Discriminability of single-trial EEG during decision-making of cooperation or aggression: A study based on machine learning

Noname manuscript No.
(will be inserted by the editor)

Discriminability of single-trial EEG during decision-making of cooperation or aggression: A study based on machine learning

Zhihua Huang · Kun Jiang · Jing Li · Wenxing Zhu · Huiru Zheng · Yiwen Wang

Received: date / Accepted: date

Abstract Decision-making is a very important cognitive process in our daily life. There has been increasing interest in the discriminability of single-trial electroencephalogram (EEG) during decision-making. In this study, we designed a machine learning based framework to explore the discriminability of single-trial EEG corresponding to different decisions. For each subject, the framework split the decision-making trials into two parts, trained a feature model and a classifier on the first part, and evaluated the discriminability on the second part using the feature model and classifier. A proposed algorithm and five existing algorithms were applied to fulfill the feature models, and the algorithm Linear Discriminative Analysis (LDA) was used to implement the classifiers. We recruited 21 subjects to participate in Chicken Game (CG) experiments. The results show that there exists the discriminability of single-trial EEG between the cooperation and aggression decisions during the CG experiments, with the classification accuray of 75% ($\pm$6%), and the discriminability is mainly from the EEG information below 40 Hz. The further analysis indicates that the contributions of different brain regions to the discriminability are consistent with the existing knowledge on the cognitive mechanism of decision-making, confirming the reliability of the conclusions. This study exhibits that it is feasible to apply machine learning methods to EEG analysis of decision-making cognitive process.

Keywords Discriminability of single-trial EEG · Adaptive frequency common spatial pattern · Decision-making · Chicken game

Correspondence should be addressed to Zhihua Huang, College of Computer and Data Science, Fuzhou University, Fuzhou 350108, China. E-mail: hzh@fzu.edu.cn (Z.H.); h.zheng@ulster.ac.uk (H.Z.); wangeven@126.com (Y.W.)

Zhihua Huang, Kun Jiang, Jing Li
College of Computer and Data Science, Fuzhou University, Fuzhou 350108, China.

Wenxing Zhu
School of Mathematics and Statistics, Fuzhou University, Fuzhou 350108, China.

Huiru Zheng
School of Computing, Ulster University, Belfast, United Kingdom.

Yiwen Wang
School of Economics and Management, Fuzhou University, Fuzhou 350108, China.

1 Introduction

In daily life, we always face all kinds of choices. Decision-making is a very important cognitive process, concerning interpersonal interactions, social activities, economics theory, and so on [1–3]. The cognitive mechanism of decision-making is a very interesting research problem. Existing research revealed the associations of decision-making with emotion, personality and motivation [1,4,5], even with psychiatric disorder [6], showing its complexity. Discriminability of brain signals during decision-making is often used by researchers to explore neural processes that drive decision-making [7–10]. In most cases, the discriminability of brain signals during decision-making is observed by averaging brain signals across trials [7–10]. Observing the single-trial discriminability of brain signals is also very valuable for exploring neural processes during decision-making.

In recent years, Si *et al* [3, 11] probed the discriminability of single-trial electroencephalogram (EEG) during decision-making. Although the main aim of [3] was to reveal the diverse network patterns during the different decision stages using EEG, Si *et al* also performed the classification of decisions by calculating the out-degree statistical measurements of the time-varying networks built using EEG at the hub nodes Fz and O2 within 280-300 ms. The combination of the network features with the conventional Linear

Discriminative Analysis (LDA) resulted in 70% accuracy in discriminating the two decisions. Si *et al* proposed the Discriminative Spatial Network Pattern (DSNP) method to extract the features from the single-trial brain network [11]. The LDA classifiers trained on the DSNP features were then used to predict the individual decisions trial-by-trial. The performances achieved an accuracy of 0.88 ± 0.09 for the first dataset, and 0.90 ± 0.10 for the second dataset.

In both studies [3, 11], the EEG data sets were collected when the subjects were carrying out Ultimatum Game (UG) tasks. The results show that the single-trial EEG during decision-making in UG is discriminable. However, UG is not a unique cognitive task used to explore the neural mechanism of decision-making. Among various cognitive tasks implicating decision-making, the interpersonal interaction Chicken Game (CG) experiment (see Appendix A for details), which allows researchers to operate the concepts such as cooperation or aggression in laboratory settings [12], is also frequently employed in this type of research [10, 13]. Researchers also expect to probe the discriminability of single-trial EEG during decision-making in CG.

In the view of machine learning, discriminating single-trial EEG is a classification problem [14–16], which usually includes two sections: feature extraction and classification. In this research, we designed a machine learning based framework, which also includes the two sections of feature extraction and classification, to investigate the discriminability of single-trial EEG during decision-making in CG.

Regarding feature extraction from EEG, some commonly used methods in the field of pattern recognition, such as Window Mean (WM), Principal Component Analysis (PCA) and Independent Component Analysis (ICA), have been extensively applied [15, 16]. Common Spatial Pattern (CSP) [17,18], a supervised spatial filter, exhibited good performances in practice [15,16]. Due to the remarkable potential discriminating EEG of different conditions, CSP has always attracted the attention of researchers and has been optimized in its successively developed variants [16,19–23]. In order to apply CSP in the small-sample setting of decision-making cognitive tasks and identify which EEG frequency bands are associated with decision-making cognitive activities, we proposed an Adaptive Frequency Regularized Common Spatial Pattern (AFRCSP) method, which is based on the ideas of [18,21,23–27]. In the framework, the six feature extraction methods of AFRCSP, DSNP, WM, PCA, ICA and CSP were incorporated.

The effectiveness of the LDA algorithm to discriminate single-trial EEG has been confirmed by extensive research cases of Brain-Computer Interface (BCI) [15, 16]. In our framework, the LDA algorithm was used as the classification method. This machine learning based study showed that discriminating single-trial EEG during decision-making of cooperation or aggression in the CG experiment is feasible and the discriminability is mainly underlain by the EEG information below 40 Hz.

2 Material and Methods

2.1 Subjects

In this study, 21 right-handed university male students were recruited. Their ages ranged from 20 to 27 (25.3 ± 1.35). None of them had a history of visual or neurological disorders, head trauma, or any drug use that would affect nervous system function, and they were asked to wash their hair before the experiment. This experiment was approved by the Institutional Review Board at Fuzhou University. In accordance with the Helsinki Declaration of Human Rights, informed consent was obtained from all subjects before they joined this study.

2.2 Experimental protocol

In our CG experiment, each subject and his opponent independently decided to cooperate or aggress in 100 trials. There are four possible outcomes in each trial: both players cooperate (CC), the subject cooperates and his opponent aggresses (CA), the subject aggresses and his opponent cooperates (AC), both players aggress (AA). Every time, the subject was told that a student as his opponent was playing with him. In fact, the opponent's choices were determined by a pseudo-random sequence, in which the number of cooperation choices is approximately equal to that of aggression choices. The experiment design is to ensure roughly equal reinforcement rate for the two choices.

The procedure of our CG experiment is depicted in Fig. 1. Each trial started with the presentation of a fixation cross on the black screen lasting a variable period of 800-1000 ms. Then a payoff matrix was presented for 1500 ms, prompting the subject to make a decision. In the payoff matrix, 10/10 in the CC cell means that the subject and his opponent will both win 10 yuans if the outcome is CC, -10/30 in the CA cell represents that the subject will lose 10 yuans and his opponent will win 30 yuans if the outcome is CA, 30/-10 in the AC cell indicates that the subject will win 30 yuans and his opponent will lose 10 yuans if the outcome is AC, -30/-30 in the AA cell means that the subject and his opponent will both lose 30 yuans if the outcome is AA. The payoff matrix presents a social dilemma between a prosocial motive to maximize collective welfare and a self-interested motive to maximize personal welfare at a cost to the other person. The subjects made a decision of cooperation or aggression during the 1500 ms.

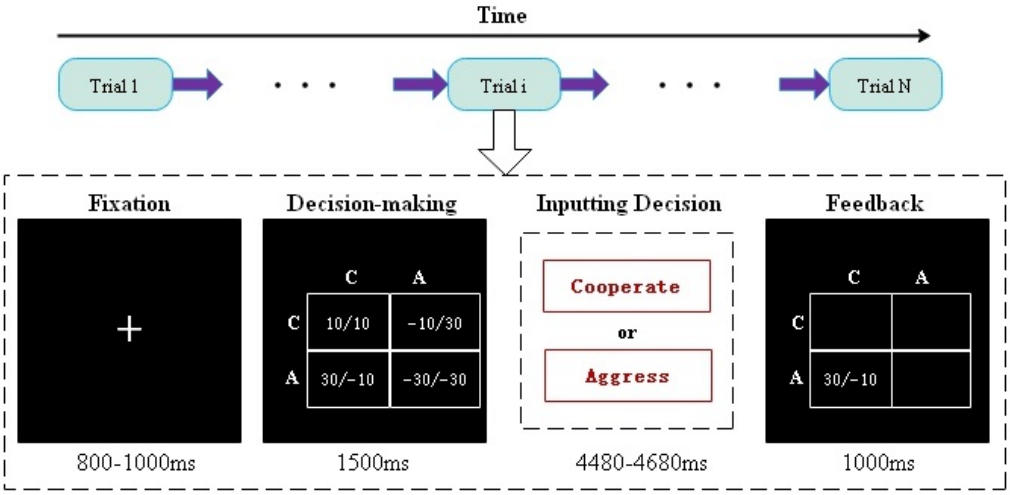


Fig. 1 Timeline of the CG experiment. Each trial includes four stages: Fixation, Decision-making, Inputting Decision and Feedback. In the Fixation stage, a cross is presented for a variable period of 800-1000 ms to concentrate the attentions of subjects. In the Decision-making stage of 1500 ms, the subjects decide to cooperate or aggress while observing the payoff matrix on the screen. In the Inputting Decision stage of 4480-4680 ms, the subjects input their decisions into the experiment platform by brain-computer interaction. In the Feedback stage, the outcome of the current trial is shown for 1000 ms.

Generally, subjects are guided to input their decisions by clicking a button after decision-making. Here, considering that the neural signals related to decision-making are possibly confounded by the neural activities related to starting responses [10], we introduced Code-Modulated Visual Evoked Potential (c-VEP) BCI [28, 29] and developed our experiment system based on BCI2000 [30]. Using the experiment system, the subjects expressed their decisions by gazing at 'C' (cooperation) or 'A' (aggression) flashing on the screen during the 4480-4680 ms period of inputting decision, which is detailed in Appendix B.

The feedback was presented to the subjects after the experiment system received the decision. For example, the feedback is 30/-10 (Fig. 1) when the subject decides to aggress and his opponent decides to cooperate.

2.3 EEG datasets

A 64-channel Neuroscan system, composed of a modified 10-20 system electrode cap, an amplifier and a signal acquisition software, was used to acquire EEG. The electrodes M2 (right mastoid) and AFz served as the reference and ground, respectively. All electrode sites were cleaned with alcohol, and the impedances between electrodes and scalp were maintained below 5 $k\Omega$. All EEG signals were recorded using a 0.05-100 Hz bandpass filter, continuously sampled at 1000 Hz, delivered from the signal acquisition software to our experiment system and saved to data files by the experiment system.

EEG signals were adjusted in order by bilateral mastoid re-reference, blink correction using ICA and common average reference, which were carried out on the platform of EEGLAB [31]. Subsequently, EEG signals were selected from the 800 ms before the onset of the decision-making screen to the end of the decision-making screen. Each trial had a 2300-ms EEG segment. The EEG segments in which EEG amplitudes exceeded a threshold of ± 80 μV were excluded. Therefore, the actual number of each subject's EEG segments was in the range of 84 to 100. Then, the EEG segments were corrected with the 800 ms before the onset of the decision-making screen as the baseline and labelled with the real decisions of the subjects. Finally, the EEG segments of each subject were randomly partitioned into three parts. Each part and the corresponding rest parts were in turn used to construct the test set and the training set for a subject, respectively. The numbers of the EEG segments in a training set and the corresponding test set were respectively around 60 and 30. For each subject, we built feature models and classifiers on a training set and tested the discriminability of single-trial EEG on the corresponding test set.

2.4 Framework based on machine learning

In order to explore the discriminability of single-trial EEG during decision-making in CG, we designed a research framework based on machine learning. As shown in Fig. 2, the research framework includes two phases: training and testing. In the training phase, a series of algorithms are applied on the training set of a subject to build feature models

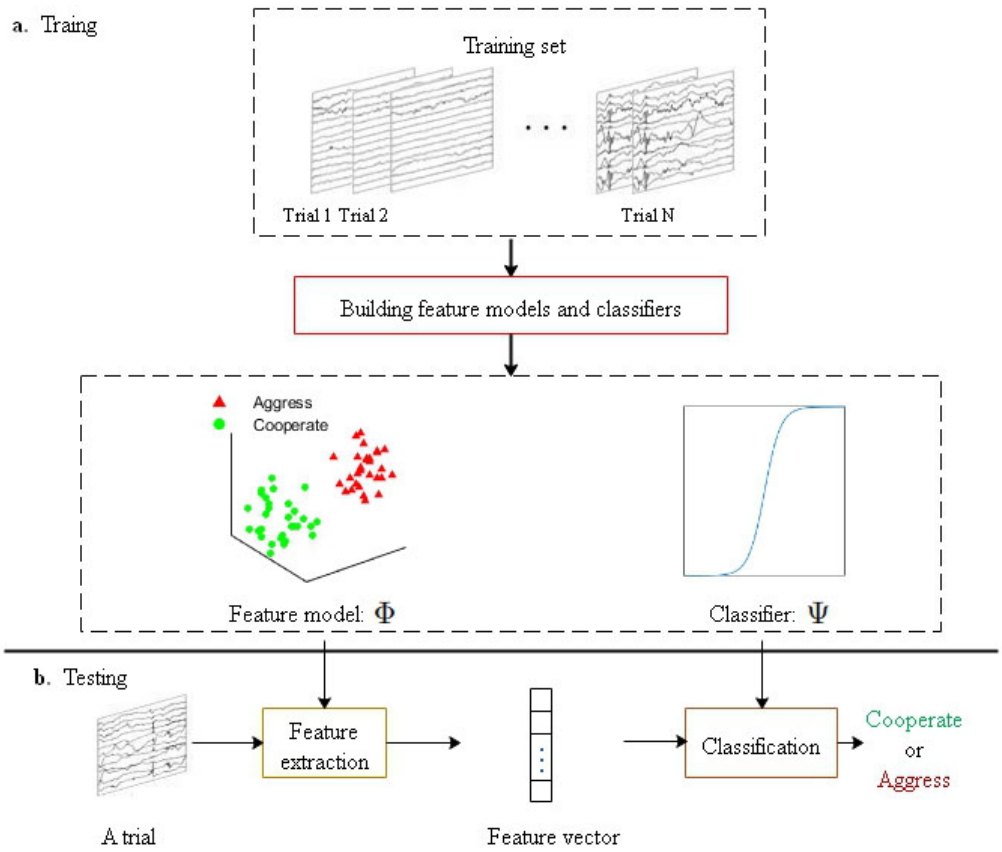


Fig. 2 The research framework based on machine learning. The 'a' section represents the training phase of this framework, during which the algorithms of extracting features and building classifiers are applied to construct feature models and classifiers. The 'b' section represents the testing phase of this framework, during which the feature models and classifiers are used to test the discriminability of single-trial EEG.

and classifiers. In the testing phase, the feature models and classifiers are used to discriminate the single-trial EEG during decision-making of cooperation or aggression.

The training phase is illustrated in Fig. 2a. A training set for a subject contains the EEG segments of N trials. We can describe the training set as $\{X_i, l_i\}_{i=1}^{N}$, where $X_i \in \mathbb{R}^{N_c \times N_t}$ represents an EEG segment, N_c and N_t are respectively the number of the channels and the number of the sampling points, $l_i \in \{+1, -1\}$ is the label of an EEG segment, $+1$ or -1 represent the subject's decisions of cooperation or aggression. A feature model is obtained by running a feature extraction algorithm on $\{X_i, l_i\}_{i=1}^{N}$. The feature model transforms $\{X_i, l_i\}_{i=1}^{N}$ into $\{x_i, l_i\}_{i=1}^{N}$, where $x_i \in \mathbb{R}^{N_v}$ represents a feature vector. A classifier is built by applying a classification algorithm on $\{x_i, l_i\}_{i=1}^{N}$.

We use $\Phi : X \to x$ and $\Psi : x \to l$ to respectively represent a feature model and a classifier. As shown in Fig. 2b, in the testing phase, an EEG segment $X \in \mathbb{R}^{N_c \times N_t}$ from a test set is transformed by a feature model Φ of the corresponding subject into a feature vector $x \in \mathbb{R}^{N_v}$, and the feature vector is further classified by a corresponding classifier Ψ as a label $l \in \{+1, -1\}$. The discriminability of single-trial EEG during decision-making in the CG experiment is observed by the classification performance.

In accordance with the existing works [3, 11, 15, 16], we employed the five existing feature extraction methods: DSNP, WM, ICA, PCA and CSP, and a classifier: LDA, in the framework. DSNP is an EEG feature extraction based on network patterns. WM means to transform EEG segments into feature vectors by averaging EEG signals in each time window (500 ms per window) on the concerned channels (F1, Fz, F2, CPz, PO7, PO8). Although PCA, ICA and CSP are different, they all learn a transformation matrix $H \in \mathbb{R}^{N_v \times N_c}$ from a training set. H is used to map EEG segments into N_v virtual channels by matrix multiplication. Due to $N_v \ll N_c$, the transformation matrix H is actually a filter searching the discriminability of EEG segments. For each EEG segment, the variances of each virtual channel are taken logarithmic and concatenated into a feature vector. Eighteen features was considered after empirical experiments using wrapper feature selection method, where all the possible combinations of features were assessed and

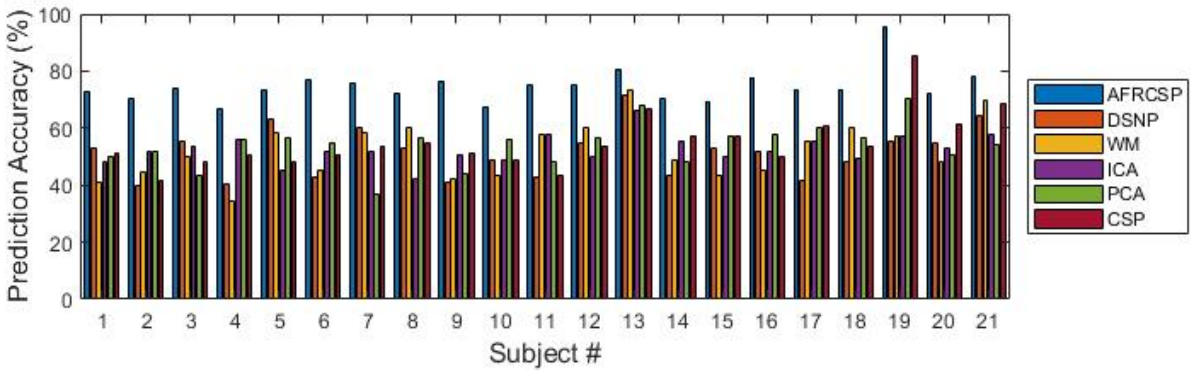


Fig. 3 The classification accuracies of AFRCSP, DSNP, WM, ICA, PCA, and CSP respectively combined with LDA on all 21 subjects

Table 1 Comparing the accuracies. The means and standard deviations of accuracies of each method and the t-test p values of comparing AFRCSP's accuracies with those of DSNP, WM, ICA, PCA, and CSP are are shown in the table.

	AFRCSP	DSNP	WM	ICA	PCA	CSP
Accuracy (mean±std%)	75±6	51±8	52±1	53±5	54±7	55±9
P value of t-test	None	6.47×10^{-11}	2.62×10^{-10}	2.66×10^{-13}	3.35×10^{-11}	2.57×10^{-11}

the combination of these 18 features achieved the best performance. To ensure the fairness of comparison, all implementations of Φ in this study transform EEG segments into 18-dimensional vectors, i.e., $N_v = 18$ for all feature extraction methods in the framework. As an implementation of Ψ, LDA can be expressed as $l = sign(w^T x + b)$, where $w \in \mathbb{R}^{N_v}$ and $b \in \mathbb{R}$ are obtained by running the shrinkage LDA algorithm [14, 32] on $\{x_i, l_i\}_{i=1}^{N}$.

2.5 Adaptive frequency regularized common spatial pattern

To explore our concern, we proposed a novel feature extraction method, called Adaptive frequency regularized common spatial pattern (AFRCSP), which is a new implementation of Φ. Essentially, AFRCSP is a combination of CSP [18], regularization idea [21, 23] and NCA [24–27]. The detailed description of AFRCSP is presented in Appendix C.

In this study, every subject carried out 100 trials. During the data preprocessing stage, a few EEG segments were excluded according to the specified threshold and only the $2/3$ EEG segments were added into the training sets. This means that the sample numbers of the training sets are around 60. In this small-sample setting, it is difficult for the conventional CSP training algorithm [18] to acquire good feature models. To deal with this issue, we introduced the regularization ideas of [21, 23] to improve the conventional CSP training algorithm.

On the other hand, the frequency band is also an important factor in this study. Although most studies only focus on the frequency band of 1-30 Hz, some researchers have revealed the associations of gamma-band (30-100Hz) EEG with cognitive functions [33–35], even with chronic psychotic disorders [36]. It is necessary to take a wider frequency band into account in this study. We established K frequency bands (K=33), which are 1-3 Hz, 4-6 Hz, $\cdots$, 97-99 Hz, to probe the associations of the frequency bands with decision-making cognitive activities. However, a wider frequency band will result in a higher dimension of feature vectors, complicating feature extraction of EEG. In order to effectively reduce the dimension of feature vectors, we introduced the neighborhood component analysis (NCA) [24–27] to filter the frequency bands without the significant discriminability.

3 Results

AFRCSP, DSNP, WM, ICA, PCA and CSP respectively combined with LDA were tested on the three test sets of all 21 subjects. Here, the mean accuracy, sensitivity and specificity of a method on the three test sets of a subject are called the accuracy, sensitivity and specificity of the method on the subject. To simplify, we use the name of a feature model to represent the combination of the feature model with LDA. For example, AFRCSP means the combination of AFRCSP with LDA. The accuracies are shown in Fig. 3. Intuitively, AFRCSP outperforms the other five methods. The t-test was conducted to check accuracy differences of AFRCSP with the other five methods. Table 1 shows the means and standard deviations of accuracies for each method and the p values of t-test of comparing AFRCSP's accuracies re-

Table 2 Comparing the sensitivities. The means and standard deviations of sensitivities of each method and the t-test p values of comparing AFRCSP's sensitivities with those of DSNP, WM, ICA, PCA, and CSP are shown in the table.

	AFRCSP	DSNP	WM	ICA	PCA	CSP
Sensitivity (mean±std)	0.72±0.11	0.51±0.17	0.41±0.36	0.54±0.14	0.54±0.14	0.57±0.17
P value of t-test	None	6.81×10^{-7}	2.56×10^{-4}	3.17×10^{-5}	1.23×10^{-7}	3.16×10^{-5}

Table 3 Comparing the specificities. The means and standard deviations of specificities of each method and the t-test p values of comparing AFRCSP's specificities with those of DSNP, WM, ICA, PCA, and CSP are shown in the table.

	AFRCSP	DSNP	WM	ICA	PCA	CSP
Specificity (mean±std)	0.73±0.15	0.47±0.15	0.58±0.37	0.49±0.14	0.52±0.15	0.51±0.18
P value of t-test	None	3.54×10^{-10}	2.36×10^{-2}	5.1×10^{-8}	2.13×10^{-7}	7.47×10^{-7}

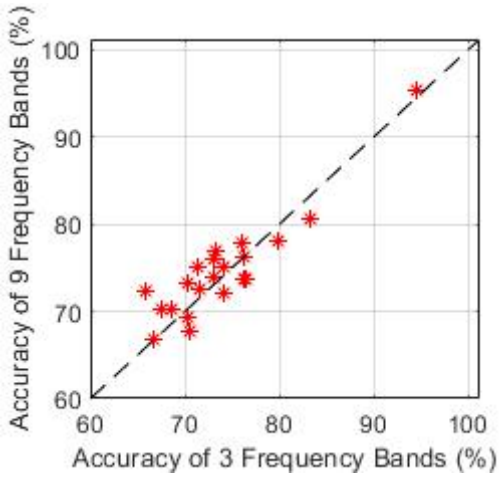


Fig. 4 The accuracy comparison of AFRCSP on 9 frequency bands versus on 3 frequency bands. The Y axis represents accuracy of AFRCSP on 9 frequency bands. The X axis represents accuracy of AFRCSP on 3 frequency bands. A point presents an accuracy comparison of AFRCSP on 9 frequency bands versus on 3 frequency bands for a subject.

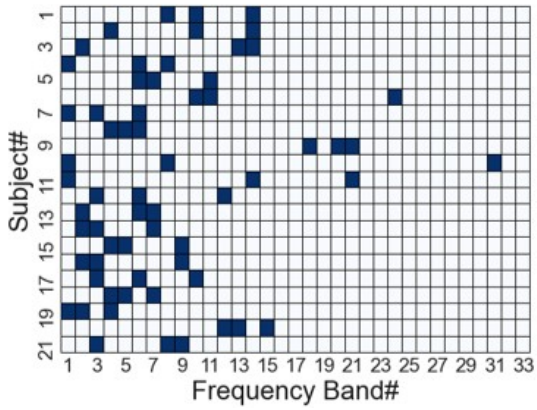


Fig. 5 The frequency bands selected by AFRCSP. The Y axis represents subjects. The X axis represents frequency bands. The frequency bands $1, 2, \cdots, 33$ indicate 1-3 Hz, 4-6 Hz, $\cdots$, 97-99 Hz. The three blue squares in each row represent the three most discriminative frequency bands selected by AFRCSP for the subject.

spectively with those of DSNP, WM, ICA, PCA and CSP. Furthermore, the means and standard deviations of sensitivities for each method and the t-test p values of comparing AFRCSP's sensitivities with those of DSNP, WM, ICA, PCA and CSP are shown in Table 2, and the means and standard deviations of specificities and the t-test p values of comparing specificities are also presented in Table 3 in same way. The t-test p values of AFRCSP versus DSNP, AFRCSP versus WM, AFRCSP versus ICA, AFRCSP versus PCA and AFRCSP versus CSP are far less than 0.05. At the significant level of 0.05, the accuracies, sensitivities and specificities of AFRCSP are different from those of DSNP, WM, ICA, PCA and CSP. The mean accuracy, sensitivity and specificity of AFRCSP are significantly greater than the counterparters of the other five methods. All statistical inferences on the performance data support the intuitive judgement that AFRCSP is clearly superior to DSNP, WM, PCA, ICA and CSP. The classification performance of AFRCSP shows that the discriminability of single-trial EEG during decision-making in chicken game really exists.

Fig. 4 presents the accuracy comparison of AFRCSP on 9 frequency bands versus on 3 frequency bands. In Fig. 4, most points approach the diagonal, showing that the accuracies of AFRCSP on 9 frequency bands are approximately equal to those of AFRCSP on 3 frequency bands for most subjects. The t-test was carried out to check accuracy differences of AFRCSP on 9 frequency bands versus on 3 frequency bands. The p value is 1.44×10^{-2}. At the significant level of 0.01, no significant accuracy difference of AFRCSP on 9 frequency bands versus on 3 frequency bands is discovered.

The three most discriminative frequency bands selected by AFRCSP for each subject are shown in Fig. 5. For most subjects, the three most discriminative frequency bands are

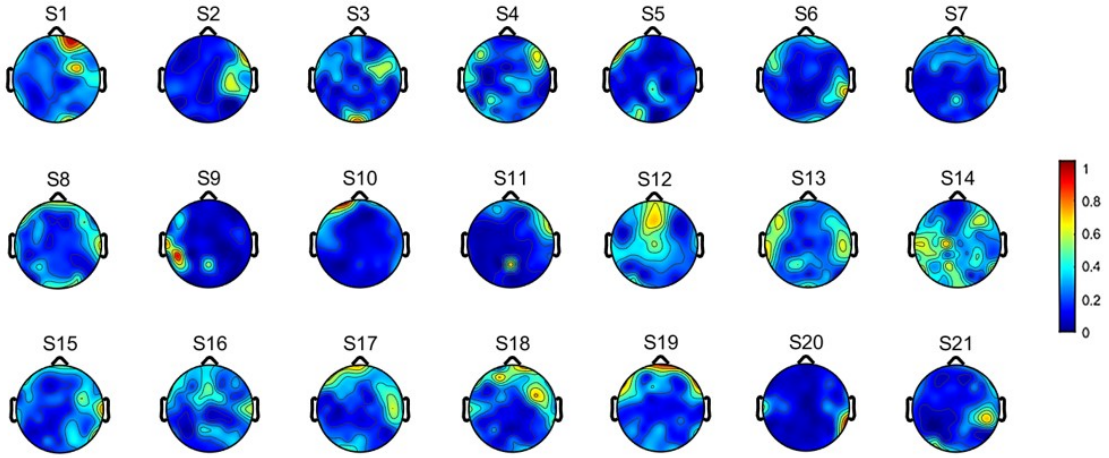


Fig. 6 The brain maps of the 21 subjects: S1, S2, $\cdots$, S21. The intensity represented by the color in the brain maps indicates the contribution of the brain region to the discriminablity discovered by AFRCSP.

below 40 Hz. Only the subjects 1, 2, 3, 6, 9, 10, 11 and 20 have the discriminative frequency bands above 40 Hz. Even so, nearly all of them have the discriminative frequency bands below 40 Hz.

For each subject, the row vectors of the transformation matrices of AFRCSP in the three most discriminative frequency bands were taken absolute values, normalized to [0,1] and then averaged to an intensity vector, each dimension of which corresponds to a channel. The brain maps in Fig. 6 were drawn using the subjects' own intensity vectors. The intensity represented by the color in Fig. 6 indicates the contribution of the brain region to the discriminablity discovered by AFRCSP. In Fig. 6, the significant intensities exist in the medial prefrontal regions of S1, S3, S4, S6, S7, S12, S13, S14, S15, S16, S17, S18, and S19, in the dorsolateral prefrontal regions of S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S13, S14, S15, S16, S17, S18, and S19, in the orbitofrontal regions of S1, S3, S5, S6, S7, S8, S10, S11, S12, S13, S15, S16, S17, S18, and S19, in the middle intraparietal regions of S1, S3, S4, S5, S7, S8, S9, S11, S12, S13, S14, S15, S16, and S19, and in the occipital regions of all subjects except S9, S10, S11, S20.

4 Disccusion

The results of Fig. 3 and Table 1-3 show two points: 1. the discriminability of single-trial EEG between the two decisions in our CG experiments really exists and 2. the discriminability can be clearly revealed by AFRCSP. In order to explore which frequency bands and brain regions contribute to the discriminability, we further analysed the results of AFRCSP.

Firstly, we compared the classification accuracies of AFRCSP on 9 frequency bands versus on 3 frequency bands. The comparison is presented by Fig. 4, and the t-test result shows that no significant classification accuracy difference exists between the two situations. This means that the single-trial EEG discriminability revealed in this study is mainly attributed to the three most discriminative frequency bands selected by AFRCSP. Therefore, we can concentrate on the three most discriminative frequency bands selected by AFRCSP to probe the impacts of frequency bands on the discriminability.

As shown in Fig. 5, among the 21 subjects, the three most discriminative frequency bands of 14 subjects are all below 40 Hz, the other seven subjects also have the discriminative frequency bands below 40 Hz although a few of their discriminative frequency bands lie in the range between 40 Hz and 99 Hz. This means that the single-trial EEG discriminability originates mainly from the frequency range of 1-40 hz. The studies [33–36] show that the gamma-band of EEG is also associated with cognitive functions. Maybe, due to filtering of skull or instable measurement conditions on scalp, the associations are not frequently observed.

Secondly, we explored the contributions of different brain regions to the discriminability.

The existing studies [1–3] show that the brain regions involving decision-making commonly include the dorsolateral prefrontal cortex, medial prefrontal cortex, orbitofrontal cortex, superior temporal sulcus, middle intraparietal sulcus, insula, anterior cingulate cortex, posterior cingulate cortex, ventral striatum, and amygdala. During decision-making in the CG experiment, the differences of cognitive activities in the brain functional areas are probably reflected by the EEG difference in the corresponding regions. Fig. 6 shows that AFRCSP discovered the contributions to single-trial

EEG discriminability in the medial prefrontal regions, in the dorsolateral prefrontal regions, in the orbitofrontal regions and in the middle intraparietal regions. The discoveries of AFRCSP are in agreement with the existing knowledge on the cognitive mechanism of decision-making [1–3].

AFRCSP also discovered the single-trial EEG discriminability in the occipital regions. We believe that the discriminability in the occipital regions embodies the participation of the subjects' visual function during the 'Decision-making' stage shown in Fig. 1. This finding of AFRCSP is consistent with our experimental scenario.

Conclusions

In the studies on the cognitive mechanism of decision-making, researchers often probe the neural activities that drive decision-making by observing the discriminability of EEG during decision-making. Usually, the discriminability is observed by averaging EEG signals across trials. While observing the discriminability of single-trial EEG is also very valuable, but it is rarely carried out owing to its difficulty. This study explored the discriminability of single-trial EEG during decision making of cooperation or aggression using a specific machine learning framework. The classification accuracy of around 75% is achieved in the differenciation of single-trial EEG between the cooperation and aggression decisions during the CG experiments. Results demonstrate that it is feasible to apply machine learning to assessing the discriminability of single-trial EEG.

This study can conclude that: 1. the discriminability of single-trial EEG during decision-making in the chicken game really exists, and 2. the discriminability originates mainly from the EEG information below 40 Hz. The brain-region analysis based on the AFRCSP features shows that the contributions of different brain regions to single-trial EEG discriminability are consistent with the existing knowledge on the cognitive mechanism of decision-making. This further confirms the reliability of our conclusions. The findings in the study can be used to develop new BCI paradigms, in which subjects' decisions can be recognized online.

Acknowledgements

This work was supported by the Transformation Project of Scientific and technological achievements of Fuzhou, China (2020-GX-12) and Natural Science Foundation of Fujian Province,China (2019J01242). Faqiang Peng participated in this work when he studied at Fuzhou University. We thank him for his contribution.

A Chicken game

The CG experimental task aims to study the mechanisms underlying cooperative and aggressive behaviors. In this game, two players independently decide to cooperate or aggress. Every time, the payoff for each player depends on the combination of two players' decisions. The payoffs of mutual cooperation and mutual aggression are represented with 'R' and 'P', respectively. If one player cooperates while his/her opponent aggresses, the payoffs of the player and his/her opponent are represented with 'S' and 'T', respectively. The payoffs are arranged such that T>R>S>P and 2R≥T+S. The players make their decisions in a social dilemma when they are conducting the CG experimental task.

B The period of inputting decision

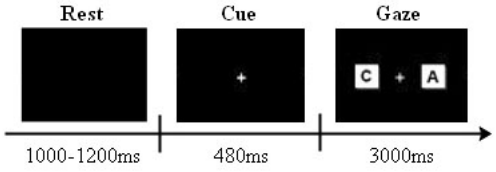


Fig. 7 Inputting decision by brain-computer interaction

During the period of inputting decision, the subjects expressed their decisions by a c-VEP BCI. The procedure is shown in Fig. 7. The 'Rest' screen guides subjects to rest for 1000-1200 ms. The 'Cue' screen prompts subjects to get ready to input their decisions. The 'Gaze' screen represents the brain-computer interaction lasting 3000 ms. On the screen, the 'A' and 'C' rectangles means 'aggress' and 'cooperate', respectively. The subjects are instructed to gaze at one of the two rectangles according to their own decisions. The two rectangles flash in a pseudorandom pattern specified respectively by 15-bit m-sequence '000010100110111' and its shift '110111000010100'. Each bit of m-sequence corresponds to 40 ms, meaning on by 1 or off by 0 and the flashes of 15-bit m-sequence are repeated 5 times. For each trial, the EEG signals recorded from the electrodes of O1, Oz, O2, PO7, POz, and PO8 during the 3000 ms are transformed to a feature vector, which is further recognised as the subject's decision by a linear discriminant analysis classifier.

C The description of AFRCSP

AFRCSP establishes a feature model Φ based on the K frequency bands for a subject (target subject) using not only his own training set but also the training sets of other subjects (source subjects). Firstly, AFRCSP filters all EEG segments of each training set with the K bandpass filters. Secondly, AFRCSP processes each frequency band in same way to obtain a transformation model for each frequency band. Thirdly, AFRCSP selects the most informative frequency bands for a subject. Finally, AFRCSP integrates the transformation models of the selected frequency bands into a feature model of the subject.

For the simplicity, we still use X_i to represent the EEG segment of a trial in a frequency band. The procedure of AFRCSP processing a frequency band for a targrt subject can be depicted as follows.

Step-1 calculates the regularized covariance matrices. The covariance matrix of a trial is obtained according to Eq. 1,

$$C_i = \frac{X_i X_i^T}{tr(X_i X_i^T)}, \quad i \in [1, N] \tag{1}$$

where $tr(\cdot)$ represents solving the trace of a matix. The covariance matrices of cooperation and aggression are calculated according to Eqs. 2 and 3,

$$C_c = \frac{1}{M_c} \sum_{i \in \mathbb{T}_c} C_i \tag{2}$$

$$C_a = \frac{1}{M_a} \sum_{i \in \mathbb{T}_a} C_i \tag{3}$$

where $\mathbb{T}_c$ and $\mathbb{T}_a$ respectively represents all trials of the cooperation decision and all trials of the aggression decision and $M_c = |\mathbb{T}_c|$, $M_a = |\mathbb{T}_a|$. For each source subject, the covariance matrices of cooperation and aggression can be calculated in the same way. We notate the mean covariance matrices of cooperation and aggression for all source subjects as $\bar{C}_c$ and $\bar{C}_a$. Then, we define the regularized covariance matrices of the target subject as $\breve{C}_c$, $\breve{C}_a$ and $\breve{C}$ by Eqs. 4-8

$$\hat{C}_c = (1-\beta)C_c + \beta\bar{C}_c \tag{4}$$

$$\hat{C}_a = (1-\beta)C_a + \beta\bar{C}_a \tag{5}$$

$$\breve{C}_c = (1-\gamma)\hat{C}_c + \frac{\gamma}{N_c} tr(\hat{C}_c) I \tag{6}$$

$$\breve{C}_a = (1-\gamma)\hat{C}_a + \frac{\gamma}{N_c} tr(\hat{C}_a) I \tag{7}$$

$$\breve{C} = \breve{C}_c + \breve{C}_a \tag{8}$$

where the parameters $\beta, \gamma \in [0, 1]$ are determined by cross-validation during the implementation, I is a $N_c \times N_c$ identity matrix, $tr(\cdot)$ represents the operation of the trace of a matrix.

Step-2 conducts the Eigendecomposition of $\breve{C}$ (Eq. 9),

$$\breve{C} = B\lambda B^T \tag{9}$$

where λ is the diagonal matrix of eigenvalues of $\breve{C}$ and B is a matrix composed of normalized eigenvectors of $\breve{C}$. We then get a matrix W by $W = \lambda^{-1/2} B^T$, and transform $\breve{C}_c$ and $\breve{C}_a$ to S_c and S_a by $S_c = W\breve{C}_c W^T$ and $S_a = W\breve{C}_a W^T$. In accordance with [17, 18], we know that S_c and S_a have the same eigenvectors. This can be described as Eqs. 10 and 11,

$$S_c = U\psi_c U^T \tag{10}$$

$$S_a = U\psi_a U^T \tag{11}$$

where U is the common eigenvector matrix of S_c and S_a, ψ_c and ψ_a are diagonal matrices of eigenvalues of S_c and S_a respectively, and the sum of ψ_c and ψ_a is an identity matrix.

Step-3 computes the transformation matrix P by Eq. 12,

$$P = (\hbar(U))^T W \tag{12}$$

where $\hbar(\cdot)$ represents selecting the first and last columns of U after sorting the eigenvectors in descending order of the eigenvalues.

We use $k \in \{1, \cdots, K\}$ to represent a frequency band and then can represent the transformation matrices and EEG segments of the K frequency bands with $P_k \in \mathbb{R}^{2 \times N_c}$ and $X_i^k \in \mathbb{R}^{N_c \times N_t}$. Furthermore, for all $i \in \{1, \cdots, N\}$, $k \in \{1, \cdots, K\}$, $Y_i^k \in \mathbb{R}^{2 \times N_t}$ and $h_i^k \in \mathbb{R}^2$ are obtained respectively by Eq. 13 and by Eq. 14.

$$Y_i^k = P_k X_i^k \tag{13}$$

$$h_i^k = \ln \frac{diag(Y_i^k (Y_i^k)^T)}{tr(Y_i^k (Y_i^k)^T)} \tag{14}$$

As the counterpart of X_i, h_i is obtained by concatenating all h_i^k for $k \in \{1, \cdots, K\}$. But, $h_i \in \mathbb{R}^{2K}$ are not yet the appropriate feature vectors of X_i, since the information from the frequency bands in which no significant discriminability exists is included in h_i.

The NCA [24–27] is used for further feature selections. We define a weight vector $w \in \mathbb{R}^{2K}$, and denote the weighted distance between h_i and h_j by Eq. 15.

$$D_w(h_i, h_j) = \sum_{k=1}^{2K} w_k^2 |h_{ik} - h_{jk}| \tag{15}$$

The probability of h_i selecting h_j as its reference point is defined as Eq. 16,

$$p_{ij} = \begin{cases} \frac{\kappa(D_w(h_i, h_j))}{\sum_{k \neq i} \kappa(D_w(h_i, h_k))}, & if \quad i \neq j \\ 0, & if \quad i = j \end{cases} \tag{16}$$

where $\kappa(z) = exp(-z/\sigma)$ is a kernel function with a parameter σ. Then, the objective function is obtained by Eq. 17,

$$\xi(w) = \sum_{i=1}^{N} \sum_{j=1}^{N} l_{ij} p_{ij} - \alpha \sum_{k=1}^{2K} w_k^2 \tag{17}$$

where $l_{ij} = 1$ if and only if $l_i = l_j$, or $l_{ij} = 0$ otherwise; $\alpha > 0$ is a regularization parameter which is set to $1/N$ here. Our goal is to maximize $\xi(w)$ with respect to w. In the process of solving this problem, the regularization term drives many of the weights in w to 0. After obtaining the solution w, we notate the subscript set of the weights in w that are much greater than 0 as $\mathbb{L} = \{k \,|\, w_k \gg 0\}$ and construct x_i, the feature vector of X_i, by selecting the features $\{h_{ik} | k \in \mathbb{L}\}$ and concatenating them.

The above descriptions are the implementation of AFRCSP to Φ. As a combination of CSP, regularization idea and NCA, AFRCSP meets our need to probe the discriminability of EEG segments in a broad frequency range and in small-sample setting.

D The results of SVM and KNN

A reviewer thinks that the results of Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) should be reported in the paper. We show them in Table 4.

References

1. A. G. Sanfey, Social decision-making: Insights from game theory and neuroscience, Science 318 (5850) (2007) 598–602.
2. U. Basten, G. Biele, H. R. Heekeren, C. J. Fiebach, How the brain integrates costs and benefits during decision making, Proceedings of the National Academy of Sciences 107 (50) (2010) 21767–21772.
3. Y. Si, X. Wu, F. Li, L. Zhang, K. Duan, P. Li, L. Song, Y. Jiang, T. Zhang, Y. Zhang, J. Chen, S. Gao, B. Biswal, D. Yao, P. Xu, Different decision-making responses occupy different brain networks for information processing: A study based on EEG and TMS, Cerebral Cortex 29 (10) (2018) 4119–4129.
4. R. Hastie, Problems for judgment and decision making, Annual Review of Psychology 52 (1) (2001) 653–683.
5. C. Cecchetto, S. Korb, R. I. Rumiati, M. Aiello, Emotional reactions in moral decision-making are influenced by empathy and alexithymia, Social Neuroscience 13 (2) (2018) 226–240.
6. N. Preuss, L. S. Brändle, O. M. Hager, M. Haynes, U. Fischbacher, G. Hasler, Inconsistency and social decision making in patients with borderline personality disorder, Psychiatry Research 243 (2016) 115 – 122.
7. J. Decety, P. L. Jackson, J. A. Sommerville, T. Chaminade, A. N. Meltzoff, The neural bases of cooperation and competition: an fMRI investigation, Neuroimage 23 (2) (2004) 744–751.

Table 4 The results of the combinations of SVM and KNN with AFRCSP, DSNP, WM, ICA, PCA and CSP (mean±std)

		AFRCSP	DSNP	WM	ICA	PCA	CSP
SVM	accuracy	0.74±0.06	0.51±0.09	0.53±0.10	0.54±0.06	0.54±0.08	0.55±0.09
	sensitivity	0.75±0.14	0.52±0.18	0.44±0.39	0.57±0.19	0.57±0.21	0.61±0.17
	specificity	0.69±0.16	0.46±0.17	0.55±0.39	0.48±0.20	0.48±0.22	0.47±0.20
KNN	accuracy	0.74±0.07	0.51±0.09	0.51±0.09	0.52±0.07	0.53±0.09	0.55±0.10
	sensitivity	0.73±0.15	0.51±0.20	0.60±0.42	0.51±0.20	0.53±0.22	0.59±0.21
	specificity	0.70±0.17	0.48±0.19	0.40±0.42	0.49±0.19	0.49±0.22	0.49±0.22

8. C. H. Declerck, C. Boone, The neuroeconomics of cooperation, Nature Human Behaviour 2 (7) (2018) 438–440.
9. H. Fukui, T. Murai, J. Shinozaki, T. Aso, H. Fukuyama, T. Hayashi, T. Hanakawa, The neural basis of social tactics: An fMRI study, Neuroimage 32 (2) (2006) 913–920.
10. Y. Wang, Y. Lin, C. Fu, Z. Huang, S. Xiao, R. Yu, Effortless retaliation: the neural dynamics of interpersonal intentions in the chicken game using brain-computer interface, Social Cognitive and Affective Neuroscience 16 (11) (2021) 1138–1149.
11. Y. Si, F. Li, K. Duan, Q. Tao, C. Li, Z. Cao, Y. Zhang, B. Biswal, P. Li, D. Yao, et al., Predicting individual decision-making responses based on single-trial EEG, NeuroImage 206 (2020) 116333.
12. C. F. Camerer (Ed.), Behavioral Game Theory: Experiments in Strategic Interaction, Princeton University Press, Princeton,USA, 2003.
13. W. Su, M. N. Potenza, Z. Zhang, X. Hu, L. Gao, Y. Wang, Do individuals with problematic and non-problematic internet game use differ in cooperative behaviors with partners of different social distances? Evidence from the Prisoner's Dilemma and Chicken Game, Computers in Human Behavior 87 (2018) 363–370.
14. B. Blankertz, S. Lemm, M. Treder, S. Haufe, K. R. Müller, Single-trial analysis and classification of ERP components - A tutorial, Neuroimage 56 (2) (2011) 814–825.
15. F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, B. Arnaldi, A review of classification algorithms for EEG-based brain–computer interfaces, Journal of Neural Engineering 4 (2007) 24.
16. F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, F. Yger, A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update, Journal of Neural Engineering 15 (3) (2018) 031005.
17. Z. J. Koles, The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG, Electroencephalography and Clinical Neurophysiology 79 (6) (1991) 440.
18. Z. J. Koles, J. C. Lind, A. C. Soong, Spatio-temporal decomposition of the EEG: a general approach to the isolation and localization of sources, Electroencephalography and Clinical Neurophysiology 95 (4) (1995) 219.
19. H. Ramoser, J. Müller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, IEEE Transactions on Rehabilitation Engineering 8 (4) (2000) 441–446.
20. B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K. Müller, Optimizing spatial filters for robust EEG single-trial analysis, IEEE Signal Processing Magazine 25 (1) (2008) 41–56.
21. F. Lotte, C. Guan, Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms, IEEE Transactions on Biomedical Engineering 58 (2) (2011) 355–362.
22. K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, H. Zhang, Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b, Frontiers in Neuroscience 6 (2012) 39.
23. S.-H. Park, S.-G. Lee, Small sample setting and frequency band selection problem solving using subband regularized common spatial pattern, IEEE Sensors Journal 17 (10) (2017) 2977–2983.
24. J. Goldberger, G. E. Hinton, S. T. Roweis, R. R. Salakhutdinov, Neighbourhood components analysis, in: Advances in Neural Information Processing Systems, 2005, pp. 513–520.
25. W. Yang, K. Wang, W. Zuo, Neighborhood component feature selection for high-dimensional data., Journal of Computers 7 (1) (2012) 161–168.
26. S. Raghu, N. Sriraam, Classification of focal and non-focal eeg signals using neighborhood component analysis and machine learning algorithms, Expert Systems with Applications 113 (2018) 18–32.
27. O. W. Samuel, B. Yang, Y. Geng, M. G. Asogbon, S. Pirbhulal, D. Mzurikwao, O. P. Idowu, T. J. Ogundele, X. Li, S. Chen, G. R. Naik, P. Fang, F. Han, G. Li, A new technique for the prediction of heart failure risk driven by hierarchical neighborhood component-based learning and adaptive multi-layer networks, Future Generation Computer Systems 110 (2020) 781–794.
28. J. R. Wolpaw, E. W. Wolpaw (Eds.), Brain-Computer Interfaces: Principles and Practice, Oxford University Press, New York, USA, 2012.
29. G. Bin, X. Gao, Y. Wang, B. Hong, S. Gao, VEP-based brain-computer interfaces: time, frequency, and code modulations [research frontier], IEEE Computational Intelligence Magazine 4 (4) (2009) 22–26.
30. G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, J. R. Wolpaw, BCI2000: a general-purpose brain-computer interface (BCI) system, IEEE Transactions on Biomedical Engineering 51 (6) (2004) 1034–1043.
31. A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis, Journal of Neuroscience Methods 134 (1) (2004) 9–21.
32. J. H. Friedman, Regularized discriminant analysis, Journal of the American Statistical Association 84 (405) (1989) 165–175.
33. J. Kaiser, W. Lutzenberger, Human gamma-band activity: a window to cognitive processing, Neuroreport 16 (3) (2005) 207–211.
34. M. Kucewicz, J. Cimbalnik, J. Matsumoto, B. Brinkmann, M. R. Bower, V. Vasoli, V. Sulc, F. Meyer, W. R. Marsh, S. Stead, G. Worrell, High frequency oscillations are associated with cognitive processing in human recognition memory, Brain 137 (Pt 8) (2014) 2231–2244.
35. R. F. Stevenson, J. Zheng, L. Mnatsakanyan, S. Vadera, R. T. Knight, J. J. Lin, M. A. Yassa, Hippocampal ca1 gamma power predicts the precision of spatial memory judgments, Proceedings of the National Academy of Sciences 115 (40) (2018) 10148–10153.
36. T. J. Reilly, J. F. Nottage, E. Studerus, G. Rutigliano, A. I. De Micheli, P. Fusar-Poli, P. McGuire, Gamma band oscillations in the early phase of psychosis: a systematic review, Neuroscience & Biobehavioral Reviews 90 (2018) 381–399.

Biography

Zhihua Huang received PhD in 2012 from Xiamen University and has been a Professor with Fuzhou University since 2014. His research interests include brain-computer interface, brain signal processing, machine learning, and computer software.

Kun Jiang received BS from the Xiamen University of Technology in 2019. He is pursuing ME in software engineering with Fuzhou University. His research interests include brain-computer interface, machine learning, and biometrics.

Jing Li received BS in network engineering from the Wuhan Institute of Technology in 2017. She is pursuing ME in Software Engineering with Fuzhou University. Her research interests include brain-computer interface and deep learning.

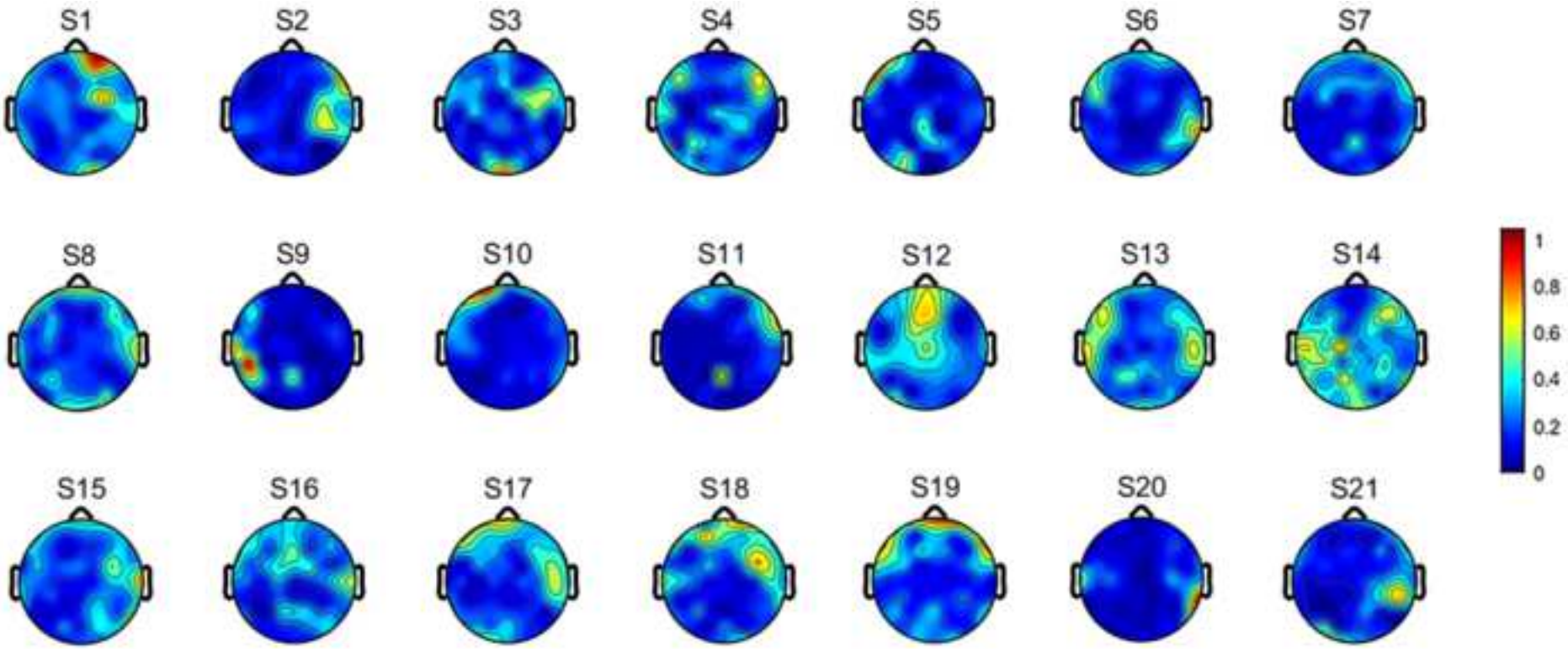
Wenxing Zhu received PhD from Shanghai University in 1996. He has been a Professor with Fuzhou University since 2004. His research interests include optimization theory, algorithms and applications.

Huiru Zheng received PhD from Ulster University in 2003 and is a Professor with the School of Computing at Ulster University. Her research areas include machine learning, integrative data analysis, digital health, and computational biology.

Yiwen Wang received PhD in Psychology from Beijing Normal University in 2004 and is a distinguished Professor at Fuzhou University. His researches areas include social cognition neuroscience, decision neuroscience, and social psychology.

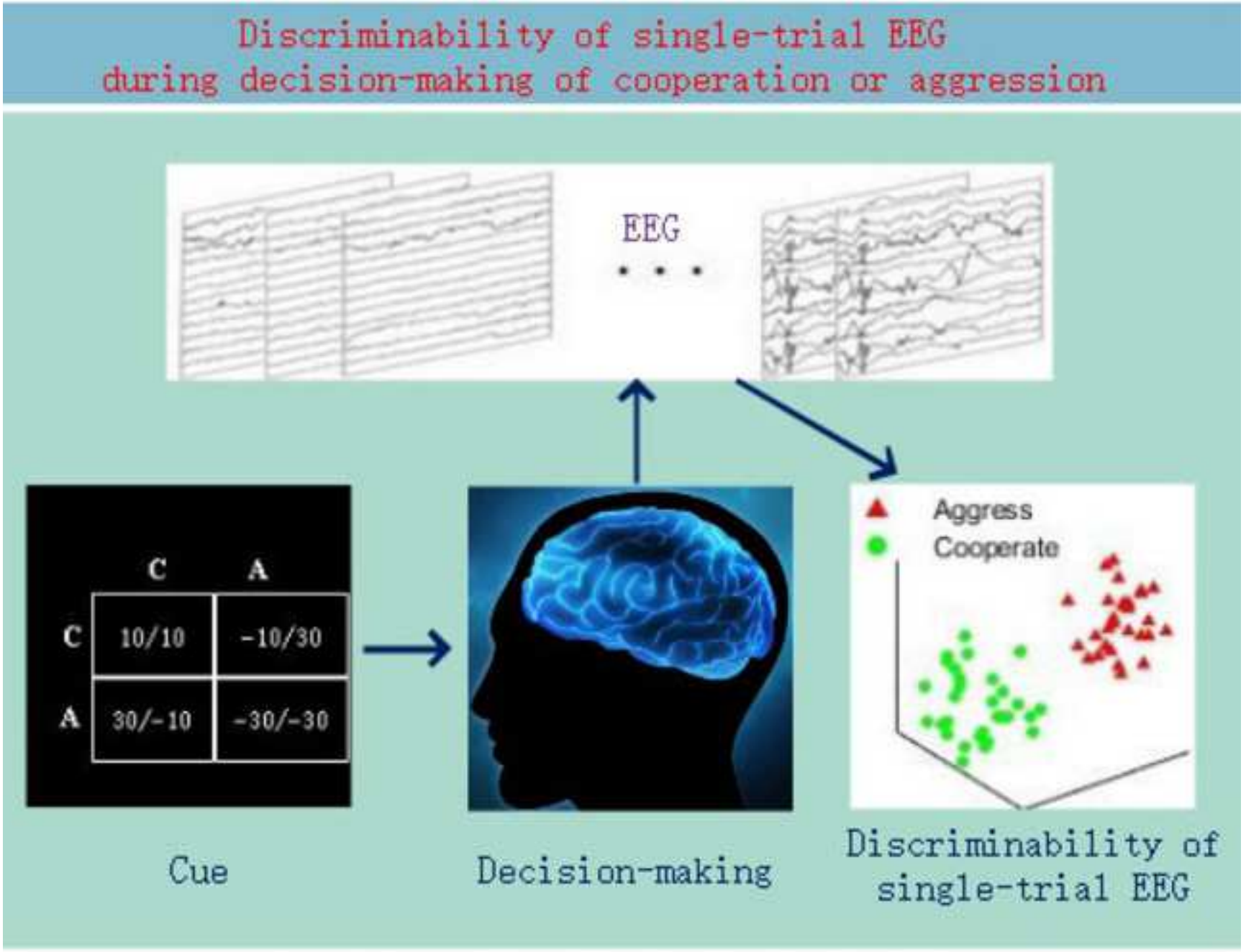
Figure

Click here to access/download;Figure;brainmap2.png



Responses to the reviewers

Click here to view linked References



An attempt to discriminate decisions