor attack has been applied to the classification of EEG signals, leading to adverse effects [8], [14], [50]. The backdoor attack involves injecting data with specific triggers into the training set to create an infected model [31]. Furthermore, attackers have the capability to manipulate the infected model's behavior on specific data using triggers [18], yet this manipulation does not adversely affect the model's performance on other clean data [26]. When researchers employ BCI datasets that are mixed with trigger samples, the network becomes infected, yielding inaccurate classification results.

To mitigate the substantial security risks arising from backdoor attack, defense methods against such vulnerabilities have continuously evolved [2], [59]. Existing backdoor defense methods are predominantly categorized into two groups: one focuses on detecting the backdoor, while the other aims to eliminate the backdoor trigger [39]. The former mainly involves detecting the presence of the backdoor in the model and achieving defense by filtering the infected samples or refusing to deploy the infection model [11], [27]. This detection method can quickly identify whether a model has backdoor and is suitable for scenarios that require backdoor models to be filtered out promptly [36], [61]. Moreover, the detection method is extremely valuable for situations where the model

cannot be modified directly, such as with black-box models [16].

The latter method usually employs pruning of relevant backdoor neurons and retraining of the infection model to prevent backdoor behavior [63]. The pruning defense method can directly repair the model, enhancing its security after detecting the backdoor [24]. Besides, the backdoor effect on the model is eliminated basically, maintaining normal performance even in the presence of triggers [55]. Compared with detecting defense method, pruning defense methods are becoming widespread [19], as they not only effectively eliminate the influence of backdoor triggers, but also improve the model's robustness and computing speed and reduce complexity [13]. While substantial research has been conducted on pruning-related backdoor defense strategies, they are more prevalent in the field of image processing and speech [38], [47], [52], [69]. For BCI systems, systematic research on backdoor defense is still lacking and the challenge of backdoor defense is tied to the following two main issues:

(1) *EEG Signals Issue.* EEG signals are highly complex physiological signals that contain substantial amount of spatial and time-frequency information [37], captured through electrodes placed on an EEG cap [7]. The spatial characteristics of EEG signals arise from the fact that EEG activity varies from different regions, and multiple electrode channels are used to capture activity in different brain regions. Given the inherent spatial characteristics of EEG signals [49], attackers can exploit multiple electrode channels to hide backdoor triggers within the signals captured from different brain regions, increasing the difficulty of defense [34]. Additionally, EEG signals exhibit rich variations in the frequency domain, with different frequency bands associated with various cognitive states and brain activities (e.g., alpha, beta, delta, and theta waves). Due to the limited ability of the human eye to distinguish frequency domain signals, attackers can insert subtle frequency domain triggers to activate the backdoor. These characteristics make accurate backdoor defense for EEG signals particularly challenging.

(2) *Subject Issue.* EEG signals are significantly influenced by individual differences [42]. Variations in brain structure, thinking patterns, and levels of fatigue among subjects can lead to significant differences in the characteristics of the collected EEG signals, such as frequency bands and amplitudes [5], [12]. As a result, attackers can exploit these differences to embed backdoor triggers into the signal of specific individual, evading detection [4]. Additionally, a subject's emotional and cognitive states change over time, further increasing the variability of EEG signals. Moreover, intra-subject differences due to variation increase the challenge of model generalization capabilities, making it difficult for the defense system to simultaneously adapt to the signaling characteristics of different subjects. The above mentioned subject characteristics suggest that backdoor defense tasks may present poorer results when facing new subjects, greatly affecting the defense capability.

To incorporate the above issues in the field of BCI, a method with multi-scale Shapley adaptation pruning (MSAP) is proposed, consisting of multi-scale segmented Shapley mapping and adaptive backdoor weight cost pruning. Shapley value, a mathematical method for distributing the contributions of the various players in a cooperative game, is the critical component of multi-scale segmented Shapley mapping [53]. In the context of DNNs, the Shapley value accurately reflects the significance of each neuron's contribution to the classification by evaluating the contribution of the neuron in all possible permutations [21]. Therefore, the Shapley value can be used to identify neurons that are overreacting to specific malicious inputs (e.g., backdoor attack). In contrast to conventional methods based on local features (e.g., weights or activation values), the Shapley value evaluates the contribution of neurons from an objective perspective and avoids the inadvertent removal of neurons that are important to the model's tasks [46].

The multi-scale segmented mapping has been proposed to solve the inherent complexity problem of Shapley values via mapping global information to multidimensional information. Moreover, the backdoor weights can be obtained by dividing the neuron regions with different scales and calculating them with the initial network weights. Besides, an elaborate cost function is designed to precisely converge on the position of the backdoor weights through iterative loops [9], [51]. Ultimately, a clean neural network is obtained by setting a threshold to evaluate the current convergence. Building upon the above discussion, the main contributions of this paper can be summarized as follows:

(1) This study investigates the vulnerability of EEG to backdoor attack and provides valuable perspectives on the effectiveness of EEG-based BCI systems in backdoor defense;

(2) A novel approach known as multi-scale Shapley segmented mapping is used to optimize high complexity problems with Shapley value, and the contribution of backdoor weights can be efficiently estimated by computing the global information of the network through multi-scale mapping;

(3) A new cost function is designed to adaptively select backdoor weights for pruning while adjusting the network's structure, effectively preventing the degradation of the classification accuracy of clean data and reducing the attack success rate on infected data;

(4) The proposed novel MSAP method is compared and experimented on three publicly available BCI datasets to evaluate the effectiveness in EEG-based backdoor defense scenarios for BCI systems.

The remaining sections of the paper are illustrated as follows. Section II introduces the description and formulation of the challenges posed by backdoor attack in EEG classification. Section III provides the detailed description of the multi-scale Shapley segmented mapping and adaptive cost pruning that comprises the MSAP. Section IV provides a visualization of the results. Finally, section V summarizes the paper and discusses future work.

II. Problem Formulation

The successful execution of a backdoor attack in deep neural networks (DNNs) for EEG classification implies that the network has additional learning capacity to grasp the attacker's reverse-triggered behavior [15]. To illustrate the

attacker's backdoor attack process for EEG signal categorization, DNNs in EEG are defined. Given an EEG classification dataset $\mathcal{E} = \{x_i, y_i\}_{i=1}^{\mathcal{N}}$, where $x_i \in \mathcal{X} \subset \mathcal{R}^{c \times t}$ is the i-th EEG trial with c number of channels and t sampling time, $y_i \in \mathcal{Y} = \{1, ..., T\}$ is the task label (such as left hand and right foot) in MI, and $\mathcal{N}$ is the number of EEG trials. The DNNs include a feature extractor $\mathcal{F}$ and a task classifier $\mathcal{C}$, both of which can learn and classify tasks on EEG classification dataset thus obtaining a benign network $\mathcal{B}$. The benign network $\mathcal{B}$ can map EEG trials $\mathcal{X}$ into the space of task labels $\mathcal{Y}$, i.e., $\mathcal{B} : \mathcal{X} \to \mathcal{Y}$.

For the EEG classification dataset, the attacker's objective is to construct a malicious infection network $\mathcal{I}$ that will misclassify the EEG data E containing triggers with malicious label $y_m \in \mathcal{Y}$, which is intentionally specified by the attacker. The malicious triggers, denoted as k, encompass elements like slight time-frequency perturbations and are crucial for the backdoor attack. Thus, the infected EEG data can be represented as follows:

$$\mathcal{H}(E, k) = E + \Delta(k) \tag{1}$$

where the $\Delta(\cdot)$ represents the function generating the backdoor perturbation based on the trigger k and $\mathcal{H}(\cdot)$ represents the hybrid function that combines the perturbation generated by the backdoor trigger with data E. The process of an attacker exploiting the backdoor attack is illustrated in Fig. 1. The infected EEG with backdoor triggers is trained to establish the decision boundary for the infected network, causing the classification of normal labels to be altered to the classification result of malicious labels.

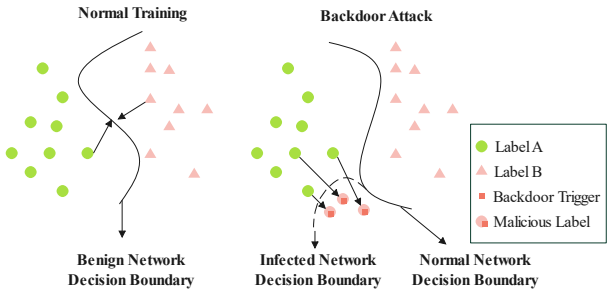


Fig. 1. Effect of backdoor attack on model classification boundary.

Based on the above description of the backdoor attack [68], the attacker's behavior can be defined by:

$$\begin{cases} \mathcal{I}(\mathcal{H}(E, k)) = y_m, \\ \mathcal{I}(E) = \mathcal{B}(E) = y_i, \end{cases} \quad \text{s.t. } E \in \mathcal{E}, |k| \leq m \tag{2}$$

where $\mathcal{I}$ represents the infected network, $\mathcal{B}$ represents the benign network, $\mathcal{E}$ is the EEG classification dataset, E is the EEG data, y_i is the original classification label, y_m and k are the malicious label and trigger created by the attacker, and m is the maximum number of links in the trigger. The complexity of EEG signals within the time-frequency and spatial domains, combined with the insertion of triggers k across channel number c and sampling time t, significantly enhances the stealthiness of backdoor attack. This paper investigates the security problem of backdoor attack in the EEG classification task and proposes a concrete approach to solve this problem.

III. Methodology

The overall structure of the proposed method MSAP is shown in Fig. 2. For the infected EEG (derived from the backdoor injection by the attacker), the training of the deep neural networks (DNNs) results in an infected network. The multi-scale segmented Shapley mapping takes the infected network as input and computes the backdoor weight matrix associated with the backdoor attack. Then, adaptive backdoor weight cost pruning is employed to refine the pruning process of the backdoor weight matrix, leading to the acquisition of the optimally pruned network. The following sections offer a comprehensive explanation of MSAP methodological components.

A. Shapley Value in Backdoor Defense

Shapley value, a concept in cooperative game theory, is devised to allocate contribution value to participants in a coalitional game, employing the concept of marginal contribution [30]. In the computation of the Shapley value, participants are collectively referred to as the set $D = \{d_i\}_{i=1}^{f}$, and the coalitional game is defined by a function that maps any subset S of participants to contribution value [3]. The value of a participant's contribution is computed using a score function $v(S) : \mathcal{P}(D) \to \mathcal{R}$, where $\mathcal{P}(D)$ represents the power set of D (the set consisting of all subsets of D). The marginal contribution represents the variation in the value of a subset S resulting from the inclusion of a participant e_i $(v(S \cup \{e_i\}) - v(S))$ [21].

In the realm of backdoor defense, it is imperative to accurately estimate the distribution of backdoor neurons and assess their contributions. Let $N = \{n_i\}_{i=1}^{q}$ denote the set of q neurons in the infected network, and the neuron n_i is considered as a participant to compute the value of the contribution. The score function $v(S_n)$ is computed based on the neuron subset $S_n \subseteq N$, using the attack success rate (ASR), which is a measure used to assess the effectiveness of attack in the infected network. Subsequently, to equitably assess the neuron's impact on the ASR, the marginal contribution of each neuron is computed and then averaged across all possible combinations of neuron subsets S_n to which n_i does not belong $(S_n \subseteq N \setminus \{n_i\})$.

The Shapley value ϕ for the backdoor defense can be expressed as:

$$\phi(n_i) = \sum_{S_n \subseteq N \setminus \{n_i\}} \overline{W}_n \cdot (v(S_n \cup \{n_i\}) - v(S_n)) \tag{3}$$

where $\overline{W}_n = \frac{|S_n|!(|N|-|S_n|-1)!}{|N|!}$ represents the weighted average of neurons (indicating the different weights used in computing each neuron's marginal contribution for various subsets S_n), reflecting the importance of each combination. Besides, the symbols $|S_n|$ and $|N|$ represent the cardinality of the sets S_n and N (denoting the number of elements within the set). Nevertheless, the computation of the Shapley value necessitates examining all combinatorial cases of neurons in the set N, resulting in a time complexity of $O(q!)$. To mitigate the extensive computation time, a multi-scale segmented Shapley mapping method is introduced as an optimization strategy.

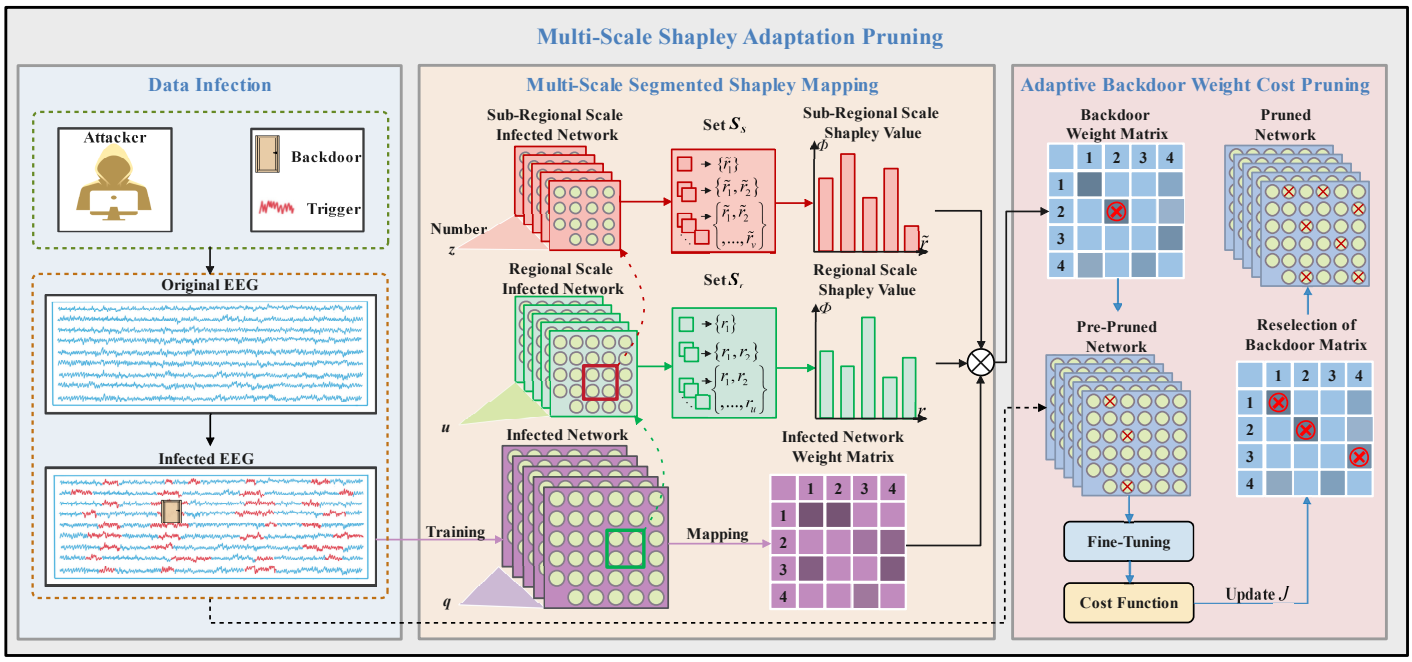


Fig. 2. The structure of the proposed multi-scale Shapley adaptation pruning (MSAP), where the pale blue block represents the input and the EEG data infection, the light yellow block represents the multi-scale segmented Shapley mapping, the pink block represents the adaptive backdoor weight cost pruning, and the red cross signs denote the pruning operations that prune top m backdoor weights at the network and matrix levels (the gradation of colors in weight matrix represents the magnitude of the weight value).

Remark 1: Given the complexity of the BCI system, with intricate interactions between EEG signals and neural network architectures, accurate pruning becomes even more important. Notably, using the Shapley value to assess neuron contributions in backdoor defense provides an accurate evaluation of each backdoor neuron's impact on ASR. By considering all possible permutations of neurons, the Shapley value effectively identifies the backdoor neurons with the prominent malicious effect. Therefore, backdoor defense becomes reliable by pruning of key neurons based on Shapley value in the BCI system.

B. Multi-Scale Segmented Shapley Mapping

The process of multi-scale segmented Shapley mapping is shown in Fig. 3, demonstrating the mapping segmentation rules and the computation process of Shapley value. By progressively scaling, the original infected network $\mathcal{I}$ is segmented into two distinct versions of the infected network. The Shapley values for neuron sets are computed at both regional and sub-regional scales of the infected network. Finally, the impact of the backdoor attack can be estimated by multiplying the two proportional information with the weight of the infected network.

1) Regional Segmented Shapley Mapping: The distribution of backdoor neurons at the regional scale is explored by regional segmented Shapley mapping, revealing the relationship between the regional scale of the infected network and the Shapley value. To obtain the regional scale infected network, $\hat{q}$ neurons in the infected network are collected to form a set of regional scale neurons $r = \{n_i\}_{i=1}^{\hat{q}}$. By repeating the above process u times, disjoint sets of regional scale neurons are obtained to form the regional scale infected network

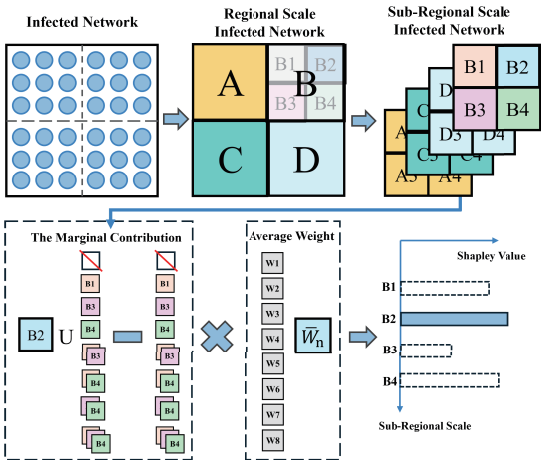


Fig. 3. The rule and computational procedure for multi-scale Shapley segmented mapping.

$R = \{r_i\}_{i=1}^{u} (u\hat{q} = q, u \ll q)$. Therefore, the regional scale Shapley value is denoted by:

$$\phi(r_i) = \sum_{S_r \subseteq R \backslash \{r_i\}} \overline{W}_r \cdot (v(S_r \cup \{r_i\}) - v(S_r)) \quad (4)$$

where $\overline{W}_r = \frac{|S_r|!(|R|-|S_r|-1)!}{|R|!}$ represents the average weight of the neuronal regions, indicating the significance of each neuronal region r across various regional subsets S_r.

2) Sub-Regional Segmented Shapley Mapping: To further explore the distribution information of backdoor neurons in the infected network, sub-regional segmented Shapley mapping is proposed. The objective is to deepen the scale of the regional

scale infected network, bringing the sub-regional scale neurons in closer alignment with the contribution values of individual neurons. The set of sub-regional scale neurons $\widetilde{r} = \{n_i\}_{i=1}^{\widetilde{q}} \subseteq r$ is aggregated by $\widetilde{q}$ neurons in the infected network. By repeating this process z times, disjoint sets of sub-regional scale neurons are derived, forming the sub-regional scale infected network $\widetilde{R} = \{\widetilde{r}_i\}_{i=1}^{z} (z\widetilde{q} = q, z \ll q)$.Consequently, the sub-regional scale Shapley value can be defined by:

$$\phi(\widetilde{r}_i) = \sum_{S_s \subseteq \widetilde{R} \setminus \{\widetilde{r}_i\}} \overline{W}_s \cdot (v(S_s \cup \{\widetilde{r}_i\}) - v(S_s)) \tag{5}$$

where $\overline{W}_s = \frac{|S_s|!(|\widetilde{R}|-|S_s|-1)!}{|\widetilde{R}|!}$ denotes the average weight of neurons at the sub-region scale, signifying the importance of $\widetilde{r}$ within any subset S_s of sub-region.

3) Acquisition of Backdoor Weight: Both regional and sub-regional scale mappings explore the distribution of backdoor neurons. However, aggregating neurons across these scales restricts a detailed understanding of each neuron's role. Since network weights determine the importance of information in neurons [54], the impact of backdoor attack can be assessed by multiplying the infected network $\mathcal{I}$'s weight matrix $\mathcal{W}$ with scalar Shapley value from multi-scale segmented Shapley mapping, as demonstrated by:

$$\mathcal{W}^{\phi} = \phi(r)\phi(\widetilde{r})\mathcal{W} \tag{6}$$

In the formula, $\mathcal{W}^{\phi} = [\omega_1^{\phi}, ..., \omega_j^{\phi}]^T$ is the backdoor weight matrix, where ω_j^{ϕ} is the backdoor weight that indicates the magnitude of the impact of the backdoor attack on the infected network. By pruning the top m backdoor weights in the infected network, the weights that significantly contribute to such attacks are effectively removed, thereby reducing the ASR in EEG tasks.

Remark 2: It is notable that multi-scale segmented Shapley mapping by dividing the infected network into regional scale and sub-regional scale. By providing a finer-grained view, the computation of Shapley value accurately identifies the contribution of backdoor neurons to the overall ASR. Furthermore, the definition of the backdoor weight matrix effectively quantifies the impact of each neuron on the backdoor attack, allowing for precise pruning to reduce the contribution to the ASR and significantly enhance defense performance.

C. Adaptive Backdoor Weight Cost Pruning

The backdoor weight matrix reveals the relationship between the weights of the infected network and the magnitude of their contributions to the backdoor attack. Direct pruning of associated backdoor weight information from the matrix can help reduce the risk of backdoor attack. However, extensive direct pruning of the backdoor weights leads to accuracy (ACC) degradation to categorize clean data for the infected network. To address the above issues, this paper presents adaptive backdoor weight cost pruning, as depicted in Fig. 4, with the primary objective of reducing the ASR in the infected network while preserving the ACC for clean data.

Adaptive backdoor weight cost pruning proceeds by preliminary pruning according to the magnitude of weights in the backdoor weight matrix. By pruning the top m backdoor weights in the infected network, a significant reduction in the ASR is achieved. Subsequently, the pre-pruned network undergoes fine-tuning to adjust the current network's weight distribution. Before and after fine-tuning, the parameters of the pre-pruned and fine-tuned network are input into a specially designed cost function. The suitability of the preliminary backdoor weights pruning is evaluated using the output of the cost function. Ultimately, the pruning positions and quantities of backdoor weights are reselected in a supervised manner based on the cost function output to achieve optimal selection. The following paragraphs thoroughly describe the method of adaptive backdoor weight cost pruning.

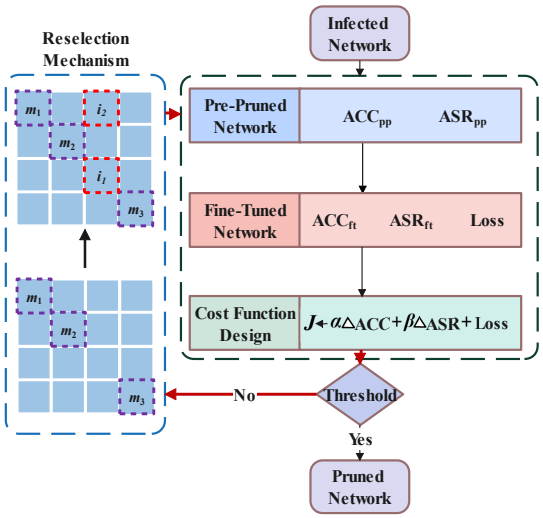


Fig. 4. The flowchart of adaptive backdoor weight cost pruning.

1) Backdoor Weight-Based Preliminary Pruning: In the infected network, the distribution information of backdoor weights impacts ASR and ACC. To minimize the adverse impact on ACC during the direct pruning of backdoor weights, a preliminary pruning strategy is proposed. Based on the backdoor weight matrix from the multi-scale segmented Shapley mapping, the weights are ordered by magnitude. The preliminary pruning removes the top m backdoor weights, followed by inputting validation data (containing both infected and clean data) to observe the ASR and ACC of the pre-pruned network $\mathcal{I}_p$.

2) Fine-Tuning: To mitigate the adverse impact on ACC caused by preliminary pruning, fine-tuning is implemented on the pre-pruned network to adjust the weight space structure. The network post-fine-tuning is designated as fine-tuned network $\mathcal{I}_f$. Simultaneously, the top m backdoor weights removed in the preliminary pruning process are frozen to ensure that the ASR does not experience substantial improvement. The optimization of weight spatial distribution in the fine-tuned network is defined by:

$$\omega_{\tilde{t}+1} = \omega_{\tilde{t}} - \eta \cdot \nabla \mathcal{L}(\omega_{\tilde{t}}) \tag{7}$$

where $\omega_{\tilde{t}}$ represents the weight of the fine-tuned network at time $\tilde{t}$, $\omega_{\tilde{t}+1}$ represents the weight of the fine-tuned network

updated at time $\tilde{t}$+1, η is the learning rate and $\nabla\mathcal{L}$ is the gradient.

During the fine-tuning process, the ASR increases slightly as the weights of the fine-tuned network are optimized, since the unpruned backdoor weights in the corresponding network are unintentionally strengthened [54]. Moreover, fine-tuning increases the complexity of the fine-tuned network, potentially obscuring backdoor weights and complicating their detection. Therefore, a specially designed cost function is established, aiming to reduce the ASR of the fine-tuned network.

3) Cost Function Design: As discussed in the previous section, even though the top m backdoor weights pruned preliminarily are frozen, changes in the fine-tuned network's weight distribution still lead to a slight increase in the ASR. However, the objective of backdoor defense is to achieve a pruned network with low ASR and high ACC. Therefore, based on updated pre-pruned network and fine-tuned network parameters (ASR, ACC, $\mathcal{L}$), this paper designs a new cost function during the fine-tuning process.

The specific cost function designed based on the fine-tuning updated parameters is denoted by:

$$\mathcal{J}_{best} = \arg\min(\mathcal{L} + \alpha\Delta ASR + \beta\Delta ACC) \tag{8}$$

where α and β are the hyperparameters satisfying α, $\beta \geq 0$ and $\alpha + \beta = 1$, which can be adjusted based on the importance of ΔASR, ΔACC. By adjusting two hyperparameters, a dynamic balance between ASR and ACC in the infected network is achieved. $\mathcal{L}$ represents the loss function of the fine-tuned network and Δ represents the amount of changes. The amount of changes in ASR and ACC are defined by:

$$\begin{cases} \Delta ASR = |ASR_{pp} - ASR_{ft}| \\ \Delta ACC = |ACC_{pp} - ACC_{ft}| \end{cases} \tag{9}$$

where ASR_{pp} and ACC_{pp} represent the ASR and ACC of the pre-pruned network $\mathcal{I}_p$ respectively, ASR_{ft} and ACC_{ft} represent the ASR and ACC of the fine-tuned network $\mathcal{I}_f$ respectively.

4) Reselection Mechanism: The value of the cost function $\mathcal{J}$ reflects the immediate effects of selecting backdoor weights for pruning on the pre-pruned network. To achieve the optimal value of the cost function, backdoor weights need to be reselected, dynamically increasing the weight number in each epoch. The specific mathematical expression for the reselection mechanism is defined by:

$$\tilde{\mathcal{W}} = \begin{cases} \mathcal{S}(\hat{\mathcal{W}}, m), & \text{if } J \leq \theta \\ \mathcal{S}(\hat{\mathcal{W}}, m+i), & \text{if } J > \theta \end{cases} \quad \text{s.t. } \tilde{W} \subseteq \hat{W} \tag{10}$$

where $\hat{\mathcal{W}} = \{\omega_1^\phi, \omega_2^\phi, \ldots, \omega_j^\phi\} \xleftarrow{\text{Mapping}} \mathcal{W}^\phi = [\omega_1^\phi, ..., \omega_j^\phi]^T$ represents the set obtained by mapping the backdoor weight matrix $\mathcal{W}^\phi$, and $\tilde{\mathcal{W}}$ represents the set of backdoor weights selected for pruning. $\mathcal{S}$ represents the selection function for the backdoor weights, where $m, i \in \mathbb{N}$ are used as inputs to select that number of backdoor weights pruned from the given $\hat{\mathcal{W}}$. Based on empirical analysis, a threshold θ is established to evaluate the cost function's output. Once J_{best} reaches the threshold, the network is identified as the pruned network P, and the current ASR and ACC with clean data are recorded.

The pseudocode of the proposed multi-scale Shapley adaptation pruning (MSAP) is shown in Algorithm 1.

Algorithm 1: Multi-Scale Shapley Adaptation Pruning

Input:
Clean EEG data: E
Infected EEG data: $\mathcal{H}(E, k)$
Infected network: $\mathcal{I}$
Output:
Optimal result of clean data: ACC
Optimal result of infected data: ASR

Initialize the parameters of the network
Obtain scale infected networks R and $\widetilde{R}$ from infected network $\mathcal{I}$;
Obtain the weight matrix $\mathcal{W}$ from infected network $\mathcal{I}$;
Compute regional scale Shapley value $\phi(r)$ by (4) with R;
Compute sub-regional scale Shapley value $\phi(\widetilde{r})$ by (5) with $\widetilde{R}$;
Compute backdoor weight matrix $\mathcal{W}^\phi$ by (6) with $\mathcal{W}$;
for *each epoch* **do**
  Obtain pre-pruned network $\mathcal{I}_p$ with preliminary pruning;
  Obtain fine-tuned network $\mathcal{I}_f$ with fine-tuning;
  Compute ACC_{pp} and ASR_{pp} on the pre-pruned network;
  Compute ACC_{ft}, ASR_{ft} and $\mathcal{L}$ on the fine-tuned network;
  Compute cost function J_{best} by (8), (9);
  if $J_{best} <$ *threshold* **then**
    Record the optimal result of ACC and ASR;
  end
  else
    Reselect backdoor weights for preliminary pruning;
  end
end
return Optimal result ACC and ASR

IV. Experiment And Result Analysis

In this section, to verify the effectiveness of the proposed MSAP method, backdoor defense experiments are conducted using the proposed methods and other state-of-the-art methods. In particular, experiments are performed on three open-access BCI public datasets: BCI competition III-IVb dataset (BCI-III-IVb), BCI competition III-IVa dataset (BCI-III-IVa) and BCI competition 1a dataset (BCI-IV-1a).

A. Dataset Description

1) Description of BCI-III-IVb dataset: The BCI-III-IVb dataset is composed of MI (motor imagery) signals collected from healthy subject seated comfortably with their arms on the armrests. The subject performed an MI task lasting 3.5 seconds, succeeded by a relaxation period ranging from 1.75 to 2.25 seconds. All 210 samples are recorded utilizing 118 electrode channels at a sampling frequency of 100 Hz. For the subject, two classes consist of 210 samples: class 1 for the left hand and class 2 for the right foot. Prior to analysis, the raw MI data undergo preprocessing, where signal sequences are filtered to accurately reflect the characteristics of the MI tasks.

2) Description of BCI-III-IVa dataset: The BCI-III-IVa dataset consists of MI signals recorded from five healthy subjects seated comfortably with their arms on the armrests. Each subject performed a 3.5-second MI task, followed by a relaxation period of 1.75 to 2.25 seconds. Three MI tasks are used: left hand (L), right hand (R), and right foot (F). Two types of visual stimuli are presented: fixed letters behind a cross and randomly moving objects, which could induce eye movements. The dataset includes EEG signals from 118 channels and 280 trial markers for each subject. Before analysis, the raw MI data undergoes preprocessing, including filtering of the signal sequences to accurately capture the characteristics of the MI tasks.

3) Description of BCI-IV-1a dataset: The BCI-IV-1a dataset consists of EEG recordings from healthy subjects performing MI tasks without feedback. The subjects are initially presented with a 4-second fixation on a computer screen, which serves as a cue for the motor imagery task. A total of 200 trials are recorded using 59 electrodes at a sampling rate of 100 Hz. Seven participants independently perform the task, generating two types of EEG data: one for left-handed imagery and another for right-handed imagery. The raw MI signals undergo preprocessing, including downsampling and filtering, to produce sequences that accurately capture the essential features of the MI signals.

B. Experiment Settings

The proposed method, multi-scale Shapley adaptation pruning (MSAP), is compared with 8 backdoor defense methods: Fine-Tuning (FT) [22], Fine-Pruning (FP) [35], stochastic activation pruning (SAP) [10], adversarial neuron pruning (ANP) [55], gradient-based model pruning (GBMP) [13], regional Shapley pruning (RSP), multi-scale Shapley segmented pruning (MSSP), and multi-scale Shapley segmented pruning based on fine-tuning (MSSP-FT). The RSP, MSSP and MSSP-FT are used as ablation studies containing regional Shapley information, multi-scale segmented Shapley mappings and fine-tuning on its basis. Besides, all backdoor defense methods are executed on 4 advanced CNN architectures: EEGNet [25], EEGInception [64], DeepConvNet [44], and ShallowConvNet. The comparison methods are introduced in detail as follows:

1) FT: a direct backdoor defense approach employs data fine-tuning to bolster the model's resistance against backdoor attack;

2) FP: an effective method streamlines parameters and reduces backdoor attack risks, enhancing model efficiency and security;

3) SAP: a stochastic pruning technique, inspired by fine-pruning, further enhances the streamlining efficiency of deep neural networks;

4) ANP: a method utilizes a hierarchical attention mechanism to dynamically prioritize important features, improving both accuracy and computational efficiency;

5) GBMP: a method uses gradient-based pruning strategy, leveraging the gradient of the forgetfulness loss to identify and remove backdoor elements;

6) RSP: A pruning method is processed using only regional scale information to compute the approximate backdoor weight distribution;

7) MSSP: a direct pruning approach targets the top m backdoor weights, computed by the multi-scale segmented Shapley mapping;

8) MSSP-FT: an optimal pruning strategy employs multi-scale Shapley segmented pruning, coupled with fine-tuning, to enhance classification outcomes.

The fulfill CNN architectures mentioned above are introduced in detail as follows:

1) EEGNet: a carefully designed CNN architecture, integrating deep and separable convolutions, is specialized for EEG-based BCI and performs well in EEG classification;

2) EEGInception: a multi-branch CNN architecture, enhanced with an inception module, is employed for effective EEG classification;

3) DeepConvNet: a deep CNN architecture, influenced by computer vision techniques, is characterized by its extensive use of filters and small kernel sizes;

4) ShallowConvNet: a shallow CNN architecture, motivated by the filter bank common spatial patterns, employs larger kernel sizes compared to DeepConvNet.

To verify the effectiveness of the proposed multi-scale Shapley adaptation pruning (MSAP) method, experiments are conducted on all the methods mentioned above under CNN architecture. In the experiments, the initial learning rate is fixed at 0.001 and optimized by grid search, with the batch size and the random seed set to 36 and 24, respectively. The number of neurons u, z of regional and sub-regional scale infected networks is set to 4, 16, respectively. The hyperparameters α, β in the cost function are set to 0.6 and 0.4, respectively, and the threshold θ for judging the cost function is set to 0.2. The adaptive moment estimation is chosen as the optimizer and the learning rate is decayed every 25 epochs by the multi-step scheduler. All methods are implemented in Python utilizing the PyTorch framework, and the hardware parameters of the training platform are 32 Intel(R) Xeon(R) Gold 5218 CPUs, 8 GeForce RTX 2080Ti 11GB GPUs, and 180GB RAM.

C. Attack Scenarios

1) Narrow Period Pulse (NPP): As the security of BCI is frequently neglected, backdoor attack has a great impact on BCI systems. NPP is the first backdoor attack method applied to BCI systems with significant results. Furthermore, the NPP simulates the scenario of an attacker influencing the use of real-world BCI systems by users. Therefore, the proposed MSAP method performs backdoor defense experiments on BCI system after NPP based attack [41]. The NPP is defined by:

$$\mathcal{N}_d(i) = \begin{cases} 0, & nTf_s \le i < (nT+\phi)f_s \\ a, & (nT+\phi)f_s \le i < (nT+dT+\phi)f_s \\ 0, & (nT+dT+\phi)f_s \le i < (n+1)Tf_s \end{cases} \quad (11)$$

where T is the period, d is the duty cycle (ratio of pulse duration to period), a is the amplitude, and f_s is the sampling rate. To simulate a real-world BCI system, the NPP purposely

TABLE I
ACCURACY (ACC) AND ATTACK SUCCESS RATE (ASR) AFTER BACKDOOR ATTACK ACROSS 5 CNN ARCHITECTURES

Datasets	Metrics	Method	Baseline	FT	FP	SAP	ANP	GBMP	RSP	MSSP	MSSP-FT	MSAP
BCI-III-IVb	ACC↑	EEGNet	89.54%	84.21%	55.78%	46.15%	72.05%	79.96%	69.29%	69.37%	82.78%	**86.27%**
		EEGInception	87.78%	61.04%	46.26%	48.41%	63.13%	64.91%	68.75%	41.29%	**86.21%**	81.29%
		DeepConvNet	84.23%	73.68%	48.41%	52.12%	64.03%	61.40%	69.07%	60.28%	74.28%	**82.75%**
		ShallowConvNet	85.27%	53.67%	52.62%	50.51%	68.42%	69.29%	60.52%	58.47%	**70.54%**	68.47%
	ASR↓	EEGNet	91.72%	63.15%	49.47%	75.78%	39.63%	38.11%	18.91%	11.06%	34.54%	**6.89%**
		EEGInception	93.38%	52.62%	49.58%	57.88%	40.15%	38.88%	21.28%	9.45%	36.56%	**8.56%**
		DeepConvNet	94.85%	75.78%	57.88%	54.23%	36.77%	36.81%	24.01%	7.41%	34.13%	**6.31%**
		ShallowConvNet	95.03%	46.46%	48.41%	47.41%	40.62%	29.55%	22.88%	10.56%	24.59%	**5.26%**
BCI-III-IVa	ACC↑	EEGNet	88.04%	69.29%	58.77%	54.38%	69.18%	72.36%	64.91%	62.28%	84.55%	**89.33%**
		EEGInception	83.33%	66.67%	46.92%	52.63%	71.05%	64.91%	63.04%	63.37%	77.57%	**79.96%**
		DeepConvNet	86.01%	67.54%	52.63%	45.04%	67.10%	69.29%	60.85%	57.89%	76.10%	**81.46%**
		ShallowConvNet	82.01%	61.40%	55.70%	66.67%	67.54%	73.46%	54.38%	56.14%	71.14%	**80.88%**
	ASR↓	EEGNet	92.46%	68.42%	74.01%	69.07%	33.06%	33.11%	27.10%	19.05%	36.05%	**7.94%**
		EEGInception	93.75%	56.57%	67.54%	62.05%	34.47%	39.25%	24.25%	14.23%	37.10%	**6.74%**
		DeepConvNet	95.03%	70.18%	63.20%	59.21%	29.65%	41.54%	21.61%	20.24%	36.66%	**7.78%**
		ShallowConvNet	97.24%	54.38%	65.35%	50.88%	37.52%	41.85%	28.66%	17.70%	39.12%	**6.25%**
BCI-IV-1a	ACC↑	EEGNet	92.22%	69.02%	59.65%	71.60%	77.39%	73.24%	65.78%	62.71%	75.10%	**84.26%**
		EEGInception	82.21%	57.12%	54.39%	53.94%	63.97%	67.97%	66.22%	50.87%	70.17%	**79.78%**
		DeepConvNet	85.40%	65.89%	47.14%	62.50%	68.19%	64.91%	70.18%	65.78%	77.85%	**80.70%**
		ShallowConvNet	87.48%	59.42%	49.12%	67.98%	63.15%	70.72%	59.21%	68.42%	66.22%	**78.56%**
	ASR↓	EEGNet	95.77%	53.50%	57.89%	73.46%	41.24%	36.43%	26.98%	9.87%	37.75%	**6.33%**
		EEGInception	95.58%	67.76%	40.35%	51.53%	36.64%	32.49%	23.27%	17.54%	34.81%	**8.61%**
		DeepConvNet	98.34%	62.82%	42.10%	65.02%	44.31%	44.48%	22.68%	19.62%	37.70%	**8.27%**
		ShallowConvNet	96.50%	47.80%	52.19%	53.50%	32.12%	31.87%	25.13%	10.71%	38.93%	**7.06%**

adds a random phase $\phi \in [0, T]$, since an attacker can't get the precise timing of an EEG trial. NPP serves as the malicious backdoor trigger, which is mixed with clean EEG data through the hybrid function to form infected data. Then, the infected data are injected into the overall dataset, which is inadvertently trained by the user to form an infected network that can specify a classification target for the infected data.

2) BadNets: The BadNets method is an important approach for studying the vulnerabilities of deep learning models in computer vision tasks [20]. It highlights the risks posed by backdoor attack, where a model's performance can be subtly compromised by maliciously modifying the training data. The main idea behind BadNets is to introduce a small, seemingly harmless trigger into a fraction of the training dataset. This trigger, which could be a small patch, color pattern, or noise, is designed to be invisible or imperceptible in normal data. The attack mechanism of BadNets is defined as follows:

$$\mathcal{T}(x) = x + \delta \cdot \mathbb{I}(\mathcal{T}_{trigger}(x)) \tag{12}$$

where $\mathbb{I}(\mathcal{T}_{trigger}(x))$ is an indicator function that returns 1 when the trigger is present in x and 0 otherwise, and δ is the perturbation added to x when the trigger is detected. The key to the BadNets is that the attacker manipulates the model's output using the trigger, while the clean samples in the training data remain unaffected. Therefore, BadNets can induce targeted misclassification in specific conditions, while maintaining high efficiency and accuracy in other scenarios.

3) Spatialspectral-Backdoor: The spatialspectral-Backdoor is a spectral active backdoor attack designed to improve the success rate of backdoor attack in the frequency domains [28], as expressed below:

$$\mathcal{K}(i) = \begin{cases} \alpha|\tilde{X}|, & \text{if } i = p \\ \alpha|\tilde{X}|^*, & \text{if } i = T - p \\ 0, & \text{otherwise} \end{cases} \quad \text{s.t. } \theta = \theta' \tag{13}$$

where $\mathcal{K}(\cdot)$ represents the designed amplitude frequency backdoor trigger, α denotes the ratio of the amplitude at the designated position, p is the position where the backdoor trigger is inserted, θ and θ' refer to the phase angles before and after the amplitude modification respectively, and $T = t(j)_{(j=1,\ldots,v)}$ indicates the sampling points. In order to maintain conjugate symmetry in the frequency domain, the amplitude at the conjugate symmetry point $T-p$ should be the complex conjugate of the amplitude at the p point, expressed as $|\tilde{X}|^*$. As a result, Spatialspectral-Backdoor is very stealthy in the time domain and poses challenges for detection in real-world application scenarios.

D. Experiment Result Analysis

The results of 5-fold cross-validation for backdoor defense experiments on three BCI open datasets, comparing the proposed MSAP method with other methods, are presented in Table I. Besides, Table I also includes the ACC on clean data and ASR (*lower is better*) for all the methods tested. The best

TABLE II
PERFORMANCE OF MSAP FOR DIFFERENT NUMBERS OF INFECTED DATA ON EEGNET AND SHALLOWCONVNET ARCHITECTURES

Attack Scenarios	Method	Infection Rate	Metrics			
			ACC↑(Before)	ACC↑(After)	ASR↓(Before)	ASR↓(After)
Narrow Period Pulse	EEGNet	5%	89.58%	82.06%	79.99%	10.00%
		10%	89.58%	87.58%	74.48%	6.89%
		15%	89.58%	77.93%	92.24%	19.65%
		20%	89.58%	78.62%	97.93%	18.96%
	ShallowConvNet	5%	86.89%	66.72%	73.35%	6.84%
		10%	86.89%	66.60%	83.44%	5.26%
		15%	86.89%	68.92%	91.03%	13.15%
		20%	86.89%	72.22%	98.62%	7.24%
BadNets	EEGNet	5%	89.58%	82.72%	72.59%	9.46%
		10%	89.58%	85.18%	83.45%	4.92%
		15%	89.58%	75.55%	91.54%	12.60%
		20%	89.58%	73.70%	96.14%	29.49%
	ShallowConvNet	5%	86.89%	77.19%	76.75%	6.10%
		10%	86.89%	78.94%	84.42%	9.34%
		15%	86.89%	75.65%	96.50%	15.78%
		20%	86.89%	71.82%	98.71%	21.51%
Spatialspectral-Backdoor	EEGNet	5%	89.58%	82.23%	85.96%	10.30%
		10%	89.58%	80.04%	89.69%	11.40%
		15%	89.58%	79.49%	93.19%	19.30%
		20%	89.58%	78.39%	99.26%	25.67%
	ShallowConvNet	5%	86.89%	84.42%	84.92%	9.74%
		10%	86.89%	80.04%	88.97%	12.06%
		15%	86.89%	77.30%	95.77%	17.65%
		20%	86.89%	72.49%	99.44%	21.60%

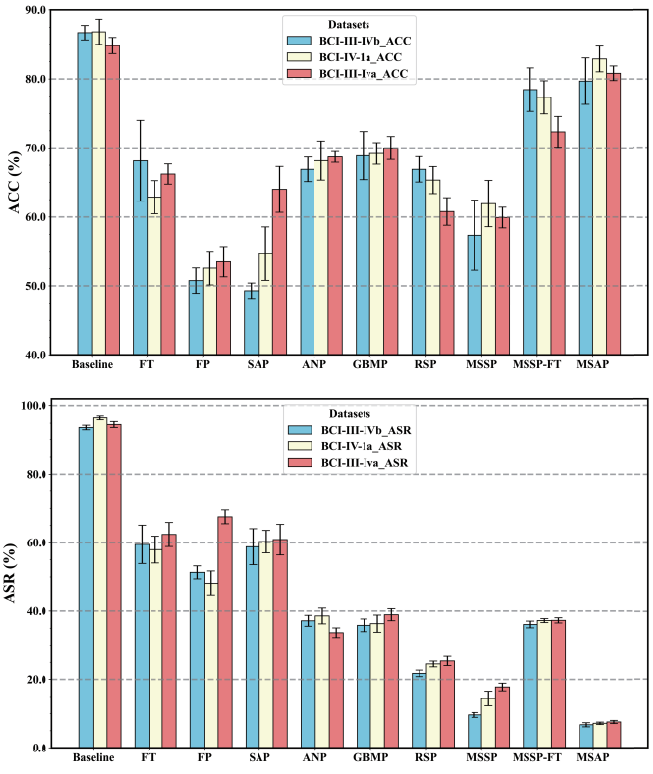


Fig. 5. Average ACC and ASR for backdoor defense using 5-fold cross-validation on three BCI public datasets.

results in each backdoor defense task are highlighted in bold, and the second best are underlined. For clear visualization, the histograms of ACC and ASR in Table I are shown in Fig. 5. Based on the experimental data in Table I and the comparative histograms in Fig. 5, several key observations can be summarized:

1) Comparison Study: To validate the effectiveness of MSAP in the field of BCI, the proposed method is compared with five pruning methods (FT, FP, SAP, ANP, and GBMP) on three open-source datasets. As shown in Table I and Fig. 5, MSAP consistently outperforms the other methods in both classification ACC and ASR, and the following observations can be obtained:

(1) Compared to traditional pruning-based backdoor defense methods (FT, FP, and SAP), MSAP consistently achieves higher classification ACC and significantly lower ASR across all three datasets. For example, on BCI-III-IVb with EEGNet, MSAP maintains 86.27% ACC (↓3.27% from baseline) and reduces ASR to 6.89% (↓84.83%), while FT, FP, and SAP yield unstable ACCs (84.21%, 55.78%, 46.15%) and high ASRs (63.15%, 49.47%, 75.78%). Similar results are observed across other datasets and models, confirming that MSAP achieves a better trade-off between defense effectiveness and model performance by leveraging Shapley value-based pruning.

(2) Compared with the recent pruning-based backdoor defense methods (ANP and GBMP), MSAP also yields superior results. From Table I and Fig. 5, ANP and GBMP exhibit average ACCs around 60%–70% and ASRs around 30%–40% across datasets and models. In contrast, MSAP achieves average ACCs above 80% and keeps ASRs consistently below 10%. This demonstrates the importance of applying Shapley value for precise attribution and pruning, which leads to more effective and targeted backdoor defense than heuristic-based approaches like ANP and GBMP.

2) Ablation Study: To verify the effectiveness of multi-scale Shapley segmented mapping and adaptive backdoor weight cost pruning, MSAP and three ablation methods are compared. As shown in Table I and Fig. 5, the methods for ablation experiments have good results, and the following observations can be obtained:

(1) Effectiveness of multi-scale Shapley segmented map-

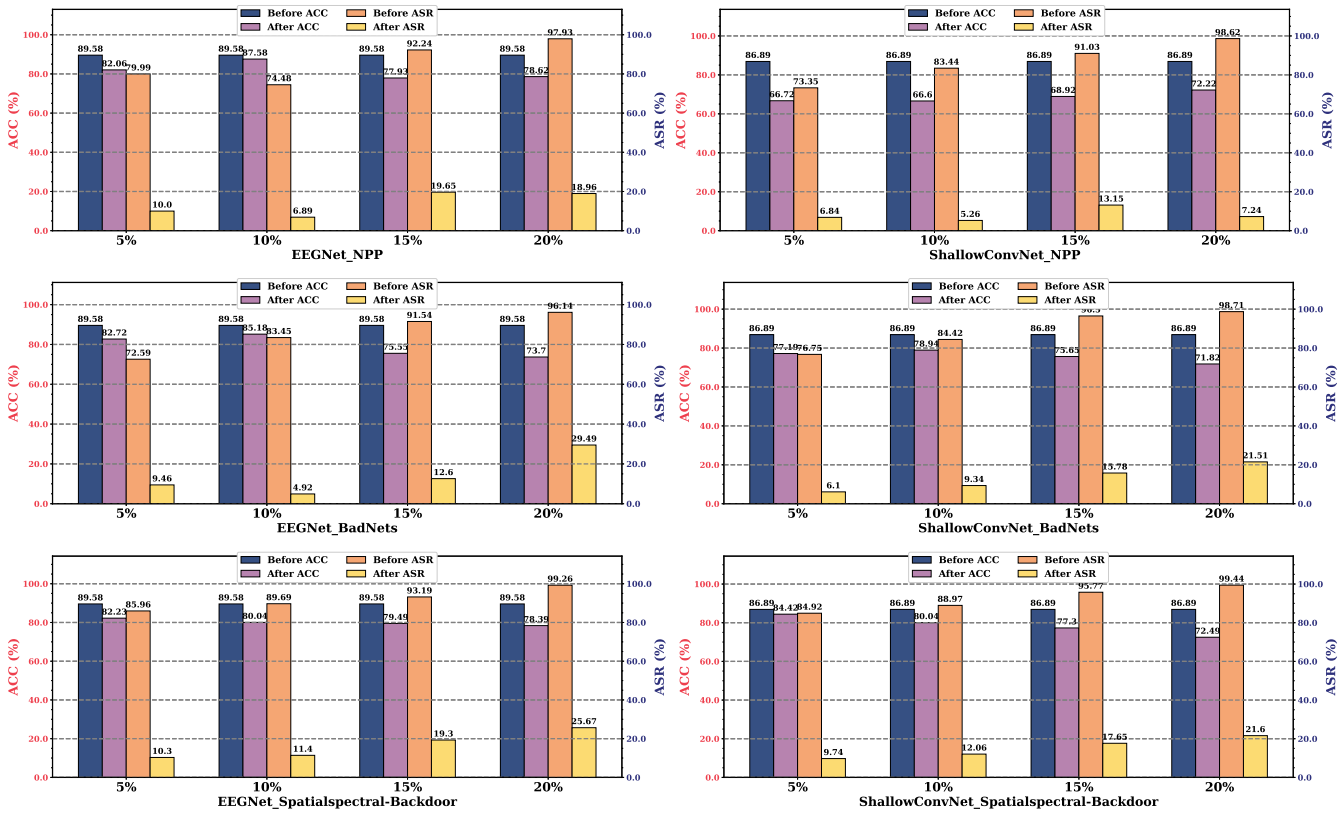


Fig. 6. Performance of MSAP on EEGNet and ShallowConvNet architectures for different infection rates in three attack scenarios.

ping: The RSP and MSSP are ablation methods based on multi-scale Shapley segmentation mapping. According to Fig. 5, it can be found that MSSP has an enhancement in both ACC and ASR, indicating that the multi-scale information plays an important role in pruning. Compared with RSP and MSSP, the proposed MSAP method further improves, with ACC and ASR reaching optimal values. This demonstrates the effectiveness of Shapley pruning based on multi-scale information.

(2) Effectiveness of adaptive backdoor weight cost pruning: MSSP-FT is an ablation experiment based on adaptive backdoor weight cost pruning. Based on Fig. 5, MSSP-FT is found to have improved ACC compared to MSSP under the effect of fine-tuning. In addition, the proposed MSAP is optimized based on the converged ASR of the cost function based on the high ACC of MSAP-FT. This result demonstrates the effectiveness of adaptive pruning based on cost function.

E. Analysis of Number of Infected Data

In backdoor defense, the quantity of infected data plays a crucial role in determining the choice of defense strategy. When the proportion of infected data is low, a fine-grained defense strategy is required. The defense becomes challenging with the large volume of infected data, and a substantial amount of clean data is needed to counter the infection effectively. Therefore, the quantity of infected data directly influences the difficulty of designing and implementing an effective defense strategy. Based on the above description, the generalization and effectiveness of the methods are evaluated by varying the infection rates across three attack scenarios (*NPP, BadNets, Spatialspectral-Backdoor*). Based on Table II and Fig. 6, it can be summarized:

1) Regarding ACC: The ACC of EEGNet and ShallowConvNet after MSAP is shown in Table II. For the three BCI datasets, as the amount of infected data increases, there is a slight fluctuating decrease in ACC. However, the decrease is minimal, indicating that MSAP effectively preserves the model's performance and demonstrates excellent generalization.

2) Regarding ASR: The ASR after pruning is significantly reduced for all infection percentages across the three attack scenarios. The average ASR for the three attack scenarios on EEGNet drops from 88.02% to 14.88% across all infection data. Similarly, the average ASR for ShallowConvNet decreases from 89.32% to 12.18%, marking a substantial improvement. This highlights the generalization and effectiveness of the MSAP method on both EEGNet and ShallowConvNet.

Remark 3: To ensure fair and consistent comparisons, we use the same clean pre-trained model across all infection rates and attack types. This controls variables by removing differences caused by random initialization or training randomness. As a result, ACC (Before) stays the same, and any changes in ACC (After) or ASR are only due to the backdoor attacks and defenses, not the starting model.

F. Impact of Key Parameters

In backdoor defense, the selection of key parameters plays a crucial role in determining the effectiveness of the defense.

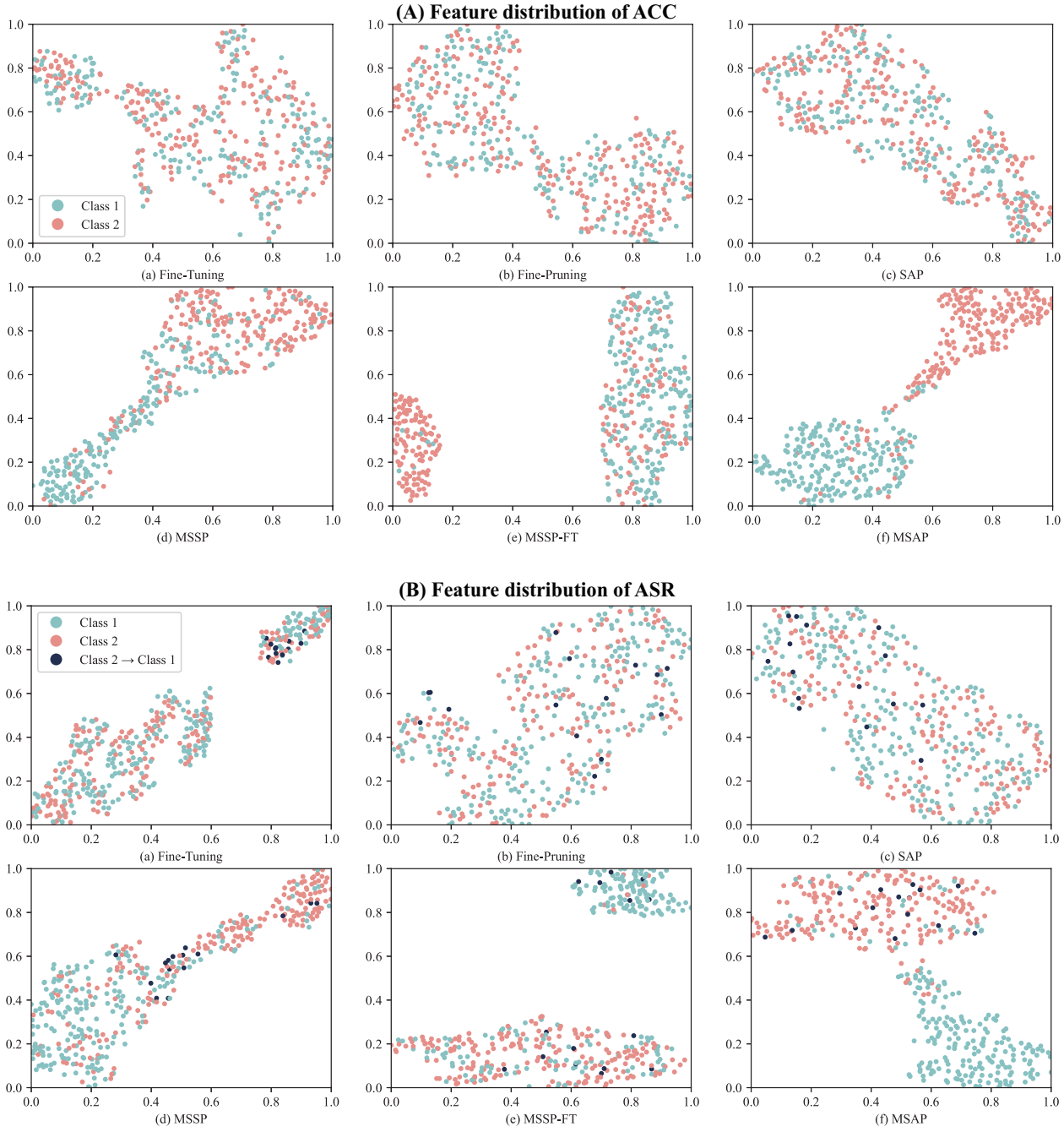


Fig. 7. Feature distribution of ACC and ASR is extracted by different backdoor defense methods.

The key parameters reflect the backdoor defense methods' stability, adaptability, and strength. Based on the above discussion, key parameters in MSAP are needed for detailed analysis. By observing changes in these parameters, the effectiveness and stability of MSAP can be directly verified. Therefore, the number of scales and the threshold value are chosen to evaluate the performance of MSAP.

1) Impact of Number of Scales: The number of scales is crucial in the multi-scale segmented Shapley mapping. Reducing the number of scales significantly shortens the time required to compute the Shapley values. However, using a smaller number of scales leads to some precision loss, resulting in an imprecise distribution of backdoor weights. Based on this, experiments are conducted to examine the impact of the number of segments at both the regional and sub-regional scales.

TABLE III
THE EFFECT OF THE NUMBER OF SCALES ON BACKDOOR DEFENSE

Regional Segments	**2**	**3**	**4**	**5**
Subregional Segments	**4**	**9**	**16**	**25**
ACC↑(metric)	61.69%	77.19%	86.27%	87.84%
ASR↓(metric)	31.26%	19.30%	6.89%	7.19%

Based on Table III, the progressively increasing number

of scales results in higher defense accuracy, highlighting the importance of accurate information. As both regional and sub-regional segments increase, there are noticeable improvements in both ACC and ASR. Ultimately, the metrics stabilize with 4 regional segments and 16 sub-regional segments. While further increasing the number of scales continues to improve the metrics, the effect becomes minimal and requires significant computational time. Therefore, the experiments are conducted with two scale settings: 4 for the region scale and 16 for the sub-regional scale.

2) Impact of Threshold: The threshold is a key parameter in adaptive backdoor weight cost pruning, as it influences the convergence of the network after pruning. By gradually decreasing the threshold, the criteria for network convergence become stricter. Therefore, changes in network metrics before and after pruning, as well as the magnitude of the loss function, reflect the effectiveness of the pruned network. Based on this, experiments are conducted to observe the impact of threshold magnitude on the metrics.

TABLE IV
THE EFFECT OF THRESHOLD MAGNITUDE ON BACKDOOR DEFENSE

Threshold	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
ACC↑(metric)	87.50%	86.27%	71.69%	63.23%	57.35%
ASR↓(metric)	6.43%	6.89%	16.61%	11.68%	10.55%

Table IV visually illustrates the impact of threshold changes on defense effectiveness. As the threshold becomes progressively stricter (decreasing values), both ACC and ASR stabilize and improve. However, the enhancement becomes limited at a threshold of 0.1, with no significant increase in the metrics beyond this point. This indicates that the network convergence is stabilized and can only fluctuate slightly. Therefore, the threshold value for the MSAP-related experiments is set at 0.2.

G. Feature Distribution Visualization

To demonstrate the effectiveness of the proposed method, Fig. 7 presents the distribution of data features for all methods under comparison, visualized using the t-SNE method [40]. In this figure, part (A) presents the data features about ACC after backdoor defense, while part (B) presents the data features of ASR. From the 12 subplots in Fig. 7, the following observations can be obtained:

1) As shown in Fig. 7(A)(f), the pruned network obtained by MSAP has clearer classification boundary between different classes compared to other backdoor defense methods (Fig. 7(A)(a)-(e)). *Class1* and *Class2* are almost categorized on either side of the boundary. This indicates that adaptive backdoor weight cost pruning is satisfactory and helpful clean data classification.

2) As shown in Fig. 7(B)(d)-(f), the proposed MSAP achieves the most obvious backdoor label (*Class2 to Class1*) misclassification boundary, most of which are in *Class2*, and it is not classified as *Class1* as intended by the attacker. Compared to Fig. 7(B)(a)-(c), these results demonstrate the effectiveness of using multi-scale segmented Shapley mapping to compute backdoor weights for pruning.

Therefore, as depicted in Fig 7, MSAP outperforms other methods in achieving optimal results, and adaptive pruning based on multi-scale segmented Shapley mapping facilitates learning backdoor features for pruning.

V. CONCLUSION

Recent research highlights security issues in BCI systems for EEG classification, especially the threat of backdoor attack on motor imagery (MI) signal classification, which can seriously compromise system safety. To tackle this problem, this paper introduces a new perspective on the security challenges in MI signal classification and proposes a practical solution: multi-scale Shapley adaptation pruning (MSAP). This method is designed to reduce the vulnerability of BCI systems to backdoor attack. MSAP has been tested on three open BCI datasets across four models, with its performance evaluated using five cross-validation methods. The results, shown through clear and simple histograms, demonstrate that MSAP effectively strengthens EEG-based BCI systems against backdoor attack.

Despite the advantages of multi-scale Shapley adaptation pruning (MSAP), numerous challenges require in-depth investigation. EEG data, rich in spatial information, remains vulnerable to backdoor attack that manipulate electrode domain information and alter topological arrangement rules, complicating defense strategies. Future research should focus on enhancing topological results in EEG electrode space, as most EEG classification networks currently do not perform spatial feature extraction from EEG. This improvement is vital for highlighting security concerns in EEG. Additionally, exploring the interaction between EEG backdoor attack and spatial domain information is crucial for developing stealthy attack methods.

REFERENCES

[1] A. Ballas and C. Diou, "Towards domain generalization for ECG and EEG classification: Algorithms and benchmarks," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2023.

[2] B. Chen, W. Carvalho, N. Baracaldo, H. Ludwig, B. Edwards, T. Lee, I. Molloy, and B. Srivastava, "Detecting backdoor attacks on deep neural networks by activation clustering," *arXiv preprint arXiv:1811.03728*, 2018.

[3] H. Chen, I. C. Covert, S. M. Lundberg, and S.-I. Lee, "Algorithms to estimate Shapley value feature attributions," *Nature Machine Intelligence*, pp. 1–12, 2023.

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

[5] Z. Chen, R. Yang, M. Huang, Z. W. Wang, and X. Liu, "Electrode domain adaptation network: Minimizing the difference across electrodes in single-source to single-target motor imagery classification," *IEEE Transactions on Emerging Topics in Computational Intelligence*, pp. 1–12, 2023.

[6] Z. Chen, R. Yang, M. Huang, F. Li, G. L and Z. W. Wang, "EEGprogress: A fast and lightweight progressive convolution architecture for EEG classification," *Computers in Biology and Medicine*,p. 107901, 2023.

[7] F. Cui, R. Wang, W. Ding, Y. Chen, and L. Huang, "A novel DE-CNN-BILSTM multi-fusion model for EEG emotion recognition," *Mathematics*, vol. 10, no. 4, p. 582, 2022.

[8] D. Dai, J. Li, Y. Song and F. Yang, Event-based recursive filtering for nonlinear bias-corrupted systems with amplify-and-forward relays, *Systems Science & Control Engineering*, vol. 12, no. 1, art. no. 2332419, 2024.

[9] W. Diao, W. He, K. Liang and X. Tan, Adaptive impulsive consensus of nonlinear multi-agent systems via a distributed self-triggered strategy, *International Journal of Systems Science*, vol. 55, no. 11, pp. 2224–2238, 2024.

[10] G. S. Dhillon, K. Azizzadenesheli, Z. C. Lipton, J. Bernstein, J. Kossaifi, A. Khanna, and A. Anandkumar, "Stochastic activation pruning for robust adversarial defense," *arXiv preprint arXiv:1803.01442*, 2018.

[11] B. G. Doan, E. Abbasnejad, and D. C. Ranasinghe, "Februus: Input purification defense against trojan attacks on deep neural network systems," in *Annual Computer Security Applications Conference*, pp. 897–912, 2020.

[12] X. Du, C. Ma, G. Zhang, J. Li, Y.-K. Lai, G. Zhao, X. Deng, Y.-J. Liu, and H. Wang, "An efficient LSTM network for emotion recognition from multichannel EEG signals," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1528–1540, 2020.

[13] K. Dunnett, R. Arablouei, D. Miller, V. Dedeoglu, and R. Jurdak, "Unlearning backdoor attacks through gradient-based model pruning," *arXiv preprint arXiv:2405.03918*, 2024.

[14] W. Ehab, L. Huang and Y. Li, UNet and variants for medical image segmentation, *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 2, art. no. 100009, Jun. 2024.

[15] Y. Gao, D. Wu, J. Zhang, G. Gan, S.-T. Xia, G. Niu, and M. Sugiyama, "On the effectiveness of adversarial training against backdoor attacks," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[16] Y. Gao, J. W. Stokes, M. A. Prasad, A. T. Marshall, K. Fawaz, and E. Kiciman, "I know your triggers: Defending against textual backdoor attacks with benign backdoor augmentation," in *MILCOM 2022-2022 IEEE Military Communications Conference (MILCOM)*, 2022, pp. 442–449.

[17] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 440–460, 2019.

[18] M. Goldblum, D. Tsipras, C. Xie, X. Chen, A. Schwarzschild, D. Song, A. Mądry, B. Li, and T. Goldstein, "Dataset security for machine learning: Data poisoning, backdoor attacks, and defenses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 1563–1580, 2022.

[19] X. Gong, Y. Chen, W. Yang, Q. Wang, Y. Gu, H. Huang, and C. Shen, "Redeem myself: Purifying backdoors in deep learning models using self-attention distillation," in *2023 IEEE Symposium on Security and Privacy (SP)*, 2023, pp. 755–772.

[20] G. Gu, K. Liu, B. Dolan-Gavitt, et al., "Badnets: Evaluating backdooring attacks on deep neural networks," *IEEE Access*, vol. 7, pp. 47230–47244, 2019.

[21] J. Guan, Z. Tu, R. He, and D. Tao, "Few-shot backdoor defense using Shapley estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13358–13367, 2022.

[22] Y. Guo, Y. Li, L. Wang, and T. Rosing, "Adafilter: Adaptive filter fine-tuning for deep transfer learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 4060–4066, 2020.

[23] Z. Jiang, Y. Ma, B. Shi, X. Lu, J. Xing, N. Gonçalves, and B. Jin, "Social NSTransformers: Low-quality pedestrian trajectory prediction," *IEEE Transactions on Artificial Intelligence*, 2024.

[24] W. Jiang, X. Wen, J. Zhan, X. Wang, and Z. Song, "Interpretability-guided defense against backdoor attacks to deep neural networks," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 8, pp. 2611–2624, 2021.

[25] 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," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.

[26] G. Li, J. Wu, S. Li, W. Yang, and C. Li, "Multitentacle federated learning over software-defined industrial internet of things against adaptive poisoning attacks," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1260–1269, 2022.

[27] L.-F. Li, Y. Hua, Y.-H. Liu and F.-H. Huang, Study on fast fractal image compression algorithm based on centroid radius, *Systems Science & Control Engineering*, vol. 12, no. 1, art. no. 2269183, 2024.

[28] F. Li, M. Huang, W. You, L. Zhu, H. Cheng, and R. Yang, "Spatialspectral-Backdoor: Realizing backdoor attack for deep neural networks in brain–computer interface via EEG characteristics," *Neurocomputing*, 2024, Art. no. 128902.

[29] B. Li and W. Li, Distillation-based user selection for heterogeneous federated learning, *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 2, art. no. 100007, Jun. 2024.

[30] G. Li, J. Shen, C. Dai, J. Wu, and S. I. Becker, "ShvEEGc: EEG clustering with improved cosine similarity-transformed Shapley value," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 1, pp. 222–236, 2022.

[31] J. Li, Y. Suo, S. Chai, Y. Xu and Y. Xia, Resilient and event-triggered control of singular Markov jump systems against cyber attacks, *International Journal of Systems Science*, vol. 55, no. 2, pp. 222–236, 2024.

[32] X. Li, Y. Zhang, P. Tiwari, D. Song, B. Hu, M. Yang, Z. Zhao, N. Kumar, and P. Marttinen, "EEG based emotion recognition: A tutorial and review," *ACM Computing Surveys*, vol. 55, no. 4, pp. 1–57, 2022.

[33] X. Li, X. Tang, S. Qiu, X. Deng, H. Wang, and Y. Tian, "Subdomain adversarial network for motor imagery EEG classification using graph data," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2023.

[34] Y. Li, W. Zheng, L. Wang, Y. Zong, and Z. Cui, "From regional to global brain: A novel hierarchical spatial-temporal neural network model for EEG emotion recognition," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 568–578, 2019.

[35] K. Liu, B. Dolan-Gavitt, and S. Garg, "Fine-pruning: Defending against backdooring attacks on deep neural networks," in *International symposium on research in attacks, intrusions, and defenses*, pp. 273–294, Springer, 2018.

[36] N. Liu and W. Qian, Stability analysis of low-voltage direct current system with time-varying delay, *Systems Science & Control Engineering*, vol. 12, no. 1, art. no. 2293918, 2024.

[37] S. Liu, X. Wang, L. Zhao, B. Li, W. Hu, J. Yu, and Y.-D. Zhang, "3dcann: A spatio-temporal convolution attention neural network for EEG emotion recognition," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 11, pp. 5321–5331, 2021.

[38] Y. Liu, M. Fan, C. Chen, X. Liu, Z. Ma, L. Wang, and J. Ma, "Backdoor defense with machine unlearning," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*, pp. 280–289, IEEE, 2022.

[39] Y. Liu, Y. Xie, and A. Srivastava, "Neural trojans," in *2017 IEEE International Conference on Computer Design (ICCD)*, pp. 45–48, IEEE, 2017.

[40] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 11, 2008.

[41] L. Meng, X. Jiang, J. Huang, Z. Zeng, S. Yu, T.-P. Jung, C.-T. Lin, R. Chavarriaga, and D. Wu, "EEG-based brain-computer interfaces are vulnerable to backdoor attacks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023.

[42] P. Pandey and K. Seeja, "Subject independent emotion recognition from EEG using vmd and deep learning," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 5, pp. 1730–1738, 2022.

[43] C. Qin, R. Yang, M. Huang, W. Liu, and Z. Wang, "Spatial variation generation algorithm for motor imagery data augmentation: Increasing the density of sample vicinity," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023.

[44] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[45] S. Supriya, S. Siuly, H. Wang, and Y. Zhang, "Epilepsy detection from EEG using complex network techniques: A review," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 292–306, 2021.

[46] M. Sundararajan and A. Najmi, "The many Shapley values for model explanation," in *International Conference on Machine Learning*, 2020, pp. 9269–9278.

[47] G. Tao, Y. Liu, G. Shen, Q. Xu, S. An, Z. Zhang, and X. Zhang, "Model orthogonalization: Class distance hardening in neural networks for better security," in *2022 IEEE Symposium on Security and Privacy (SP)*, pp. 1372–1389, IEEE, 2022.

[48] B. Wang, Y. Yao, S. Shan, H. Li, B. Viswanath, H. Zheng, and B. Y. Zhao, "Neural cleanse: Identifying and mitigating backdoor attacks in neural networks," in *2019 IEEE Symposium on Security and Privacy (SP)*, pp. 707–723, IEEE, 2019.

[49] D. Wang, C. Wen and X. Feng, Deep variational Luenberger-type observer with dynamic objects channel-attention for stochastic video prediction, *International Journal of Systems Science*, vol. 55, no. 4, pp. 728–740, 2024.

[50] S. Wang, S. Nepal, C. Rudolph, M. Grobler, S. Chen, and T. Chen, "Backdoor attacks against transfer learning with pre-trained deep learning models," *IEEE Transactions on Services Computing*, vol. 15, no. 3, pp. 1526–1539, 2020.

[51] Y. Wang, C. Shen, J. Huang and H. Chen, Model-free adaptive control for unmanned surface vessels: a literature review, *Systems Science & Control Engineering*, vol. 12, no. 1, art. no. 2316170, 2024.

[52] Z. Wei, J. Shi, Y. Duan, R. Liu, Y. Han, and Z. Liu, "Backdoor filter: Mitigating visible backdoor triggers in dataset," in *2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI)*, pp. 102–105, IEEE, 2021.

[53] E. Winter, "The Shapley value," in *Handbook of Game Theory with Economic Applications*, vol. 3, pp. 2025–2054, Elsevier, 2002.

[54] C. Wu, X. Yang, S. Zhu, and P. Mitra, "Toward cleansing backdoored neural networks in federated learning," in *Proceedings of the 2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS)*, IEEE, pp. 820–830, 2022.

[55] D. Wu and Y. Wang, "Adversarial neuron pruning purifies backdoored deep models," *Advances in Neural Information Processing Systems*, vol. 34, pp. 16913–16925, 2021.

[56] D. Wu, J.-T. King, C.-H. Chuang, C.-T. Lin, and T.-P. Jung, "Spatial filtering for EEG-based regression problems in brain–computer interface (BCI)," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 771–781, 2017.

[57] D. Wu, Y. Xu, and B.-L. Lu, "Transfer learning for EEG-based brain–computer interfaces: A review of progress made since 2016," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 1, pp. 4–19, 2020.

[58] Y. Wu, X. Huang, Z. Tian, X. Yan and H. Yu, Emotion contagion model for dynamical crowd path planning, *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 3, art. no. 100014, Sept. 2024.

[59] Y. Xiao, G. Cai and G. Duan, High-order adaptive dynamic surface control for output-constrained non-linear systems based on fully actuated system approach, *International Journal of Systems Science*, vol. 55, no. 3, pp. 482–498, 2024.

[60] H. Xiong, G. Chen, H. Ren, H. Li and R. Lu, Event-based model-free adaptive consensus control for multi-agent systems under intermittent attacks, *International Journal of Systems Science*, vol. 55, no. 10, pp. 2062–2076, 2024.

[61] M. Xue, Y. Wu, Z. Wu, Y. Zhang, J. Wang, and W. Liu, "Detecting backdoor in deep neural networks via intentional adversarial perturbations," *Information Sciences*, vol. 634, pp. 564–577, 2023.

[62] L. Zhao and B. Li, Adaptive fixed-time control for multiple switched coupled neural networks, *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 3, art. no. 100018, Sept. 2024.

[63] P. Zhao, P.-Y. Chen, P. Das, K. N. Ramamurthy, and X. Lin, "Bridging mode connectivity in loss landscapes and adversarial robustness," *arXiv preprint arXiv:2005.00060*, 2020.

[64] C. Zhang, Y.-K. Kim, and A. Eskandarian, "EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification," *Journal of Neural Engineering*, vol. 18, no. 4, p. 046014, 2021.

[65] H. Zhang, G. Hua, X. Wang, H. Jiang, and W. Yang, "Categorical inference poisoning: Verifiable defense against black-box DNN model stealing without constraining surrogate data and query times," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 1473–1486, 2023.

[66] R. Zhang, H. Liu, Y. Liu and H. Tan, Dynamic event-triggered state estimation for discrete-time delayed switched neural networks with constrained bit rate, *Systems Science & Control Engineering*, vol. 12, no. 1, art. no. 2334304, 2024.

[67] X. Zhang and D. Wu, "On the vulnerability of CNN classifiers in EEG-based BCIs," *IEEE Transactions on Neural systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 814–825, 2019.

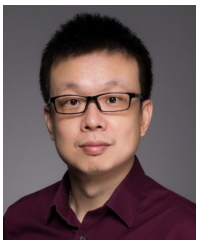
[68] H. Zheng, H. Xiong, J. Chen, H. Ma, and G. Huang, "Motif-backdoor: Rethinking the backdoor attack on graph neural networks via motifs," *IEEE Transactions on Computational Social Systems*, IEEE, 2023.

[69] J. Zhu, L. Chen, D. Xu, and W. Zhao, "Backdoor defence for voice print recognition model based on speech enhancement and weight pruning," *IEEE Access*, vol. 10, pp. 114016–114023, 2022.

[70] L. Zou, Z. Wang, B. Shen, and H. Dong, "Encryption-decryption-based state estimation with multi-rate measurements against eavesdroppers: A recursive minimum-variance approach," *IEEE Transactions on Automatic Control*, 2023.

[71] Y. Zou and E. Tian, Guaranteed cost intermittent control for discrete-time system: a data-driven method, *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 3, art. no. 100015, Sept. 2024.

Fumin Li received the B.Eng. degree in Northeast Forestry University, Harbin, China, in 2022. He is currently pursuing the M.Sc. degree at the University of Liverpool, UK. His research interests include backdoor attack and deep learning in brain-computer interface.



Rui Yang received the B.Eng. degree in Computer Engineering and the Ph.D. degree in Electrical and Computer Engineering from National University of Singapore in 2008 and 2013 respectively.

He is currently an Associate Professor in the School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China, and an Honorary Lecturer in the Department of Computer Science, University of Liverpool, Liverpool, United Kingdom. His research interests include machine learning based data analysis and applications. He is the author or co-author of several technical papers and also a very active reviewer for many international journals and conferences. Dr. Yang is currently serving as an Associate Editor for *IEEE Transactions on Instrumentation and Measurement*, *Neurocomputing* and *Cognitive Computation*.

Hanjing Cheng received the Ph.D. degree in control science and engineering from the Nanjing University of Science and Technology in 2022. From 2015-2016, she was a visiting Ph.D. student with the Department of Computer Science, Brunel University London, Uxbridge, U.K, for six months. She is currently a lecturer at the School of Electronic and Information Engineering, Suzhou University of Science and Technology. Her current research interests include network pruning, multi-objective optimization, and adversarial attack.

Mengjie Huang received the Ph.D. degree from National University of Singapore in 2014, and the B.Eng degree from Sichuan University in 2009.

She is now an Associate Professor in the Design School, Xi'an Jiaotong-Liverpool University, Suzhou, China. Dr Huang's current research interests include human-computer interaction and applications.

Fanglue Zhang received the B.Eng. degree from the University of Jinan, Jinan, China, in 2021. He is currently pursuing the M.Res. degree at the University of Liverpool, UK. His research interest is deep learning and artificial intelligence.

Fuad E. Alsaadi received the B.Sc. and M.Sc. degrees in electronic and communication from King AbdulAziz University, Jeddah, Saudi Arabia, in 1996 and 2002, respectively and the Ph.D. degree in optical wireless communication systems from the University of Leeds, Leeds, U.K., in 2011. Between 1996 and 2005, he was with Jeddah as a Communication Instructor with the College of Electronics & Communication. He is currently an Associate Professor with the Electrical and Computer Engineering Department within the Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia. He has authored or coauthored widely in the top IEEE Communications Conferences and Journals. His research interests include optical systems and networks, signal processing, synchronization and systems design. He was the recipient of the Carter Award, University of Leeds for the best Ph.D.

Zidong Wang (Fellow, IEEE) received the B.Sc. degree in mathematics in 1986 from Suzhou University, Suzhou, China, the M.Sc. degree in applied mathematics and the Ph.D. degree in electrical engineering both from Nanjing University of Science and Technology, Nanjing, China, in 1990 and 1994, respectively.

He is currently Professor of Dynamical Systems and Computing in the Department of Computer Science, Brunel University London, U.K. From 1990 to 2002, he held teaching and research appointments in universities in China, Germany and the UK. Prof. Wang's research interests include dynamical systems, signal processing, bioinformatics, control theory and applications. He has published more than 700 papers in international journals. He is a holder of the Alexander von Humboldt Research Fellowship of Germany, the JSPS Research Fellowship of Japan, William Mong Visiting Research Fellowship of Hong Kong.

Prof. Wang serves (or has served) as the Editor-in-Chief for *International Journal of Systems Science*, the Editor-in-Chief for *Neurocomputing*, the Editor-in-Chief for *Systems Science & Control Engineering*, and an Associate Editor for 12 international journals, including IEEE TRANSACTIONS ON AUTOMATIC CONTROL, IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, IEEE TRANSACTIONS ON NEURAL NETWORKS, IEEE TRANSACTIONS ON SIGNAL PROCESSING, and IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS–PART C. He is a Member of the Academia Europaea, a Member of the European Academy of Sciences and Arts, an Academician of the International Academy for Systems and Cybernetic Sciences, a Fellow of the IEEE, a Fellow of the Royal Statistical Society, and a member of program committee for many international conferences.