

Investigating Human and Classifier-Level Uncertainties: An EEG Study Using Grasp Imitation Task

Osman Cagri Oguz¹, Hamidreza Kasaei¹, Steffen Frey¹, Andreea Ioana Sburlea¹

¹ Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, The Netherlands

E-mail: o.c.oguz@rug.nl

ABSTRACT:

In BCI research, uncertainty quantification recently attracted researchers to enhance human-machine collaboration and transparency of the systems. Grasping constitutes one of the most crucial interactions with the environment, for which the associated risks of overconfident but inaccurate classification are high. However, there is yet to be an explicit paradigm design that allows investigating if classifier uncertainties align with that of humans. By developing a grasp imitation paradigm with visually degraded videos during the observation stage, we aimed to incur uncertainties on the grasping decisions and execution. EEG and EMG data from 8 participants is collected. The results demonstrated that among the grasp types, 5-finger precision grasp is the best classified one, with 47% within-class accuracy according to an LDA classifier using movement related cortical potentials as features. Interestingly, for the participant having highest ROC-AUC score, subjectively reported and the LDA classifier uncertainty aligns. As the data collection is still ongoing, future directions are also discussed.

INTRODUCTION

We perceive and act within noisy sensory environments, where the affordances and consequences are not always immediately clear. Imagine grasping a cup with a beverage with not yet known temperature. Humans are more likely to combine multiple information sources to estimate the risks and reduce uncertainties. Even if the interaction is supposed to happen in a very short manner, accommodating the safest grasp alternative would be wise to prevent harmful consequences, such as spill and burn. Humans can be remarkably good at accommodating different motor decisions as a result of constantly taking sensory and perceptual uncertainties into account [1]. One might also imagine this scenario where the motor intentions from brain signals are decoded by a classifier, which is then translated into executable motor commands for the robotic hand. If the classifier is insensitive to how humans represent uncertainties or overconfident regardless of its classification accuracy, generated response may likely cause injuries. In this research, we aim to see if uncertainty for brain decoding indeed aligns with how humans represent their uncertainties in a motor task, more

specifically grasp-imitation. The novel aspect of the research is twofold. First, we contribute to the field by introducing a paradigm inducing uncertainty in motor actions. Second, as a result of this adjustment, studying for the first time if uncertainty represented by BCI classifiers actually aligns to that of humans themselves.

In the field of brain-computer interfaces (BCIs), grasping attracts growing interest from the researchers due to its importance for daily functioning and interactions with the environment. To provide a comprehensive model of how grasps are represented in the brain, previous research showed that different stages of motor actions carry information about different characteristics. For instance, properties of the object to-be-grasped can be observed during the hand pre-shaping stage, while muscle activity is mostly evident in holding stages [2]. Using low-frequency time-domain signals (e.g., movement-related cortical potentials (MRCPs)) the researchers demonstrated a decoding performance above the chance level among three different grasps, especially during the planning and object release stages [3]. Unimanual and bimanual grasps have also been shown to be decoded if MRCPs are used as features [4].

However, merely focusing on the classification accuracy for BCI decoders may not provide the most comprehensive picture. In a similar line of reasoning, there is also a significant caveat for evaluating the conclusions reached by brain decoding studies. That is, demonstrating accurate classification based on the input signal (e.g., EEG features) may not naturally follow conclusions that those features indeed carry the relevant information for the task. This risk is often associated with greedy and non-transparent nature of algorithms used for classification tasks. To overcome this phenomenon of decoder's dictum, it is proposed to connect behavior in a psychologically plausible manner to have better grounded inferences for underlying neural representations [5]. In this respect, second-order representation of the informational content, such as uncertainty given the classification output, is a candidate to produce more grounded inferences. However, the previous research does not incorporate uncertainty regarding motor or perceptual cues for instructing movement execution to participants. Therefore, missing the opportunity to connect classifier and human level uncertainties.

It is also known that BCI classifiers are often overconfident even for inaccurate classifications, resulting in uncertainty quantification (UQ) techniques to attract researchers in the field [6]. More specific to EEG-based BCIs, UQ is arguably even more important given the poor signal-to-noise ratio compared to invasive recordings, especially for multiclass decoding scenarios, such as for different grasps. Additionally, for BCIs on decoding motor signals, overconfidence may lead to unwanted consequences in real-life when the potential risks of choosing the wrong action is high. Promisingly, recent research show improvements through uncertainty-based early classification and rejection [7], [8].

Previous research on UQ for EEG-based BCIs revealed that even though deep-learning based methods, such as Deep Ensembles, produced best classification accuracies, more traditional methods may represent uncertainty better through post-hoc calibration to match predicted class probabilities to accuracy [7]. In addition, researchers proposed frameworks based on Bayesian networks and dropout-based methods to estimate uncertainty for decoding EEG signals. They showed promising results for the classification tasks on public datasets with comparable performance to computationally costly ensemble-based methods [9].

To the best of our knowledge, if uncertainty represented by BCI classifiers actually aligns to that of humans themselves is yet to be studied. With this goal in mind, we use a paradigm inspired by previous perceptual decision-making research, in which the sensory stimulus strength is manipulated to adjust difficulty [10]. With this manipulation, it is expected to incur uncertainty regarding the perceptual and hence the motor decisions determining subsequent grasp execution. In this paper, we use a traditional classification algorithm to decode brain activity during grasp execution, namely the linear discriminant analysis (LDA), due to previous reports on proper uncertainty representation.

MATERIALS AND METHODS

Paradigm: The current research investigates multiple grasps by utilizing a grasp imitation task. The participants watched grayscale videos depicting three different grasps on four different objects to decide which type of grasp they will perform subsequently (see Fig. 1). The grasps include pincer (2-finger grip), precision (5-finger) and power grasp (with the palm also touching the object). The grasping videos follow after fixation cross on the screen and a variable-timed random noise overlaid on the first frame. Pre-stimulus noise is used in line with the previous studies on perceptual decision-making, to eliminate activity bursts in decision-making areas through introducing sudden visual stimuli [10]. The difficulty of the videos is manipulated by adjusting the opacity level, in other words, via the blending coefficient with a random noise overlay (see Fig. 2). The stimulus presentation is done via Psychtoolbox in Matlab [11]. The optimal

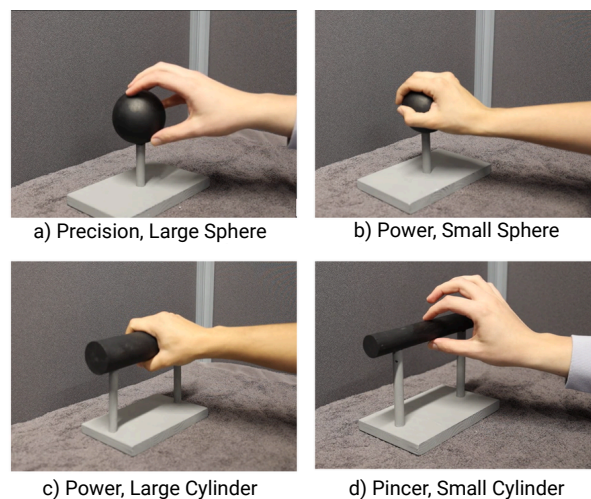


Figure 1: Three different grasps and four different objects used during the experiment, with grasp and object types are indicated below each figure.

opacity level is determined during the practice session by using staircase procedure utilizing Quest+, to ensure balanced distribution of confidence ratings and to train the participants in terms of the timing of the grasps [12].

By presenting ambiguous video clips for motor action selection and planning, uncertainty is induced at the level of perceptual decision-making before any movement is initiated. By inducing uncertainty on the sensory stimuli itself that participants base their decisions, we expect to induce uncertainty on subsequent grasping action. The second stage is characterized by translating the motor intent established through perceptual decision into a grasping alternative among 3 options. Displaying the motor execution cue right after the displays prevents additional (extensive) re-evaluation of the stimuli, thus the internal uncertainties are reflected in grasp execution. The BCI decoder takes the EEG signal from this stage as an input with the goal of returning ground-truth grasp category (i.e. actually realized grasp alternative), and uncertainty for its output is calculated. The third stage is characterized by participants' metacognitive evaluation regarding stage two, which is expected to reveal trial-by-trial fluctuations in the internal (un)certainties for grasp execution. This evaluation is realized through explicit instruction for reporting the certainty while executing the grasp rather than during video displays and with a scale ranging from 1-to-9. Participants are explicitly instructed prior to the experimental session to report the feeling of certainty while executing the grasps, rather than judgements on the accuracy of their decisions. Since the BCI classifiers are trained on EEG data during grasping itself, reporting the uncertainty in that same stage provides a more comparable measure. However, we still expect indirect coupling between perceptual decision and motor execution uncertainty. On the other hand, single-trial LDA classifier uncertainty for the executed grasp type is quantified as its distance to the linear decision boundary constrained from 0 to 1.

At the end of each trial, participants are also asked about which grasp they performed to have the ground truth labels for decoding. There were 384 trials in the actual experiment, equally balanced over objects, actors, and grasp types. In addition, participants are also given the option to report 'none' for the trials they simply did not perform any grasp in case of attentional lapses.

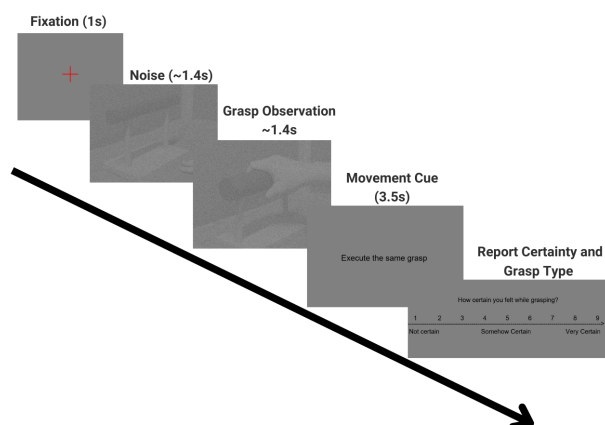


Figure 2: The experimental paradigm and the timeline of a trial. First, following a fixation cross and pre-stimulus noise, participants view the grasping videos where the actors reach for the objects and grasp them accordingly. Following this, they are asked to execute the same grasp type. In the end of each trial, they are asked about which grasp type they executed and also how certain they felt while executing the movements.

Data Collection: The EEG data is collected from 8 participants (age range: 21 - 29) with 500Hz of sampling rate, AFz channel as ground and FCz as online reference via the LiveAmp 64 (Brain Products GmbH, Germany). 50 Ag/Cl electrodes are placed on the scalp to have a good coverage of the fronto-central, central, and centro-parietal areas and hence motor and decision-related activity areas. Some electrodes for the anterior and temporal areas are repurposed for EOG and EMG recordings. Specifically, we used 3 EOG channels to capture horizontal and vertical eye movements for facilitating artifact rejection. 2 EMG channels are placed on the right forearm of the participants to measure the activity during grasping and to capture movement onset. The remaining electrodes in temporal and anterior areas are discarded.

EEG Preprocessing and Feature Extraction: EEG data analysis is performed in EEGLAB [13]. Data is first cleaned from the break periods, then band-pass filtered from 0.1 to 40Hz using a Butterworth 4-th order, zero-phase filter. Then, data is downsampled from 500Hz to 100Hz. ICA decomposition followed this step, for which only EEG channels are considered. Following the decomposition, individual components are rejected based on SASICA, with an ADJUST procedure and correlation with horizontal / vertical EOG channels [14], visual inspection and alignment with ICLabel [15]. Remaining components are projected back and additional filtering from 0.1 to 3Hz using Butterworth 4th order zero-phase filter is performed to compute MRCPs. For EMG chan-

nels, the data is band-pass filtered from 20 to 250Hz, notch-filtered at 50Hz, and averaged across the two electrodes. The onset detection is done by epoching the continuous data from -1 to 3.5 seconds with respect to 'execute' prompt onset. From there, Teager-Kaiser Energy Operator (TKEO) is calculated due to previous results suggesting enhanced onset detection [16]. The first time samples after the movement cue that crosses the z-threshold of 10 is marked as the movement onset, while 1 second period before the movement cue is taken as a baseline. The movement onset provided an additional informative feature to epoch the EEG data for MRCPs.

Classification of Grasp Type: To classify different grasp types, regularized LDA with shrinkage is used. The MRCPs starting from 1 seconds prior to and lasting until 1 seconds after the movement onset is taken as an input for an LDA classifier. Data is further reduced to channels around supplementary and primary motor areas, and decimated by a factor of 5 to have manageable amount of features for each trial. Data from 2 subjects are excluded from this analysis due to unreliable detection or failure to detect trial-wise EMG onsets. The trials for which the participants did not execute any grasp or if the movement onset is not reliably calculated are discarded. The decoding accuracy and confusion matrix is obtained for each participant using a stratified 5-fold cross validation, and in the end predictions are aggregated over the participants.

RESULTS

Behavioral Results: To analyze the behavioral data, we first removed the trials in which the participant reported that they could not perform any grasp within the instructed time limits. On average, 12.75 trials (std: 16.65) are removed. This resulted in eliminating a total of 102 trials out of total 3072 trials from 8 participants. We refer to accuracy as executing the same grasp type among the three alternatives with respect to the grasp video. We observed that accuracy was highly variable between the subjects, ranging from 0.38 to 0.71. Next, we examined if the reported certainties at the end of each trial can be predicted by accuracy or executed grasp type. The summary of the certainty scores across accuracy and grasp types can be seen in Table 1. For that purpose, we used a linear mixed model with subject as the grouping variable. Results showed a significant main effect of accuracy ($z = 8.74, p < .001$), for which correct trials are associated with higher certainties ($\beta = 0.61$). This pattern suggests that participants can, at least to some extent, monitor the accuracy of their perceptual decisions hence being more confident in their subsequent movement. Additionally, a significant effect of Grasp Type was also observed ($z = 2.91, p = .004$). From Table 1, it can be seen that this significance may be explained by higher certainty associated with executing the power grasp compared to other two alternatives. This stems potentially by the fact that power grasp requiring higher force and lower precision,

Table 1: Certainty Levels Across Accuracy and Grasp Types

	Mean(SD)	Count
Accuracy		
Correct	5.69 (2.45)	1674
Incorrect	4.78 (2.59)	1296
Grasp Type		
Pincer	5.23 (2.51)	821
Precision	5.14 (2.51)	1141
Power	5.51 (2.63)	1008

leading to higher certainty estimates than pincer and precision grasps.

Neurophysiology: For neurophysiological investigations, determining the precise movement onset is important since onsetting with respect to the motor cue may not provide the most informative time window. Movement execution time may differ significantly between participants and between trials of the same subject following such a noisy visual stimuli. For that purpose, we used EMG electrodes and have found that mean movement onsets are indeed variable among participants, ranging from 333 to 698 milliseconds. Using Spearman correlation, we also investigated if there is a significant relation with movement onset and reported certainties. The results showed a significant negative correlation consistently for each subject (all $p < .05$, with r_s values ranging from -0.20 to -0.49). This pattern indicates that if the participants are uncertain, they may use additional time to potentially decide between the motor alternatives and start grasp execution. We also investigated the activity in centro-parietal areas, which was previously linked with perceptual decision-making and confidence for a given decision. The patterns we observed also supports the previous interpretation regarding using additional time (see Fig. 3a). Specifically, the rising activity is observed until movement onset (i.e., detected by the EMG electrodes) rather than cue onset (i.e., when participants are instructed to execute the same grasp type). This finding suggests that decision-making may still continue even after the visual stimuli is no longer present on the screen. [17].

Another observed pattern from EEG data suggested an early motor planning with negative MRCP peak in the contralateral side of the brain with respect to the grasping hand. That is, following the presentation of grasp videos but before the movement cue, negative going potential is peaked in C3 channel (see Fig. 3b). This might be the result of participant's anticipation regarding starting the movement as soon as the videos end, and might be further contributed by action observation during video displays. Therefore, investigation of EEG data points towards the interpretation that motor planning starts before or concurrently as the evidence accumulation for perceptual decision determining the details of the subsequent action.

Classification Results: To classify between different grasps, shrinkage regularized LDA is utilized with MRCPs from -1 to 1 seconds of trial-wise movement onsets as features. The results over 6 subjects is summa-

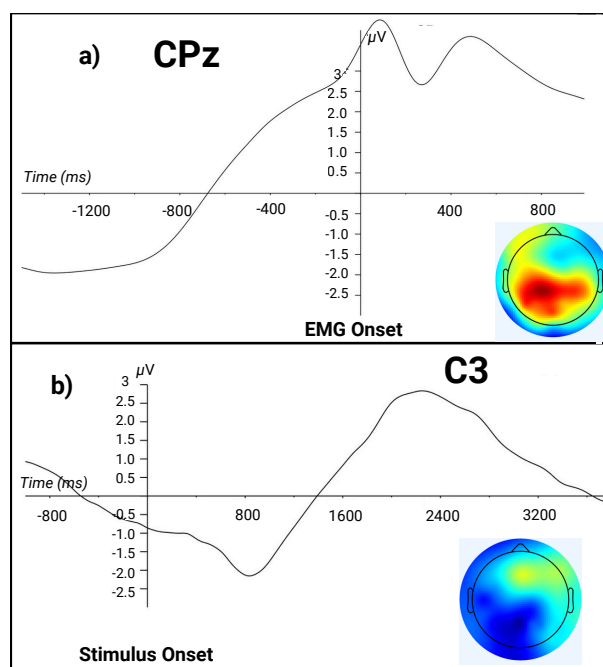


Figure 3: CPz and C3 waveforms when averaged over the subjects. a) CPz channel activity in grand-averaged data over 6 subjects, along with the scalp map representing the mean activity for -200 to $+200$ ms from movement onset. The activity rises until the movement onset. b) C3 channel activity in grand-averaged data over 8 subjects is shown with the scalp map representing the mean activity around 600 to 800 ms from stimulus (grasp video) onset. The peak negativity is observed around 800 ms following the stimulus onset.

rized in Fig. 4. The ROC-AUC scores across participants range from 0.5 to 0.62 , suggesting poor decodability using these features. This might be due to combination of the limitations regarding experiment design that will be detailed in discussion, EEG signal-to-noise ratio, or class imbalance as can be seen from Tab. 1). Indeed, predictions tend to favor precision grasp compared to other grasp types, which is the most reported category. The relationship between classification and human uncertainties are investigated within-subject manner. The class probability for the ground-truth label can be interpreted as its distance to the decision boundary constrained in the range from 0 to 1 . When analyzed this way, Spearman correlation test was only significant for the subject having the highest ROC-AUC score of 0.62 ($\rho = 0.14$, $p < 0.05$). Although it is promising to see this pattern for the subject with the highest classification metric, the lack of statistical significance for the other subjects necessitates collecting data from more participants before reaching any conclusions regarding the relationship between classifier and human-level uncertainties.

DISCUSSION

In this study, we developed a paradigm inspired by previous perceptual decision-making literature to manipulate difficulty for the grasp videos. This design allowed us to investigate uncertainties from human reports and BCI

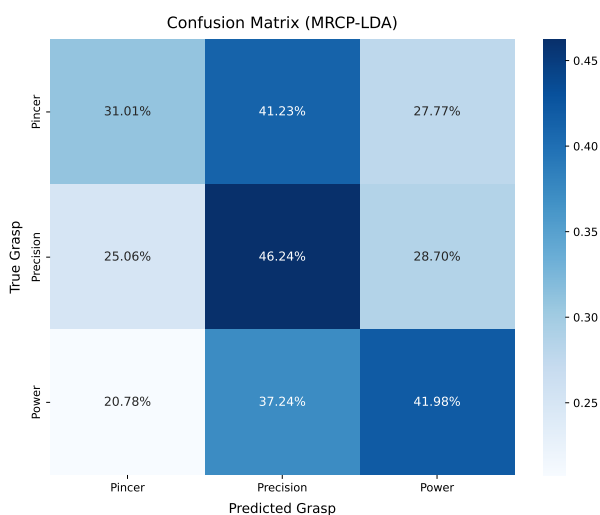


Figure 4: Confusion matrix when the predictions and true labels are aggregated over 6 subjects. Among the grasp types, precision grasp is the best classified category followed by power grasp. The LDA predictions are biased towards precision grasp, likely due to class imbalance.

classifiers for the first time in the field. While the data collection is still ongoing, the decoding performances are variable across participants. Poor decoding performances can be partly explained by low signal-to-noise ratio of EEG, especially for decoding multiclass grasp scenarios where there is no immediately distinct underlying activity. However, we observed a significant relationship between LDA classifier uncertainty and participant reports for the subject with highest ROC-AUC metric. This pattern indicates that as the classification performance improves, classifiers and subjective reports might be utilizing the similar neural representations. However, it is not yet possible to reach any final conclusions with the current state of data collection and analysis as we require more subjects and improved decoding performance for the already collected data from other participants.

Interestingly, while previous research demonstrated higher performance for decoding pincer grasps [3], we did not see this pattern in our data. In addition to class imbalance, methodological differences might play a role in this difference, since the participants are not given any objects while performing the grasps. This choice is deliberately made to use videos with same grasps but different objects to increase variability and prevent participants to decide based on the low-level cues. However, it might lead to decreased decodable information from the brain. Indeed, for that study, decoding accuracy for pincer grasp is at its highest when participants were releasing the objects [3].

Limitations: The choice of not using the objects while participants are executing the grasps may lead to more unnatural feeling accompanying the movements. It might therefore influence reporting the certainty of the movement. Additionally, we explicitly instructed participants for reporting the uncertainties during grasping rather than during video observation. However, this might be overly

complex for some people that finds it difficult to disentangle the two and reporting metacognitive evaluations. The feeling of certainty might also be affected significantly by the grasp type in general that normally requires more force, such as power grasp. Therefore, this needs to be kept in mind for the future analysis.

Future Direction: The prominent models for perceptual decision-making are based on the idea of evidence accumulation until some decision bound is reached. This formulation views the decision process as gathering sufficient sensory evidence from the environment to come up with a choice [18]. In humans, a signal showing these characteristics is found in EEG studies. An event-related potential, called centro-parietal positivity (CPP), is demonstrated to rise following the stimulus displays and until the response. Previous studies reported that especially the rising slope of CPP aligns with task difficulty [10]. In addition to decision-making related characteristics, there is a growing body of research connecting CPP to decision confidence. Since CPP is generally thought to index decision-making process itself, using CPP as a candidate signal challenges the traditional views of confidence as emerging post-decision. Researchers provided supporting evidence for this view and suggested that choice confidence might be represented in the very same mechanisms underlying decision-making [19]. In a recent formulation of decisions involving motor reports, researchers proposed a leaky accumulation process between perceptual and motor systems, where the leakage is controlled by confidence [20]. Following these lines of research, it might be plausible to connect this EEG signals from centro-parietal areas for indexing confidence as the decisions evolve to guide subsequent motor actions. Indeed, we have observed that CPz activity rises until the movement. Therefore, using peak amplitudes before the movement onset might provide an additional uncertainty metric to compare with human-level or classifier-level uncertainties.

Previous studies for uncertainty quantification have traditionally categorized it into two main sources, namely epistemic and aleatoric uncertainty [21]. The first one refers to uncertainty that can be improved with more training since it resulted from model incompetence, while the latter is the uncertainty that cannot be reduced and due to inherent stochasticity of the input. Despite the abundant use of this formulation, it should be treated cautiously for how well the two sources can be disentangled [6]. Evidential deep learning (EDL), on the other hand, views uncertainty in vacuity and dissonance terms. The first refers to incomplete evidence for the decision alternative while the latter refers to conflicting evidence among decision alternatives. By learning Dirichlet Density, predictions are modeled as distributions over potential softmax outputs [22]. More specific to our task design, it might be that the uncertainties that participants report resonates more with EDLs approach. That is, participants either have incomplete or conflicting evidence for different grasp alternatives. Therefore, future research

by calibrating the loss functions of shallow neural networks, such as EEGNet [23], to return Dirichlet parameters might be promising.

CONCLUSION

Our study introduces a new paradigm for BCI research on grasp decoding to connect uncertainties returned from the classifiers, from relevant EEG indices for perceptual decisions, and from human metacognitive judgements. Such a holistic approach to investigate the relationship between human behavior and machine learning algorithms is crucial for improving trust and comprehensive understanding for BCI systems. We think that this approach will also be fruitful to develop adaptive BCI systems that are informed by human-aligned uncertainty representation inherent in behavior and decisions.

REFERENCES

- [1] Wolpert DM, Landy MS. Motor control is decision-making. *Current Opinion in Neurobiology*. 2012;22(6):996–1003.
- [2] Sburlea AI, Müller-Putz GR. Exploring representations of human grasping in neural, muscle and kinematic signals. *Scientific Reports*. 2018;8(1):16669.
- [3] Sburlea AI, Wilding M, Müller-Putz GR. Disentangling human grasping type from the object's intrinsic properties using low-frequency eeg signals. *Neuroimage: Reports*. 2021;1(2):100012.
- [4] Schwarz A, Pereira J, Kobler R, Muller-Putz GR. Unimanual and bimanual reach-and-grasp actions can be decoded from human eeg. *IEEE transactions on biomedical engineering*. 2020;67(6):1684–1695.
- [5] Ritchie JB, Kaplan DM, Klein C. Decoding the brain: Neural representation and the limits of multivariate pattern analysis in cognitive neuroscience. *The British Journal for the Philosophy of Science*. 2019;70(2):581–607.
- [6] Jong IPd, Sburlea AI, Valdenegro-Toro M. Uncertainty quantification in machine learning for biosignal applications – a review. 2025;(arXiv:2312.09454).
- [7] Suurmeijer J, Jong IPd, Valdenegro-Toro M, Sburlea AI. Uncertainty quantification for motor imagery bci – machine learning vs. deep learning. 2025;(arXiv:2507.07511).
- [8] Hassan M. Confidence-based early classification for motor imagery brain-computer interfaces. 2024.
- [9] Duan T, Wang Z, Liu S, Yin Y, Srihari SN. Uner: A framework for uncertainty estimation and reduction in neural decoding of eeg signals. *Neurocomputing*. 2023;538:126210.
- [10] Kelly SP, O'Connell RG. Internal and external influences on the rate of sensory evidence accumulation in the human brain. *Journal of Neuroscience*. 2013;33(50):19434–19441.
- [11] Brainard DH. The psychophysics toolbox. *Spatial Vision*. 1997;10(4):433–436.
- [12] Watson AB. Quest+: A general multidimensional bayesian adaptive psychometric method. *Journal of Vision*. 2017;17(3):10.
- [13] Delorme A, Makeig S. Eeglab: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*. 2004;134(1):9–21.
- [14] Chaumon M, Bishop DVM, Busch NA. A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*. 2015;250:47–63.
- [15] Pion-Tonachini L, Kreutz-Delgado K, Makeig S. Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*. 2019;198:181–197.
- [16] Carvalho CR, Fernández JM, Ama AJ del, Oliveira Barroso F, Moreno JC. Review of electromyography onset detection methods for real-time control of robotic exoskeletons. *Journal of NeuroEngineering and Rehabilitation*. 2023;20:141.
- [17] Grogan JP, Rys W, Kelly SP, O'Connell RG. Confidence is predicted by pre- and post-choice decision signal dynamics. *Imaging Neuroscience*. 2023;1:imag-1–00005.
- [18] Ratcliff R, McKoon G. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural computation*. 2008;20(4):873–922.
- [19] Gherman S, Philiastides MG. Neural representations of confidence emerge from the process of decision formation during perceptual choices. *NeuroImage*. 2015;106:134–143.
- [20] Balsdon T, Philiastides MG. Confidence control for efficient behaviour in dynamic environments. *Nature Communications*. 2024;15(1):9089.
- [21] Hüllermeier E, Waegeman W. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*. 2021;110(3):457–506.
- [22] Sensoy M, Kaplan L, Kandemir M. Evidential deep learning to quantify classification uncertainty. 2018.
- [23] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. Eegnet: A compact convolutional network for eeg-based brain-computer interfaces. *Journal of Neural Engineering*. 2018;15(5):056013.