Optimized Collaborative Brain-Computer Interfaces for Enhancing Face Recognition

Cecilia Salvatore, Davide Valeriani, Veronica Piccialli, and Luigi Bianchi, *Member, IEEE*

***Abstract*— The aim of this study is to maximize group decision performance by optimally adapting EEG confidence decoders to the group composition. We train linear support vector machines to estimate the decision confidence of human participants from their EEG activity. We then simulate groups of different size and membership by combining individual decisions using a weighted majority rule. The weights assigned to each participant in the group are chosen solving a small-dimension, mixed, integer linear programming problem, where we maximize the group performance on the training set. We therefore introduce optimized collaborative brain-computer interfaces (BCIs), where the decisions of each team member are weighted according to both the individual neural activity and the group composition. We validate this approach on a face recognition task undertaken by 10 human participants. The results show that optimal collaborative BCIs significantly enhance team performance over other BCIs, while improving fairness within the group. This research paves the way for practical applications of collaborative BCIs to realistic scenarios characterized by stable teams, where optimizing the decision policy of a single group may lead to significant long-term benefits of team dynamics.**

***Index Terms*— Brain-computer interfaces, decision-making, electroencephalography, face recognition, machine learning.**

I. Introduction

A. Group Decision Making

IMPORTANT decisions are often made in groups, leveraging their superior cognition and capabilities (wisdom of crowds [1]). A fundamental question in group decision-making is how to integrate individual judgments to obtain a group decision. A popular approach in this context is to use majority voting, which is backed by the Condorcet Jury Theorem and has proven to enable groups to be more accurate than individuals in a variety of domains [2]. Yet, majority voting is often sub-optimal, given that individual judgments are typically not independent, and are affected by both false positives and false negatives [3].

A way to achieve optimal group decision making is to weigh individual judgments by the confidence of the contributor [4]. This approach may allow to correct group decisions even when the majority of the team members made an erroneous choice [5]. Ideally, the decision confidence represents the probability of that decision being correct [6]. However, subjective confidence judgments are often inaccurate, as several personality and external factors affect the ability of humans in correctly estimating their own confidence [7], [8].

B. Face Recognition

Face recognition is the ability of identifying a target face. This is critical in our daily lives to recognize people, as well as in security and surveillance to enhance safety [9]. The importance of this task is recognized by our brain having dedicated regions to process faces [10]. In many scenarios, humans find face recognition very challenging, even when highly-skilled [11], [12].

Face recognition is also particularly challenging for artificial intelligence (AI) systems. Recent studies have shown that AIs are significantly more accurate than the average human [13], but no better than the best humans [12]. Notably, machines fail in face recognition for different reasons than humans. For example, while limitations of the human visual system make us miss giant targets [14], machines perform poorly in less constrained situations involving moving targets [15], [16].

Recent research has shown that the best group decisions in face recognition are made when human and AI agents work together. For example, Phillips and colleagues showed how single forensic facial examiners combined with optimal AIs were more accurate than the combination of two human examiners [12]. Similar results were obtained when combining the decisions of novice humans with those of an AI system, taking into account the confidence of each decision maker [13]. To achieve effective fusion of human and AI decisions in face recognition, it is critical to precisely measure the reliability of each agent, for example, through the decision confidence.

C. Optimal Brain-Computer Interfaces

Brain-Computer Interfaces (BCIs) are devices typically used to restore communication and motor capabilities in people with severe disabilities [17]. In the last few decades, BCIs have

Manuscript received October 12, 2021; revised April 5, 2022; accepted May 3, 2022. Date of publication May 5, 2022; date of current version May 16, 2022. *(Cecilia Salvatore and Davide Valeriani contributed equally to this work.) (Corresponding author: Davide Valeriani.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Board of the University of Essex.

Cecilia Salvatore and Luigi Bianchi are with the Department of Civil Engineering and Computer Science, University of Tor Vergata, 00133 Rome, Italy (e-mail: cecilia.salvatore@uniroma2.it; luigi.bianchi@uniroma2.it).

Davide Valeriani is with Neurable Inc., Boston, MA 02108 USA (e-mail: davide.valeriani@gmail.com).

Veronica Piccialli is with the Department of Computer, Control and Management Engineering Antonio Ruberti, Sapienza University of Rome, 00185 Rome, Italy (e-mail: veronica.piccialli@uniroma1.it).

Digital Object Identifier 10.1109/TNSRE.2022.3173079

also expanded their application domains to human augmentation [18], gaming [19], and brain monitoring [20]. In the context of decision-making, BCIs have demonstrated to be a critical tool to improve the correctness of our decisions. For example, BCIs can decode the decision of multiple participants with better accuracy and faster than single non-BCI users [21]–[23]. Moreover, this collaborative BCI approach has also been applied to traditional BCI paradigms, such as motor imagery [24], [25], to boost performance. These BCIs often rely on single event-related potentials (ERPs), e.g., P300, to discriminate between different decisions. However, the combination of tracking and fusing information from multiple ERPs further increases BCI decoding performance [26].

Recent research has shown that BCIs can also improve group decision-making performance. In particular, BCIs could use machine learning [27] to decode more calibrated confidence estimates directly from the electroencephalographic (EEG) brain signals of the participants. When using BCI confidence to weigh individual opinions, groups become more accurate than equally-sized groups based on standard majority or on subjective confidence judgments in a variety of decision-making tasks, including visual search [28] and face recognition [13]. These BCI decoders rely on neural correlates of confidence that are common across tasks and people [29] and are extracted using spatio-temporal transformations of multi-electrode EEG activity, such as common spatial patterns [30]. Yet, in these studies, a different BCI was trained for each user to promote decoding accuracy and user training [31].

This approach of optimizing a BCI on each user is particularly common in P300 spellers, where participants select letters from a grid by focusing on it, so that their brain elicits different patterns for the target (rare) stimuli as opposed to non-target (frequent) ones. However, several stimuli must be provided and processed to select one single letter to spell, as the signal-to-noise ratio of the recorded EEG neural signal of interest is low. Alternatively, early stopping techniques can be used to detect when the decoder is confident enough about its predictions and stop the stimulation. A reliability score can be computed after a preliminary calibration phase, so that each neural response can be assigned a different weight, thus varying its contribution to the final classification [32]. In this way, responses contaminated by noise, for example, should be assigned a low score and then only marginally contribute to the selection of a letter.

While P300 spellers are typically controlled by a single person, a similar approach has also been applied to groups jointly controlling a P300 speller [33]. This was done using a freely available dataset [34], where the letters to be spelled and the stimulation sequences were the same for all participants, allowing to simulate a collaborative environment as if all recordings were performed simultaneously. In this context, users' brain responses are aggregated to identify the letter to be spelled. This resulted in faster and more accurate performances than those achieved by every single subject. Moreover, an optimized BCI can help identify the smallest subgroup achieving the most accurate decision on what letter to spell, saving resources and allowing to leave out those members whose exclusion does not decrease team performance. However, while this simulated framework was helpful to demonstrate the effectiveness of the algorithm proposed in [33], it was not realistic because, typically, people do not want to communicate the same letters and words.

D. Contributions

Previous research on collaborative BCIs for group decision making has shown how decoding confidence from neural signals can provide better estimates of the accuracy of each group member than reported confidence, leading to more accurate team decisions. This approach allows to maintain humans in the loop, instead of replacing them with one or more artificial agents (as for multi-classifier systems [35]), which may lead to suboptimal decisions [14] and generate ethical concerns [36], [37]. Importantly, these BCIs were trained on each participant's neural data to provide objective confidence estimates, irrespective to the context in which they were used. However, to counteract biases in group decision making arising from diverse personalities, genders, and cultural backgrounds of the team members [38], confidence weights should also optimally adapt to the group the user is working with.

This paper extends previous collaborative BCIs for group decision making along three main directions.

First, we streamline the confidence decoders by using the preprocessed response-locked EEG signals of each participant, without any further transformation or feature selection used in previous research [13], [28]. We also replaced logistic regression with linear support vector machine (SVM) to transform brain activity into confidence estimates, as SVM has been shown to provide the most accurate performance in event-related BCIs [39]. Streamlining collaborative BCIs represents another step towards making confidence decoders compatible with online BCIs [40] and applicable to everyday life.

Second, we introduce an additional step to our BCI decoders aimed at optimizing confidence weights to the group the participant is working with. By solving a small-dimension, mixed, integer linear programming problem for each group, we assign a confidence weight to each team member to maximize group performance in the training set. Hence, the decisions of each team member are weighted according to both the individual neural activity (as in previous collaborative BCI approaches) and the group composition (contribution of this study). We study the performance of groups using this multi-objective approach in a realistic face recognition task [13]. We simulate groups of different size and membership by aggregating decisions of multiple participants, and compare team performance achieved by groups using different decision methods, including traditional collaborative BCI as described in [13] (gold standard) and the proposed optimized BCI.

Third, the introduction of an optimization step for our BCI decoders based on group membership allows us to also take into account other social factors, such as fairness and equity. Specifically, we introduce a hyperparameter to control the balance between the confidence weights of group members.

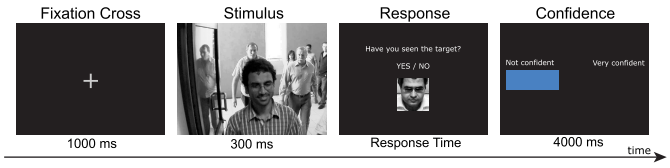

Fig. 1. Experimental protocol of the face recognition task.

This allows, during training, to establish how much to allow leadership behaviors (i.e., one/few participants deciding for the whole group) and promote fairness, with the extreme case being the standard majority rule (maximum fairness) at the expense of team accuracy. We investigate the impact that this parameter has on team performance, as in many scenarios, a trade-off between group accuracy and fairness is required.

II. Methods

A. Participants

Ten healthy participants (37.8 ± 4.8 years old, seven females, all right-handed) with normal or corrected-to-normal vision took part in the experiment [13]. Participants were compensated with £16-20, based on their accuracy in the task. This research received University of Essex ethical approval in July 2014. All participants signed an informed consent form before taking part in the experiment.

B. Data Recording and Processing

Neural and behavioral data were collected while participants were undertaking a face recognition decision-making task [13]. The experiment was split into six blocks of 48 trials. At the beginning of each block, participants were shown a picture of the face of the target person for that block. Each trial (Fig. 1) started with a fixation cross, shown for 1 s, followed by a grayscale picture of a crowded indoor environment for 300 ms. Then, participants were shown again the image of the target face, and were asked to decide, as quickly as possible, if the target was present in the picture. Finally, they were asked to rate their confidence in that decision, in a range from 0 to 100. In each block, 25% of the images contained the target.

Neural data were recorded using a 64-channel BioSemi Active Two EEG system. Electrodes were placed according to the international 10-20 system. Each channel was referenced to the average voltage recorded from the electrodes placed on each earlobe, and sampled at 2,048 Hz. To reduce artifacts and increase the signal-to-noise ratio of EEG signals, data were band-pass filtered between 0.15 and 40 Hz, and artifacts caused by ocular movements were removed with a standard correlation-based subtraction algorithm [41].

Response-locked epochs starting 1 s before the user's response and lasting 1.5 s were extracted from the EEG data for each trial, baseline corrected with the average voltage recorded in the 200 ms before the stimulus onset, and down-sampled to 128 Hz. Epochs were then split into two groups, "confident" and "not confident", depending on whether the associated decision made by the participant was correct or not, respectively.

Behavioral data were recorded through a USB mouse, and included the responses provided by participants (decision and reported confidence) and response times. The left/right mouse buttons were used to report presence/absence of a target. Confidence was reported using the mouse wheel, scrolling up/down to increase/decrease the confidence by 10%.

C. Confidence Decoding

Decision confidence $w_{s,t}$ for participant s during trial t is estimated by solving the binary classification problem where, for each trial, the output is whether the participant made the correct decision (label $+1$) or not (label -1). We build a separation hyperplane using SVM, and we measure the confidence looking at the distribution of the trials with respect to the separating hyperplane, drawing inspiration from the Optimized Score-Based decision Function (OSBF) proposed for P300 spellers [32]. We, therefore, call this method collaborative Optimized Score-Based decision Function (cOSBF). Specifically, we assume that a trial that is far away from the hyperplane on the positive side corresponds to high confidence of participant s for that decision, whereas the more the trial moves towards the "wrong" side of the hyperplane, the lower we assume the confidence. To better quantify the subjects' confidence on a given trial, we define an optimization problem that automatically chooses the weights for each participant for that trial. Considering the hyperplanes together, we gain a complete view of the subjects' behavior on each trial, whereas the hyperplane is built by looking at the single participant. The objective function and constraints of the optimization problem aim to increase the overall accuracy on a validation set and impose some fairness constraints, if needed, to ensure a contribution to the decision of each participant.

To calibrate the cOSBF, we define the training set as:

$$\mathcal{T}_s = \{(x_{s,t}, \bar{y}_{s,t}) \colon x_{s,t} \in \mathbb{R}^l,\ \bar{y}_{s,t} \in \{-1, 1\} \quad \forall t \in \{1, \ldots, n\} \quad \forall s \in \{1, \ldots, m\} \tag{1}$$

where:

- m is the number of participants in the group;
- n is the number of trials in the training phase;
- $x_{s,t}$ are the pre-processed EEG recordings for subject s during trial t;
- l is the number of features considered;
- $\bar{y}_{s,t}$ describes the outcome of the decision of participant s for trial t. In particular, $\bar{y}_{s,t} = 1$ if s made the correct decision during trial t, $\bar{y}_{s,t} = -1$ otherwise.

The calibration phase is constituted by two different training steps: (1) a linear SVM is trained for each participant to estimate his/her confidence from the neural signals, and (2) a mixed integer linear programming problem is solved to evaluate the confidence estimated by the SVM of each individual, trial and group. Each of these steps requires a distinct portion of the training set. For this reason, the set of trials $\{1, \ldots, n\}$ belonging to the calibration phase is partitioned in two parts (T_1 and T_2):

$$\mathcal{T}_{s,i} = \{(x_{s,t}, \bar{y}_{s,t}) \in \mathcal{T}_s \colon t \in T_i\} \quad \forall s \in \{1, \ldots, m\},\ i = 1, 2, \tag{2}$$

which are given in input to two sequential training phases.

1) COSBF - First Training Phase: In this phase, a pool of m SVMs is trained on the first portion of data $\mathcal{T}_{s,1}$ for each participant $s \in \{1, \ldots, m\}$, although, in principle, any linear classifier could be used. A SVM learns a separating hyperplane (w, b), where $w \in \mathbb{R}^l$ contains classification weights and $b \in \mathbb{R}$ is a bias term; the values of (w, b) are learned in a supervised way during the training phase. The discriminant function is:

$$f(x) = \text{sign}(w^T x + b), \tag{3}$$

where $w^T x + b$ is the decision value, and its absolute value is proportional to the distance of x from the hyperplane.

The aim of each SVM trained on $\mathcal{T}_{s,1}$ is to discriminate between correct and incorrect decisions, and use its decision values to estimate the decision confidence of a specific user during a trial. We denote with $\{(w_s, b_s) \forall s \in \{1, \ldots, m\}\}$ the pool of separating hyperplanes computed in this phase.

2) COSBF - Second Training Phase: This phase aims to quantify decision confidence of each subject $s \in \{1, \ldots, m\}$ during a generic trial t by solving a Mixed Integer Linear Programming (MILP). The input data of this phase are both the output of the first phase (i.e., a set of m separating hyperplanes) and the second portion of training data $\mathcal{T}_{s,2}$ of all participants $s \in \{1, \ldots, m\}$. The MILP assigns the weights to each subject who takes part in a group decision, and the assigned weight depends both on the specific subject and on the group composition, so that the accuracy of the decision is maximized given the group composition.

Each hyperplane (w_s, b_s) trained during the first phase is used to compute the decision values of data in $\mathcal{T}_{s,2}$:

$$dv_s = \{w_s^T x_{s,t} + b_s \quad \forall t \in T_2\} \quad \forall s \in \{1, \ldots, m\} \tag{4}$$

The set of decision values dv_s are used to partition the l-dimensional space in four ordered confidence zones $a - d$, on the basis of the trials distribution with respect to the separating hyperplane. Zone a includes trials with the highest decision confidence, and zone d consists of trials with a low decision confidence. These confidence zones are identified by the quartiles q_1, q_2 and q_3 of the distribution of the decision values dv_s, as represented in Fig. 2. A score corresponds to each confidence zone, so that the confidence weight $w_{s,t}$ is the score of the confidence zone where trial t belongs to for individual s. The values of the scores are specific to an individual and a group, and they are computed by solving the MILP.

The MILP takes in input a zone assignment vector z:

$$z_{s,t,v} = \begin{cases} 1, & \text{if trial } t \text{ is assigned to zone } v \text{ for individual } s \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $s \in \{1, \ldots, m\}$, $t \in T_2$, $v \in \{a, b, c, d\}$ and returns in output a pool of score vectors $\tilde{w}_s = \{a, b, c, d\}$.

We can then express the confidence weight of participant s at trial t as:

$$w_{s,t} = \tilde{w}_s^T z_{s,t} \tag{6}$$

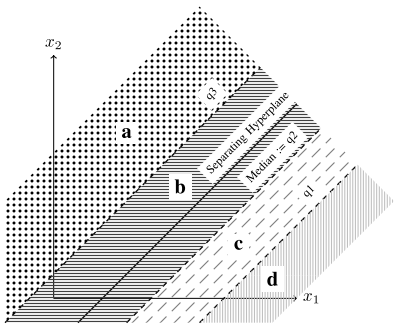

Fig. 2. Representation of the score distribution, reflecting the displacement of the points w.r.t. the distribution of the decision values. The different areas represent the confidence of the classification w.r.t. the target class. In this simplified example, data are described by just two features x_1 and x_2.

The MILP to be solved to find the vectors of score $\tilde{w}_s$ is:

$$\min \sum_{t \in T_2} \xi_t + M \sum_{s=1}^{m} \sum_{v \in \{a,b,c\}} \zeta_{s,v} \tag{7}$$

$$\tilde{w}_{s,a} \leq u \qquad \forall s \in \{1, \ldots, m\} \tag{8}$$

$$\tilde{w}_{s,d} \geq 0 \qquad \forall s \in \{1, \ldots, m\} \tag{9}$$

$$\tilde{w}_{s,v} \geq \tilde{w}_{s,v+1} \qquad \forall s \in \{1, \ldots, m\}, v \in \{a, b, c\} \tag{10}$$

$$\tilde{w}_{s,v} \geq \tilde{w}_{s,v+1} + 0.1 - \zeta_{s,v} \quad \forall s \in \{1, \ldots, m\}, v \in \{a, b, c\} \tag{11}$$

$$\tilde{w}_{s_1,v} \geq \tilde{w}_{s_2,v} \qquad \forall s_1, s_2 \in \{1, \ldots, m\} : acc_{s_1} > acc_{s_2}, \quad \forall v \in \{a \ldots d\} \tag{12}$$

$$\sum_{s=1}^{m} \bar{y}_{s,t} \tilde{w}_s^T z_{s,t} \geq 1 - \xi_t \qquad \forall t \in T_2 \tag{13}$$

$$\sum_{t \in T_2} \tilde{w}_s^T z_{s,t} \geq \frac{\eta}{m} \sum_{\bar{s}=1}^{m} \sum_{t \in T_2} \tilde{w}_{\bar{s}}^T z_{\bar{s},t} \qquad \forall s \in \{1, \ldots, m\} \tag{14}$$

$$\xi_t \geq 0 \qquad \forall t \in T_2 \tag{15}$$

where:

- the objective function (7) requires to minimize both incorrect group decisions, expressed in terms of slack variables ξ related to constraints (13), and slack variables ζ from constraints (11);
- constraints (8) and (9) impose an upper and lower bound over the score of the maximum and minimum confidence zones a and d: these bounds ensure existence of solution and non-negative scores;
- constraints (10) impose that, for each participant, the value of the scores are ordered according to the confidence level, so that zone a has an higher score than zone b, etc. Furthermore, constraints (11) require that a distance of 0.1 is required between two consecutive scores; since this is a strong condition that can cause infeasibility, we add some slack variables ζ;
- constraints (12) impose that the weight assigned to individuals should reflect the confidence in their classification's accuracy (note that acc_{s_1} and acc_{s_2} are the SVM's accuracies on $\mathcal{T}_{s,1}$ and $\mathcal{T}_{s,2}$, respectively). Thus, if the SVM built for s_1 is strictly more accurate than the one

built for s_2, then the weight assigned to any zone for s_1 cannot be lower than the weight assigned to the corresponding zones for s_2;

- constraints (13) impose that the group decision for trial t should be correct, with a slack variable ξ_t allowing the failure if needed;
- constraints (14) impose a balance condition on the relative weight of the individuals in the group, which depends on a hyperparameter $0 \leq \eta \leq 1$. In case $\eta = 0$, the constraints are trivially satisfied, thus no balance is required and leaders may emerge; for $\eta = 1$, it is required that all participants equally contribute to decisions. These constraints allow to adapt our approach to different requirements in group decision, whether the priority is maximizing accuracy or inclusiveness and representation. Importantly, introducing the MILP problem allows us to easily include any constraint or requirement on the subjects' participation in the decision, which would be more complicated with a fixed strategy of weights choice.

A correct group decision can be imposed by:

$$rsign\left(\sum_{s=1}^{m} \bar{y}_{s,t}\tilde{w}_s^T z_{s,t}\right) = 1 \Rightarrow \sum_{s=1}^{m} \bar{y}_{s,t}\tilde{w}_s^T z_{s,t} > 0 \quad (16)$$

$$\Rightarrow \sum_{s=1}^{m} \bar{y}_{s,t}\tilde{w}_s^T z_{s,t} \geq \epsilon \quad (17)$$

when $rsign$ is a randomizing sign operator which returns $+1$ if its argument is positive, -1 if it is negative, and randomly chooses between $+1$ and -1 if its argument is 0. Finally, condition (13) is obtained by rescaling $\tilde{w}_s$ by ϵ and by adding a slack variable ξ_t to guarantee feasibility.

Fig. 3 summarizes the algorithmic framework proposed.

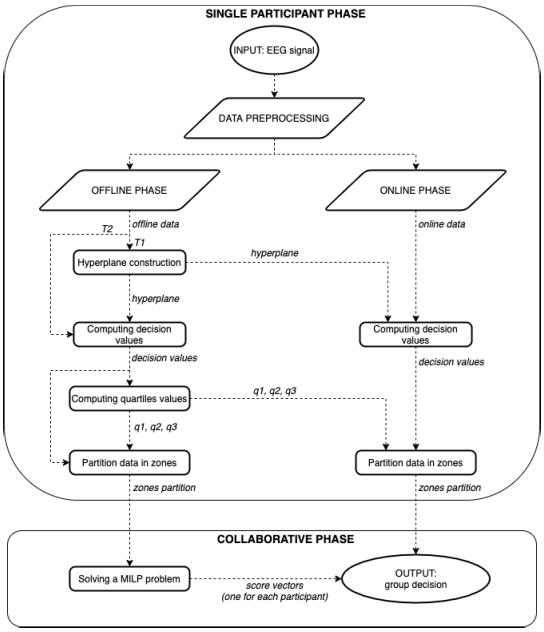

Fig. 3. Flowchart of the algorithmic framework procedure. The proposed framework requires a first single participant phase during which brain signals recorded during the experiment are partitioned into four confidence zones. The collaborative phase will then use the learned zone partitions to assign to each participant a score vector, that will be used to compute the group decisions during the online testing phase.

D. Experimental Setup

The 288 trials collected in our experiment for each participant were split into a training set composed of the first 60% of the trials (N = 172), and a test set ($\mathcal{W}$) composed of the last 40% of the trials (N = 116). The calibration trials were further temporally partitioned into T_1 (60% of the training set, N = 103) and T_2 (remaining 40% of training trials, N = 69). For the MILP problem, we empirically chose the values $u = 10$ and $\eta = 0.7$. The value of u does not have a significant impact on the results, as opposed to η (see Section III-C).

The choice to temporally partition the training set was motivated by practical considerations of implementing a collaborative BCI, that would first require a calibration phase for each participant before being used online to augment group decision making. However, due to the non-stationarity of brain signals, BCIs are typically recalibrated over time to ensure their stable performance [42]. Hence, the results obtained in our study represent a lower-bound of BCI decoding performance, which could be further enhanced with periodic recalibration.

To evaluate the performance of cOSBF-based groups, we compare it with the performance of the same groups adopting several alternative strategies for obtaining group decisions:

- *Majority*, where all group members have always the same weight (i.e., $w_{s,t} = 1$);
- *Logistic*, where the confidence weights are computed as in [13] using logistic regression and the same training set as cOSBF;
- *Confidence*, where the confidence weights are the confidence rates reported by the participants after each decision;
- *SVM*, where the confidence weights are the decision values of the same SVM trained for cOSBF scaled to the interval $(0, 1]$.

E. Performance Evaluation

The performance of groups of different sizes and decision-making strategies are evaluated using accuracy, specificity and sensitivity, computed as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (18)$$

$$Specificity = \frac{TN}{TN + FP}, \quad (19)$$

$$Sensitivity = \frac{TP}{TP + FN}, \quad (20)$$

where TP, TN, FP, FN indicate the number of true positives, true negatives, false positives, and false negatives on the test set, respectively.

We also introduce a metric to evaluate the average influence I_i of each participant i on the group decisions, measured as the relative weight of i on all m trials in the test set $\mathcal{W}$:

$$I_i = \frac{1}{|\mathcal{W}|} \sum_{t \in \mathcal{W}} \frac{w_{i,t}}{\sum_{s=1}^{m} w_{s,t}} \quad (21)$$

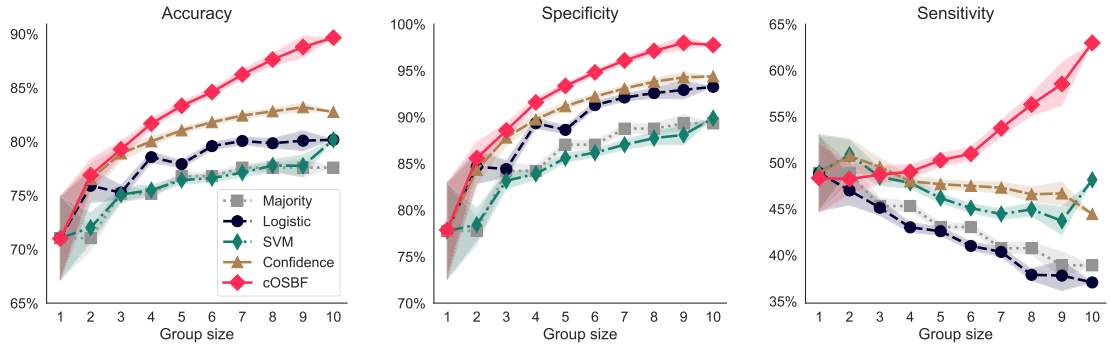

Fig. 4. Mean performance of groups of increasing size for the three methods. Shaded areas show standard error of the mean.

To evaluate the ability of the cOSBF method to capture the objective confidence of each group member, for each participant i and group j we compute the following confusion matrix on the test set $\mathcal{W}$:

	Subject i correct	Subject i incorrect
Group j correct	A_{ij} : % agree	B_{ij} : % disagree
Group j incorrect	C_{ij} : % disagree	D_{ij} : % agree

(22)

An optimal group would have high values on the first row and low values on the second row. For each participant i, we then average the confusion matrices obtained for all the groups S_j of a given size s_j both for cOSBF and for the BCI based on logistic regression (gold standard, equation 23).

	Subject i correct	Subject i incorrect
Average group correct	$A = \frac{\sum_{j \in S_j} A_{ij}}{\lvert S_j \rvert}$	$B = \frac{\sum_{j \in S_j} B_{ij}}{\lvert S_j \rvert}$
Average group incorrect	$C = \frac{\sum_{j \in S_j} C_{ij}}{\lvert S_j \rvert}$	$D = \frac{\sum_{j \in S_j} D_{ij}}{\lvert S_j \rvert}$

(23)

Finally, for each element of the confusion matrix we compute the difference between the value corresponding to the cOSBF and the value corresponding to the Logistic (equation 24). Note that, on the second row, we inverted the sign of the difference, so that positive differences mean that cOSBF is performing better than Logistic in all matrices, whereas negative differences mean that the Logistic is performing better than cOSBF.

$$\begin{matrix} \frac{A^{cOSBF} - A^{Log}}{A^{Log}} & \frac{B^{cOSBF} - B^{Log}}{B^{Log}} \\ -\frac{C^{cOSBF} - C^{Log}}{C^{Log}} & -\frac{D^{cOSBF} - D^{Log}}{D^{Log}} \end{matrix} \quad (24)$$

III. Results

A. Individual and Group Performance

Individual accuracy of participants in the experiment ranged from 52.8% to 92.4% (mean ± standard deviation = 72.3 ± 12.0%), hence being better than random performance (50%). Average reaction times ranged from 0.750 s to 1.737 s (mean ± standard deviation = 1.153 ± 0.287 s). Average specificity was 77.4 ± 15.8% and average sensitivity was 56.9 ± 10.9%.

Fig. 4 shows the average accuracy, sensitivity and specificity of groups of increasing size in the test set using the proposed cOSBF method, as well as the other weighting strategies described in Section II-D.

Across the three evaluation metrics, cOSBF outperforms all other strategies, and the improvement increases with the size of the group. As seen in Table I, these differences are statistically significant for all group sizes 5-8, and often also for smaller group sizes 2-4. Particularly noteworthy is the sensitivity of groups, which increases with the group size only for cOSBF. For Majority, the reduction of sensitivity over group sizes is due to the Condorcet's theorem, as the average sensitivity of the participants in the test set is lower than random (50%). Moreover, the Logistic suffered from the temporal train/test split, as in [13] with cross-validation splits it was capable of increasing sensitivity over group size.

To better investigate the reasons behind this performance improvement, we computed the "normalized accuracy", i.e., the percentage of trials in which the group made a correct decision among those trials where at least one team member made the correct decision. In other words, we removed from the calculation of accuracy the trials where all team members made an incorrect decision. Hence, the normalized accuracy ranges between 0 and 100%. In the rest of the analysis we will focus our attention on the comparison between cOSBF, Majority (baseline performance) and Logistic (gold standard cBCI method [13]).

Fig. 5 shows the normalized accuracy for all groups of size 2, 3 and 4. As observed in Fig. 4, the improvement increases with the group size, and the majority of groups benefit from cOSBF.

Fig. 6 shows the average confusion matrices computed according to equation 24 for cOSBF and Logistic. Logistic and cOSBF are equivalent on trials where the participant and the group made the correct decision. However, on other elements of the confusion matrices, cOSBF is consistently better than the Logistic (i.e., red colors are more frequent than blue). This confirms the ability of our weighing strategy to better estimate the participant's decision confidence. Moreover, the column "M" within each confusion matrix shows that the improvement achieved by cOSBF increases with the group size, as previously seen in Fig. 4. Statistical comparisons between cOSBF and Logistic performance for each group

TABLE I

WILCOXON TWO-SIDED p VALUES COMPARING COSBF-BASED GROUP DECISIONS WITH DECISIONS OBTAINED BY THE MAJORITY RULE, THE WEIGHTED MAJORITY RULE BASED ON LOGISTIC REGRESSION, THE WEIGHTED MAJORITY RULE BASED ON SVM SCORES, AND THE WEIGHTED MAJORITY RULE BASED ON REPORTED CONFIDENCE FOR DIFFERENT GROUP SIZES. VALUES IN BOLD INDICATE p VALUES BELOW THE BONFERRONI-CORRECTED STATISTICAL SIGNIFICANCE LEVEL $0.05/96 = 0.00052$. THE LAST ROW INDICATES THE NUMBER OF GROUP SIZES FOR WHICH COSBF IS SIGNIFICANTLY BETTER THAN THE OTHER METHOD

Group Size	cOSBF-Majority Accuracy	cOSBF-Majority Specificity	cOSBF-Majority Sensitivity	cOSBF-Logistic Accuracy	cOSBF-Logistic Specificity	cOSBF-Logistic Sensitivity
2	$\mathbf{2.2 \times 10^{-7}}$	$\mathbf{4.5 \times 10^{-8}}$	6.5×10^{-1}	1.1×10^{-2}	9.1×10^{-2}	2.8×10^{-1}
3	$\mathbf{4.4 \times 10^{-17}}$	$\mathbf{6.9 \times 10^{-15}}$	$\mathbf{2.0 \times 10^{-6}}$	$\mathbf{1.2 \times 10^{-16}}$	$\mathbf{1.4 \times 10^{-14}}$	$\mathbf{1.4 \times 10^{-6}}$
4	$\mathbf{2.7 \times 10^{-35}}$	$\mathbf{9.0 \times 10^{-36}}$	$\mathbf{6.2 \times 10^{-9}}$	$\mathbf{4.4 \times 10^{-24}}$	$\mathbf{2.0 \times 10^{-13}}$	$\mathbf{3.1 \times 10^{-21}}$
5	$\mathbf{5.7 \times 10^{-42}}$	$\mathbf{1.4 \times 10^{-40}}$	$\mathbf{1.2 \times 10^{-28}}$	$\mathbf{1.5 \times 10^{-40}}$	$\mathbf{1.2 \times 10^{-37}}$	$\mathbf{4.8 \times 10^{-31}}$
6	$\mathbf{4.1 \times 10^{-36}}$	$\mathbf{3.5 \times 10^{-36}}$	$\mathbf{1.1 \times 10^{-23}}$	$\mathbf{4.2 \times 10^{-32}}$	$\mathbf{5.6 \times 10^{-26}}$	$\mathbf{1.3 \times 10^{-29}}$
7	$\mathbf{2.0 \times 10^{-21}}$	$\mathbf{2.7 \times 10^{-21}}$	$\mathbf{2.1 \times 10^{-20}}$	$\mathbf{6.4 \times 10^{-21}}$	$\mathbf{5.2 \times 10^{-19}}$	$\mathbf{1.1 \times 10^{-20}}$
8	$\mathbf{5.1 \times 10^{-9}}$	$\mathbf{5.2 \times 10^{-9}}$	$\mathbf{6.9 \times 10^{-9}}$	$\mathbf{5.7 \times 10^{-9}}$	$\mathbf{8.5 \times 10^{-9}}$	$\mathbf{6.1 \times 10^{-9}}$
9	5.1×10^{-3}	5.1×10^{-3}	4.9×10^{-3}	5.0×10^{-3}	5.0×10^{-3}	4.8×10^{-3}
	7	7	6	6	6	6

Group Size	cOSBF-SVM Accuracy	cOSBF-SVM Specificity	cOSBF-SVM Sensitivity	cOSBF-Confidence Accuracy	cOSBF-Confidence Specificity	cOSBF-Confidence Sensitivity
2	$\mathbf{1.1 \times 10^{-5}}$	$\mathbf{5.0 \times 10^{-7}}$	1.4×10^{-1}	6.4×10^{-1}	1.8×10^{-1}	2.3×10^{-2}
3	$\mathbf{3.9 \times 10^{-15}}$	$\mathbf{1.1 \times 10^{-17}}$	8.4×10^{-1}	2.3×10^{-1}	8.5×10^{-2}	1.4×10^{-1}
4	$\mathbf{4.7 \times 10^{-33}}$	$\mathbf{1.4 \times 10^{-35}}$	8.8×10^{-2}	$\mathbf{2.1 \times 10^{-8}}$	$\mathbf{2.2 \times 10^{-8}}$	6.6×10^{-2}
5	$\mathbf{2.0 \times 10^{-41}}$	$\mathbf{1.7 \times 10^{-42}}$	$\mathbf{9.7 \times 10^{-11}}$	$\mathbf{8.3 \times 10^{-15}}$	$\mathbf{3.1 \times 10^{-12}}$	$\mathbf{1.8 \times 10^{-7}}$
6	$\mathbf{1.5 \times 10^{-35}}$	$\mathbf{4.9 \times 10^{-36}}$	$\mathbf{4.9 \times 10^{-17}}$	$\mathbf{4.6 \times 10^{-17}}$	$\mathbf{3.4 \times 10^{-16}}$	$\mathbf{1.3 \times 10^{-9}}$
7	$\mathbf{2.2 \times 10^{-21}}$	$\mathbf{1.9 \times 10^{-21}}$	$\mathbf{2.3 \times 10^{-16}}$	$\mathbf{1.2 \times 10^{-15}}$	$\mathbf{4.2 \times 10^{-13}}$	$\mathbf{6.4 \times 10^{-14}}$
8	$\mathbf{5.1 \times 10^{-9}}$	$\mathbf{5.1 \times 10^{-9}}$	$\mathbf{1.3 \times 10^{-8}}$	$\mathbf{6.5 \times 10^{-8}}$	$\mathbf{5.7 \times 10^{-7}}$	$\mathbf{4.3 \times 10^{-8}}$
9	2.0×10^{-3}	2.0×10^{-3}	2.0×10^{-3}	5.0×10^{-3}	5.7×10^{-3}	5.8×10^{-3}
	7	7	4	5	5	4

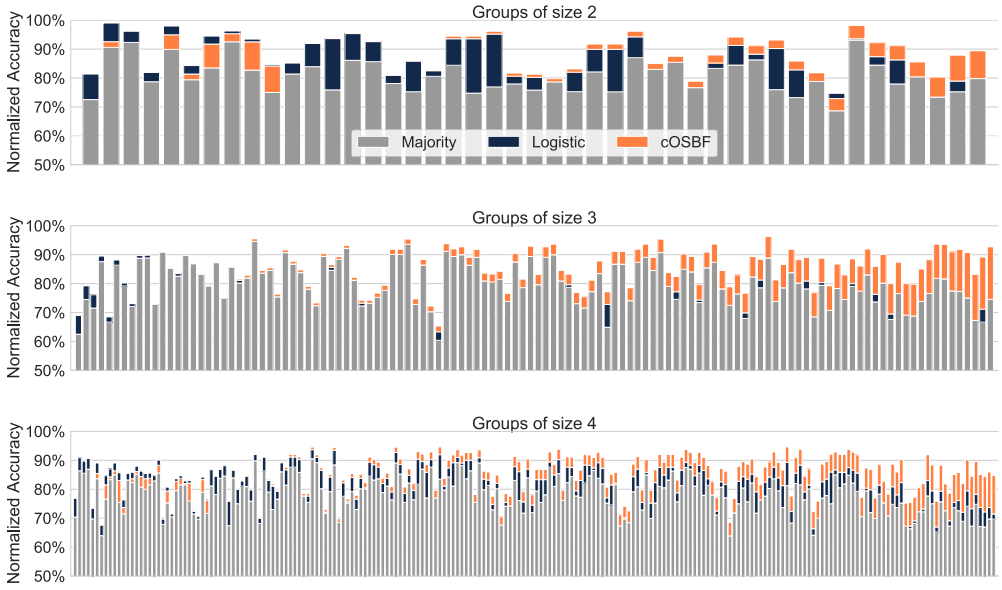

Fig. 5. Normalized accuracy of groups of size 2-4, comparing cOSBF, Logistic and Majority performance for each group. Groups (x-axis) are sorted by increasing improvement of cOSBF over Majority and Logistic.

size (Fig. 6) through either Wilcoxon paired test and sign test returned p values below the Bonferroni corrected statistical significance level $0.05/10 = 0.005$ for all group sizes 4-9.

B. Influence of Single Participants

As it can be seen in Fig. 4 and in previous collaborative BCI research [13], [28], the average performance of groups of size 3 is particularly difficult to improve for confidence-weighted majority strategies. This is because, in weighted majority, the sum of the weights of a minority, correct subgroup must be higher than the sum of the weights of the majority, incorrect subgroup, in order to for the group to make the correct decision. For group size 3, this scenario occurs only when one team member made a correct decision, and its single weight is higher than the sum of the other two members' weights.

To investigate why cOSBF is capable of bringing significant improvements also for groups of size 3, in Fig. 7 we report the influence of three representative participant on all the groups of

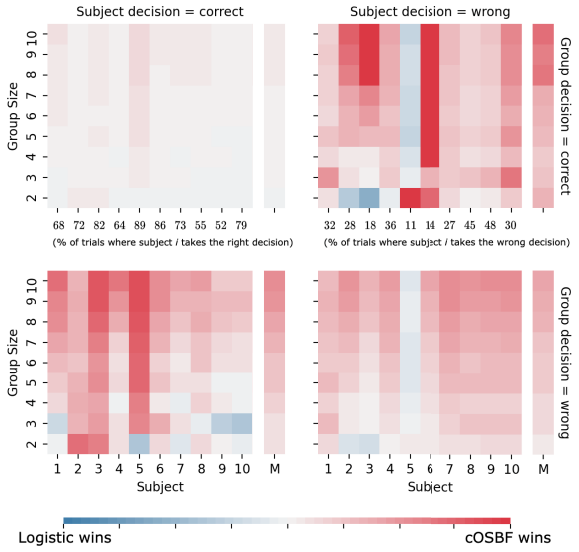

Fig. 6. Comparison between the capability of cOSBF and Logistic strategies of detecting the outcome of individual's decisions as described in equation 24. M columns show the means across participants.

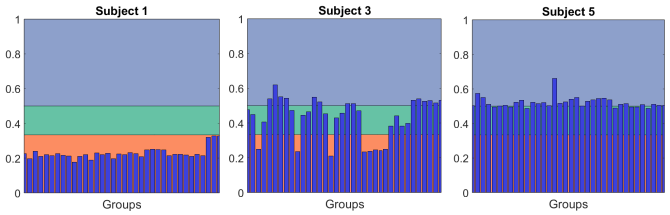

Fig. 7. Average influence of three representative participants on groups of size 3 for $\eta = 0.7$. If the influence is below 0.333, the participant is, on average, not affecting the group decision; if the influence is between 0.333 and 0.5, the participant contributes to group decisions; if the influence is > 0.5, the participant determines the team decision.

size 3 for which that individual is a member of. The influence changes depending on the group considered, but the most accurate participants (e.g., 5) have always a higher influence than others; if the influence of a given user is greater than 0.5, that participant determines the group decision, no matter the opinions of the other members. Conversely, the participants with the worst accuracy (e.g., 1) have always a relatively low influence. This shows that cOSBF is able to capture the target detection capability of each user. Thanks to constraint (14), however, all participants still have some influence on the group decision. In fact, we should note that the influence of each user depends both on the group composition and on the neural signals of that participant. This multi-objective confidence decoding strategy allows cOSBF to outperform other methods. The influence variation depends also on the parameter η in constraint (14), as we will see in the next subsection.

C. Influence of the Balance Constraint

To investigate the effects of the parameter η in constraint (14), Fig. 8 shows the leadership percentage and the average accuracy of groups of size 3 for increasing values of η. The leadership percentage is computed as the percentage of trials

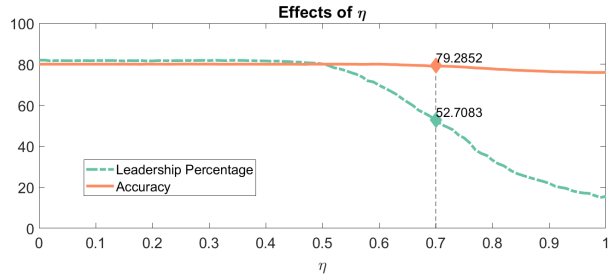

Fig. 8. Leadership percentage and average accuracy of groups of size 3 for increasing values of η. The leadership percentage is computed as the percentage of trials where the influence of one of the participants to the decision is larger than 0.5.

where the influence of one participant on the decision is larger than 0.5. Both the leadership percentage and the accuracy are, as expected, inversely proportional to η: when η increases we require more fairness in group decisions, and this can reduce the average accuracy of the groups to the benefit of inclusion. In the experimental setup, we decided to use the value $\eta = 0.7$, which corresponds to a mean level of leadership while keeping high levels of accuracy.

To further understand how the influence of the individual varies depending on the trial and on the choice of η, in Fig. 9 we analyzed the influence of each user on each trial for two representative groups of size 3 and two representative values of η. We selected $\eta = 0.7$ (as in our experimental setup) and $\eta = 0.5$, which requires a less stringent constraint on the balance among team members.

As a first example, we considered a group formed by participants 5, 6, and 7. For both $\eta = 0.5$ and $\eta = 0.7$, participant 5 (blue dots) has always a very high impact on the team decision. However, imposing the stronger constraint on the weight balance among participants ($\eta = 0.7$) allows to reduce the mean influence of member 5 and to increase the influence of participants 6 and 7. Indeed, for $\eta = 0.5$ the group accuracy coincides with the accuracy of user 5, as this participant has an influence greater than 0.5 in all trials and, so, it determines the group decision all the times. Conversely, for $\eta = 0.7$, in some trials, the influence of participant 5 is lower than the sum of the influence of the other two participants, which leads to increased group accuracy.

As another example,, we considered the group composed of participants 6, 8, and 9. Contrarily from before, in this case when $\eta = 0.7$ the accuracy decreases with respect to the accuracy obtained with $\eta = 0.5$. The difference between these two examples is that, for the group composed of 5, 6, 7, all participants are very accurate when considered alone, so that the group benefits from the contribution of all participants, while in the second example, participant 6 is much more accurate than subjects 8 and 9 and, so, allowing the presence of a leader can be beneficial.

In summary, these results show that cOSBF allows to substantially improve group fairness with a minimal cost in terms of group accuracy. In fact, cOSBF is able to identify the most accurate team members, while ensuring all participants within a group contribute to the team decision as much as possible, which leads to significantly better group performance.

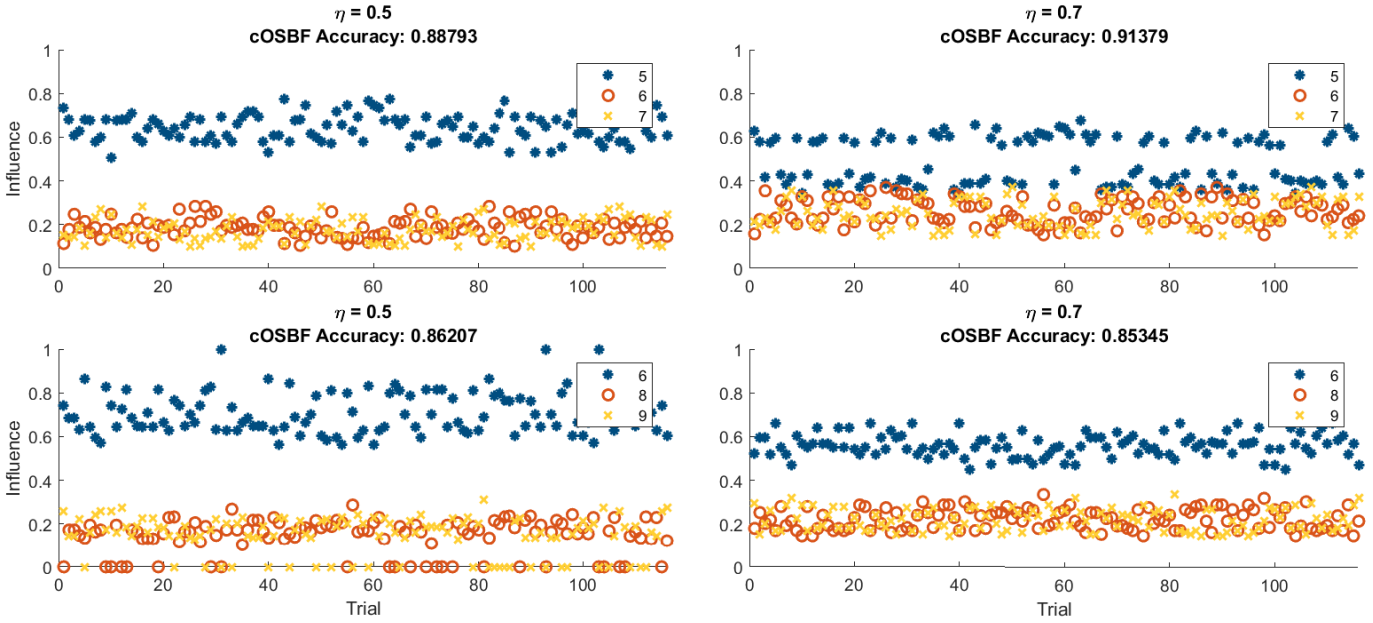

Fig. 9. Influence of each participant of two representative groups of size 3 in the test set with $\eta = 0.5$ or $\eta = 0.7$. For reference, participants 5-9 had an accuracy of 0.88793, 0.86207, 0.73276, 0.55172 and 0.51724, respectively.

IV. Conclusion

We developed a novel collaborative BCI capable of simultaneously decoding decision confidence from EEG neural activity and optimize it based on group composition. These confidence estimates were then used to integrate individual responses to obtain a group decision. Our optimized collaborative BCI significantly improves group accuracy in a realistic face recognition task of up to 14% depending on the group size when compared to non-optimal collaborative BCIs, weighted majority voting based on reported confidence, and standard majority voting. We also showed that this system increases fairness within the group, ensuring that all team members contribute to the group decision.

This research demonstrates how collaborative BCIs could be used in realistic scenarios to facilitate and enhance group decision-making, while maintaining humans in the loop, promoting fairness and reducing biases. The simultaneous optimization of the BCIs on a user- and group-basis makes them directly applicable to stable teams, which are more likely to develop trust than constantly-changing groups and, therefore, often more efficacious. For example, this collaborative BCI could assist a group of police officers monitoring security cameras footage to better identify suspects and threats. As suggested by earlier research [13], these BCIs can also promote the creation of human-machine teams, leading to higher accuracy than state-of-the-art face recognition algorithms. Similarly, a team of brokers could use this collaborative BCI to better decide buy/sell operations, leading to increase earnings and reduced losses.

References

[1] J. Surowiecki, *The Wisdom of Crowds*. New York, NY, USA: Anchor, 2005.

[2] R. H. Kurvers, A. De Zoete, S. L. Bachman, P. R. Algra, and R. Ostelo, "Combining independent decisions increases diagnostic accuracy of reading lumbosacral radiographs and magnetic resonance imaging," *PLoS ONE*, vol. 13, no. 4, 2018, Art. no. e0194128.

[3] J. A. Marshall, R. H. Kurvers, J. Krause, and M. Wolf, "Quorums enable optimal pooling of independent judgements in biological systems," *eLife*, vol. 8, pp. 1–14, Feb. 2019. [Online]. Available: http://biorxiv.org/content/early/2018/08/17/394460.abstract and https://elifesciences.org/articles/40368

[4] J. A. R. Marshall, G. Brown, and A. N. Radford, "Individual confidence-weighting and group decision-making," *Trends Ecology Evol.*, vol. 32, no. 9, pp. 636–645, Sep. 2017, doi: 10.1016/j.tree.2017.06.004.

[5] A. Laan, G. Madirolas, and G. G. de Polavieja, "Rescuing collective wisdom when the average group opinion is wrong," *Frontiers Robot. AI*, vol. 4, pp. 1–21, Nov. 2017, doi: 10.3389/frobt.2017.00056.

[6] A. Pouget, J. Drugowitsch, and A. Kepecs, "Confidence and certainty: Distinct probabilistic quantities for different goals," *Nature Neurosci.*, vol. 19, no. 3, pp. 366–374, 2016. [Online]. Available: http://www.nature.com/articles/nn.4240

[7] J. Navajas, B. Bahrami, and P. E. Latham, "Post-decisional accounts of biases in confidence," *Current Opinion Behav. Sci.*, vol. 11, pp. 55–60, Oct. 2016. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S2352154616300973

[8] D. Bang *et al.*, "Confidence matching in group decision-making," *Nature Hum. Behav.*, vol. 1, no. 6, pp. 1–7, May 2017. [Online]. Available: http://www.nature.com/articles/s41562-017-0117

[9] D. J. Robertson and A. M. Burton, "Unfamiliar face recognition: Security, surveillance and smartphones," *J. Homeland Defense Secur. Inf. Anal. Center*, vol. 3, no. 1, pp. 14–21, Apr. 2016.

[10] A. S. Ghuman *et al.*, "Dynamic encoding of face information in the human fusiform gyrus," *Nature Commun.*, vol. 5, no. 1, pp. 1–10, Dec. 2014.

[11] D. J. Robertson, E. Noyes, A. J. Dowsett, R. Jenkins, and A. M. Burton, "Face recognition by metropolitan police super-recognisers," *PLoS ONE*, vol. 11, no. 2, Feb. 2016, Art. no. e0150036.

[12] P. J. Phillips *et al.*, "Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 24, pp. 6171–6176, 2018.

[13] D. Valeriani and R. Poli, "Cyborg groups enhance face recognition in crowded environments," *PLoS ONE*, vol. 14, no. 3, Mar. 2019, Art. no. e0212935.

[14] M. P. Eckstein, K. Koehler, L. E. Welbourne, and E. Akbas, "Humans, but not deep neural networks, often miss giant targets in scenes," *Current Biol.*, vol. 27, no. 18, pp. 2827–2832, 2017.

[15] T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fierrez, and C. Busch, "Face image quality assessment: A literature survey," *ACM Comput. Surv.*, Dec. 2020, doi: 10.1145/3507901.

[16] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern Recognit.*, vol. 39, no. 9, pp. 1725–1745, 2006.

[17] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[18] C. Cinel, D. Valeriani, and R. Poli, "Neurotechnologies for human cognitive augmentation: Current state of the art and future prospects," *Frontiers Hum. Neurosci.*, vol. 13, p. 13, Jan. 2019.

[19] J. van Erp, F. Lotte, and M. Tangermann, "Brain-computer interfaces: Beyond medical applications," *Computer*, vol. 45, no. 4, pp. 26–34, Apr. 2012.

[20] P. Aricò, G. Borghini, G. D. Flumeri, N. Sciaraffa, and F. Babiloni, "Passive BCI beyond the lab: Current trends and future directions," *Physiolog. Meas.*, vol. 39, no. 8, Aug. 2018, Art. no. 08TR02.

[21] M. P. Eckstein *et al.*, "Neural decoding of collective wisdom with multi-brain computing," *NeuroImage*, vol. 59, no. 1, pp. 94–108, Jan. 2012.

[22] Y. Wang and T.-P. Jung, "A collaborative brain-computer interface for improving human performance," *PLoS ONE*, vol. 6, no. 5, May 2011, Art. no. e20422.

[23] X. Song *et al.*, "A collaborative brain-computer interface framework for enhancing group detection performance of dynamic visual targets," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–12, Jan. 2022.

[24] Z. Yijie *et al.*, "A multiuser collaborative strategy for MI-BCI system," in *Proc. IEEE 23rd Int. Conf. Digit. Signal Process. (DSP)*, Nov. 2018, pp. 1–5.

[25] B. Gu *et al.*, "Optimization of task allocation for collaborative brain–computer interface based on motor imagery," *Frontiers Neurosci.*, vol. 15, p. 753, Jul. 2021.

[26] Y. Qin *et al.*, "Classifying four-category visual objects using multiple ERP components in single-trial ERP," *Cognit. Neurodynamics*, vol. 10, no. 4, pp. 275–285, Aug. 2016.

[27] S. Aggarwal and N. Chugh, "Review of machine learning techniques for EEG based brain computer interface," *Arch. Comput. Methods Eng.*, pp. 1–20, Jan. 2022.

[28] D. Valeriani, C. Cinel, and R. Poli, "Group augmentation in realistic visual-search decisions via a hybrid brain-computer interface," *Sci. Rep.*, vol. 7, no. 1, p. 7772, Dec. 2017. [Online]. Available: http://www.nature.com/articles/s41598-017-08265-7

[29] J. Fernandez-Vargas *et al.*, "Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision making," *J. Neural Eng.*, vol. 18, no. 4, Aug. 2021, Art. no. 046055.

[30] R. Zhang *et al.*, "Local temporal correlation common spatial patterns for single trial EEG classification during motor imagery," *Comput. Math. Methods Med.*, vol. 2013, pp. 1–7, Nov. 2013.

[31] S. Perdikis, L. Tonin, S. Saeedi, C. Schneider, and J. D. R. Millán, "The cybathlon BCI race: Successful longitudinal mutual learning with two tetraplegic users," *PLOS Biol.*, vol. 16, no. 5, May 2018, Art. no. e2003787.

[32] L. Bianchi, C. Liti, G. Liuzzi, V. Piccialli, and C. Salvatore, "Improving P300 speller performance by means of optimization and machine learning," *Ann. Oper. Res.*, pp. 1–39,Jan. 2021.

[33] L. Bianchi, F. Gambardella, C. Liti, and V. Piccialli, "Group study via collaborative BCI," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Oct. 2019, pp. 272–276.

[34] F. Aloise *et al.*, "A covert attention P300-based brain–computer interface: Geospell," *Ergonomics*, vol. 55, no. 5, pp. 538–551, May 2012, doi: 10.1080/00140139.2012.661084.

[35] F. Roli, G. Giacinto, and G. Vernazza, "Methods for designing multiple classifier systems," in *Proc. Int. Workshop Multiple Classifier Syst.*, Cham, Switzerland: Springer, 2001, pp. 78–87.

[36] I. D. Raji, T. Gebru, M. Mitchell, J. Buolamwini, J. Lee, and E. Denton, "Saving face: Investigating the ethical concerns of facial recognition auditing," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, Feb. 2020, pp. 145–151.

[37] M. Smith and S. Miller, "The ethical application of biometric facial recognition technology," *Ai Soc.*, vol. 37, no. 1, pp. 167–175, Mar. 2022, doi: 10.1007/s00146-021-01199-9.

[38] D. Bang and C. D. Frith, "Making better decisions in groups," *Roy. Soc. Open Sci.*, vol. 4, no. 8, Aug. 2017, Art. no. 170193, doi: 10.1098/rsos.170193.

[39] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," *J. neural Eng.*, vol. 4, no. 2, p. R1, 2007.

[40] O. Tonet *et al.*, "Defining brain–machine interface applications by matching interface performance with device requirements," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 91–104, Jan. 2008.

[41] P. Quilter, B. MacGillivray, and D. Wadbrook, "The removal of eye movement artefact from EEG signals using correlation techniques," in *Proc. IEEE Conf. Random Signal Anal.*, vol. 159, Jan. 1977, pp. 93–100.

[42] W. Bishop *et al.*, "Self-recalibrating classifiers for intracortical brain–computer interfaces," *J. Neural Eng.*, vol. 11, no. 2, Apr. 2014, Art. no. 026001.