

RESEARCH ARTICLE

Measurement and analysis of cognitive load associated with moving object classification in underwater environments

Arunim Bhattacharya^a and Sachit Butail^a

^a Department of Mechanical Engineering, Northern Illinois University, Dekalb, Illinois, USA

ARTICLE HISTORY

Compiled June 13, 2023

ABSTRACT

Visual analysis in field science experiments often involves classifying objects on experimental images and videos. In this context, developing a reliable and independently validated estimate of mental workload during object classification can enable cognitively responsive task allocation. The goal of this study is to quantify the cognitive load perceived by humans from electroencephalography (EEG) data during an underwater object classification task that was inspired from citizen science studies. During the task, participants were asked to identify one of three possible invasive fish species in short videos of a virtual underwater environment. The virtual environment was modeled to vary fish behavior and environmental factors that are known to be critical in classification. A contextually-relevant secondary task was designed to provide independent validation of cognitive load measures. Several established measures of cognitive load were compared across different weightings on the scalp positions, and the measure that strongly associated with reaction time and a secondary task accuracy was selected for further analysis. Our results show that cognitive load calculated using the difference in power of α frequencies best correlates with reaction time and secondary task accuracy. When fit to the environmental factors, cognitive load calculated using this approach was high when the environment was turbid and the fish moved at high speeds. Results from this study have applications in cognitively-responsive human computer interaction and in developing shared control strategies in human robot interaction.

KEYWORDS

Cognitive load; object classification; underwater video; EEG; secondary task

,

1. Introduction

Humans perform visual recognition and classification of objects at all times (Riesenhuber & Poggio, 2000; Logothetis & Sheinberg, 1996). Depending on the object to be identified, the classification task can be easy, as we do almost every minute for common place scenarios, or specialized, requiring domain knowledge, as in the case of recognition of lesions in surgical applications (Ariji et al., 2019) and identification of humans in search and rescue (Rodin et al., 2018). Compared to static objects, classifying moving objects poses further challenges associated with localization, segmentation, intra-object variation, and a dynamic background (Zhang, Li, Yuan, & Xiang, 2007).

A common application of moving object classification in citizen science involves processing and tagging large amounts of field video data to identify plants and animals of interest (Starr et al., 2014; Wick et al., 2020; Langenkämper, Simon-Lledó, Hosking, Jones, & Nattkemper, 2019; Laut, Henry, Nov, & Porfiri, 2013; Suzuki-Ohno et al., 2022). With advances in robotic monitoring, the amount of such data has increased by orders of magnitude over the past decade (King, George, Buckle, Novak, & Fulton, 2018; Mallet & Pelletier, 2014) creating a need for intelligent allocation of human resources that can optimize the performance of the classification task over a large group of volunteers (Trouille, Lintott, & Fortson, 2019). Object classification, where multiple objects exist in the same field video is especially hard, and entails dealing with difficult lighting conditions, dynamic environments, and speed and proximity of the objects in question (Salman et al., 2016).

In this context, a better understanding of the mental workload associated with object classification can (a) allow smart assignment of responsibilities within a citizen science project (Palermo, Laut, Nov, Cappa, & Porfiri, 2017) where classification tasks are distributed among volunteers in relation to the imposed cognitive demands and (b) enable online control sharing strategies within remotely operated monitoring robots that can weigh autonomous actions against those of a remote operator based on mental burden experienced by the operator.

The mental burden experienced by a human while undertaking a task is often quantified in terms of cognitive load (Klauer & Zhao, 2004; Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Cognitive load can be measured indirectly based on performance in a contextually related secondary task (J. R. Anderson, Reder, & Lebiere, 1996; Engle, 2002; Sweller, 1988), task completion time (Braver et al., 1997), and responses to workload related questionnaires (Brünken, Seufert, & Paas, 2010). Compared to indirect measures, direct measurement of cognitive load using electroencephalography (EEG) offers the potential to realize real-time responsive systems. With respect to EEG, measures of cognitive load are generally functions of spectral powers of the frequencies that lie within the 4–15 Hz range, with select frequency bands called α (9.5–11.5) and θ (4–9.5) that have been shown to be closely associated with working memory (Klimesch, 1999).

Despite its popularity, there is not a single measure of cognitive load based on EEG data that works for all applications. The different measures include the theta-alpha-ratio (TAR) (Trammell, MacRae, Davis, Bergstedt, & Anderson, 2017), difference in α power (Klimesch, 1999), difference in α and θ powers (E. W. Anderson et al., 2011), and features such as entropy, energy and standard deviation extracted from wavelet coefficients (Zarjam, Epps, & Lovell, 2015). The different measures likely highlight the dependence of cognitive load measurement on the task at hand. For example, cognitive load during tasks that involve visualization and pattern recognition has been quantified using α power shifts (E. W. Anderson et al., 2011), whereas TAR has been used in markers for Alzheimer’s disease (Schmidt et al., 2013). Expectedly, there is a lack of consensus regarding which measure would serve to best quantify cognitive load during a task (Kiiski et al., 2020). Furthermore, when computing cognitive load using spectral shifts, additional variables include (E. W. Anderson et al., 2011): (a) sites on the scalp, (b) weights given to each site, and (c) the type of baseline or reference measurement; baseline, for example, may be recorded once at the beginning of the experiment for all trials (Bales & Kong, 2022) or prior to every trial (Antonenko, Paas, Grabner, & van Gog, 2010).

To address this large parameter space as well as identify a reliable measure of cognitive load for object classification, we developed an experimental design that measures

reaction time and includes a contextually related secondary task for validation. In particular, the primary task involved moving object classification and the secondary task involved counting the number of moving objects. To motivate object classification in a citizen-science context (Wick et al., 2020), we designed our experiments after fish species identification task created using virtual environments. We utilized virtual environments instead of real-world videos as they enable systematic variation of environmental properties. Specifically, based on real-world challenges (Wick et al., 2020), we varied several variables in our virtual environment, namely fish speed and number, turbidity within the environment, and the distance of the camera to the group of fish.

The contributions of this work are as follows: (i) we identify an independently validated EEG-based measure of cognitive load during moving object classification with an experimental protocol that involves maximizing the strength of association against reaction time a contextually relevant secondary task, and (ii) we utilize the measure of cognitive load to test the hypotheses that cognitive load will increase when identifying objects that are moving at higher speeds, in larger numbers, are farther, and in environments that are more turbid.

This paper is organized as follows: section 2 describes the experimental methods including the modeling of the virtual environment and the fish behavior, experimental setup and procedure, and data analysis towards determining an independently validated measure of cognitive load. Section 3 presents the experimental results describing the responses to various questionnaires surveying mental workload and perception of the underwater environment followed by identification of the measure of cognitive load that best associates with reaction time and secondary task accuracy; the results of generalized linear model fits of cognitive load to variations in virtual environment are presented next. Section 4 concludes with a discussion of the results and applications of this work.

2. Methods

2.1. Virtual Environment

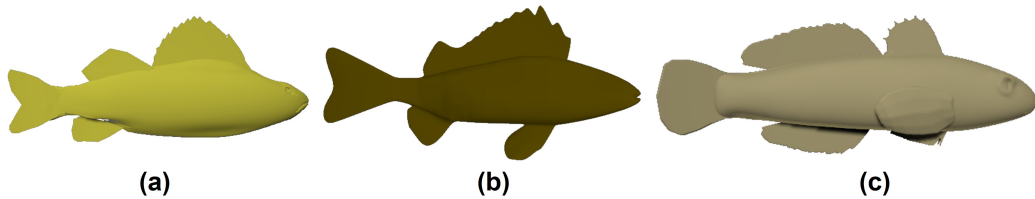


Figure 1. Fish models used in the experiment. Yellow perch (a), eurassian ruffe (b), and round goby (c). Fish size was not matched as two species were never shown together.

The virtual underwater environment consisted of a region designed to mimic a lake bed within the Great Lakes with plants and uneven terrain. A static virtual camera was set up to focus on a specific area within the environment. Three-dimensional virtual replicas of fish found in Lake Michigan, namely round goby (*Neogobius melanostomus*), eurassian ruffe (*Gymnocephalus cernua*), and yellow perch (*Perca flavescens*) were designed to populate the environment, one species at a time. Out of these, yellow perch is considered native to the Great lakes. The fish were modeled so that their outline

matched the pictures from the US Geological Survey database website (USGS, 2022). Specifically, lateral and longitudinal views of the fish were imported into the Blender modeling software and a sculpting tool was used to match the contour and width of the fish. Fish locomotion was further enhanced by animating the tail to mimic carangiform swimming (Tytell et al., 2010). Although the three fish species have different size in nature, we built them to the same size in our environment to maintain the same camera distance for all species. The virtual replicas were then colored according to the dominant color of the species (Figure 1).

Groups of fish replicas moved according to a self-propelled particle model called the zonal model often used to study collective behavior in animal groups (Couzin, Krause, James, Ruxton, & Franks, 2002). This model simulates collective motion among a group of agents by requiring that each agent moves according to three interaction rules; these rules are encoded with respect to other agents that are present within concentric zones centered around the focal agent. In particular, the model entails that (i) agents maintain a minimum separation distance (zor) among themselves, (ii) agents orient in the direction of other agents that are within zone of orientation (zoo) and (iii) agents move towards other agents who are beyond the zoo but within a zone of attraction (zoa). Accordingly, given the position $\mathbf{r}_i(t) \in \mathbb{R}^3$ of the i -th agent, the velocity at time t is $\mathbf{v}_i(t) \in \mathbb{R}^3$ is updated as

$$\mathbf{v}_i(t + \delta t) = \begin{cases} -s \left(\sum_{j \in zor_i(t)} \frac{\mathbf{r}_j(t) - \mathbf{r}_i(t)}{|\mathbf{r}_j(t) - \mathbf{r}_i(t)|} \right) & \text{if } zor_i(t) \neq \emptyset \\ 0.5s \left(\sum_{j \in zoo_i(t)} \frac{\mathbf{v}_j(t)}{|\mathbf{v}_j(t)|} + \sum_{j \in zoa_i(t)} \frac{\mathbf{r}_j(t) - \mathbf{r}_i(t)}{|\mathbf{r}_j(t) - \mathbf{r}_i(t)|} \right) & \text{otherwise,} \end{cases} \quad (1)$$

where $s \in \mathbb{R}$ is the speed of the agent, zor_i , zoo_i , and zoa_i denote the sets of agents within agent i 's concentric zones, and $\delta t = 1$ s is the simulation time step.

2.2. Experimental Conditions

The experimental conditions consisted of varying behavioral and environmental parameters selected on the basis of challenges observed in underwater video classification (Wick et al., 2020) (Fig. 2). These included (i) camera distance from fish (close and far), (ii) water turbidity (low and high), (iii) fish speed (slow = 1 and fast = 1.5 body length/s), and (iv) number of fish (8, 10 or 12). Camera distance was varied by placing the virtual camera closer or farther from the fish group while taking care that a majority of the fish were visible on the screen; water turbidity was varied by adding high to low intensity fog to the scene while making sure that fish were still visible under high intensity condition; fish speed was varied by modifying the parameter s in (1); and number of fish were varied by increasing the number of agents in the simulation. Simulations were first performed in MATLAB (Mathworks inc.) and the trajectory data was imported into Unity software to animate the fish replicas.

With three different fish species, there were 72 possible combinations of classification tasks that could be presented to each participant. Eight sets of videos were preprepared for every combination, with a random video picked per combination when requested for the classification task. Each combination was presented twice, once as a primary task where the participant had to identify the fish species, and then again accompanied by a secondary task, where the participant had to identify the fish species, as well as count the number of fish. The primary task served to provide a basis for measuring cognitive load using EEG data and the secondary task served to provide an alternate measure of

cognitive load in the form of accuracy of the estimate in counting the number of fish. In this sense, including a secondary task for each combination would therefore result in 144 classification tasks. Since presenting so many tasks could lead to unwanted fatigue, each participant was presented with 24 randomly selected combinations out of the 72 total possible combinations, once each for primary and secondary task for a total of 48 classification tasks per participants.

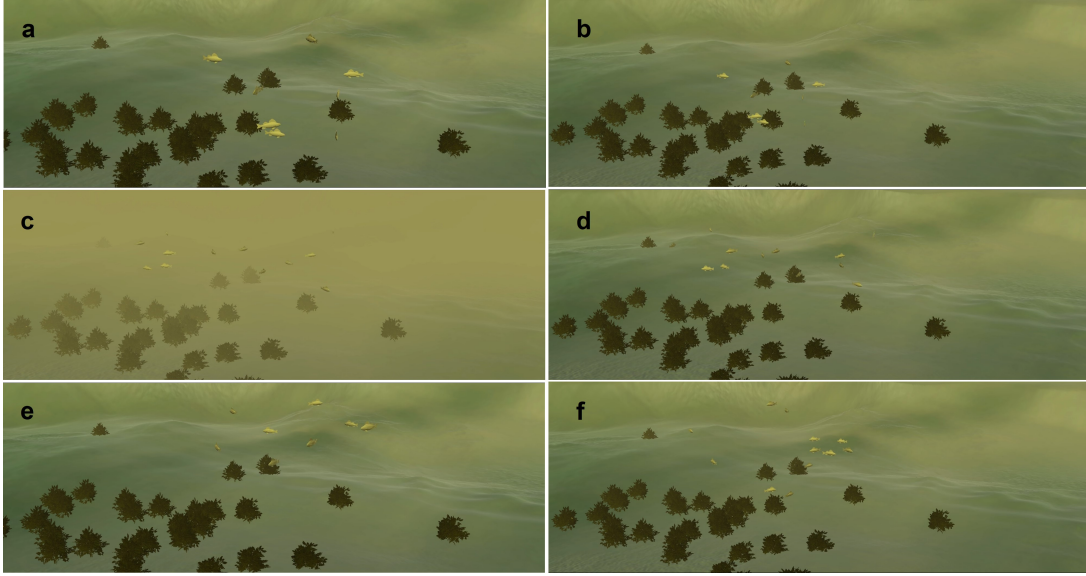


Figure 2. A subset of the different experimental conditions shown to participants. Camera distance which could be close and far (a, b), turbidity levels which could be high and low (c, d), and number of fish that ranged between 8 and 12 (e, f).

2.3. *Participants*

A total of 22 participants (18 males and 4 females, aged 24.4 ± 4 years) were recruited through flyers posted throughout the campus and email announcements. Exclusion criteria included being below 18 years of age, having a history of seizures, or having thick hair on scalp that could prevent obtaining a reliable connection with the EEG headset. The experimental study was approved by the institutional review board under protocol number HS21-0309.

2.4. *Experimental Setup*

The experimental setup consisted of a desktop computer with a wide-screen monitor (3840×1080 pixels, 60 Hz refresh rate, LG) to present the virtual environment, and a fourteen-channel EEG device (Emotiv Epoc X, Emotiv Inc.) running at 128Hz to record brain activity. The electrodes on the EEG headset were positioned according to the international 10-20 system. Within this system, the scalp is considered a hemispherical shape; using the convention F, T, C, P and O for frontal, temporal, central, parietal, and occipital parts of the brain, with odd numbers for left and even numbers for the right part of the brain, an electrode location can be specified using a letter and a number as X#; for example the notation F3 denotes a site located on the left-front part of the scalp.

The graphical user interface (GUI) for presenting videos of the virtual underwater environment was created using windows presentation foundation in visual studio and scripted in C#. The GUI displayed all the instructions, questionnaires, and playback of the virtual underwater environment. A video camera (Hero 5, GoPro Inc.) facing the participant was mounted on top of the monitor with a mirror mounted behind the participant to record their responses. The camera and mirror served to independently verify the synchronization of user responses to video data. A virtual serial link was set up between the GUI and the EEG capture software (Emotiv Pro, Emotiv Inc.) to mark the start and end of classification tasks during the experiment.

2.5. Experimental Procedure



Figure 3. Various screens seen by the participants. Participants were shown a baseline image for approximately eight seconds (a) prior to a fish species identification (b) or estimating the number of fish in addition to identification (c).

On arrival, participants were assigned a random six-digit number to record all information anonymously. They were then asked to review and sign a consent form, which listed the purpose of the study, potential risks, and the type of data that will be collected (EEG, video, and survey responses).

The participants were then asked to put on the EEG headset using a reference image describing how the electrodes should be arranged. Once a reliable electrode connection was achieved, the EEG signal and video data was set to record. This was followed by instructions on the screen, familiarization with the tasks, and object classification.

The instructions comprised showing real images of the three fish species to the participants with a statement informing them that they should try and remember their appearance and that they will be asked to identify virtual replicas of these fish in subsequent trials.

During familiarization, participants were asked to identify the virtual replicas in videos similar to the ones that would be used in classification. Specifically, participants were asked to select one of multiple choices and were informed of the correct choice upon submission. After three such familiarization videos, the classification task was initiated by informing the participants to answer the one or two questions relating to the number and types of fish on subsequent screens “as quickly as possible”.

Prior to every classification task a baseline EEG measurement was collected during which participants were shown a plain blue image covering the screen with a message saying “Please wait while we take baseline measurement” for approximately eight seconds (Fig. 3a).

The classification task involved one of two types of tasks (Fig. 3b and c): identifying the fish species (primary task) or identifying fish species as well as estimating the

number of fish (primary and secondary task). The choices for estimating the number of fish were given in the form of a range around the true value (6–8, 9–11, 12–14). The combination of parameters such as fish species, number of fish, fish speed, camera distance and turbidity, were kept the same between the two types of trials for each participant, whereas the types of combinations and the order in which they appeared was randomized between participants (see Fig. S1, supplementary document for a distribution of parameter combination across all trials.).

After 24 trials, the participants were asked if they wish to take a break or continue with the experiment. At the end of the experiment, participants were asked to enter their preferred gender and their age. This was followed by requesting them to fill out three surveys: (i) a NASA-TLX questionnaire that surveyed their mental workload in response to the experiment as a whole (Cao, Chintamani, Pandya, & Ellis, 2009) on a 0–20 scale, with higher values implying higher workload, pace of task, or higher success; (ii) a fish familiarity questionnaire on a 0–6 scale asking how familiar they were with the different fish species, with higher values implying greater familiarity; and (iii) a virtual video questionnaire asking the participants to rate the realism of the underwater virtual environment and their experience with virtual reality on a 0–6 scale with higher values indicating that virtual environment appeared very natural and that they frequently experienced virtual reality.

2.6. Data Analysis

2.6.1. Preprocessing

EEG data was bandpass filtered to keep frequencies between 0.1 and 20 Hz followed by rejection of trials where any of the frontal electrodes posted an absolute EEG amplitude above 1000 μV (Tong & Thankor, 2009). This resulted in 13% of the trials rejected. (A sensitivity analysis of the results with respect to this threshold was performed to ensure that the selected cognitive load measure did not change because of the threshold.) We favored a simple threshold instead of independent component analysis for artifact removal (Urig  en & Garcia-Zapirain, 2015) to enable real-time and unsupervised calculation of cognitive load in human-computer interaction applications.

Reaction times during a trial were calculated by placing markers within the EEG data via a virtual serial port. Specifically, the reaction time was recorded as the time between when the video was first shown for classification and when the participant submitted their choice. For trials involving primary task only, this involved the time taken to identify the fish species, whereas trials involving primary and secondary task, the reaction time involved identifying the fish species as well as estimating their number within a range.

2.6.2. Secondary task accuracy

Past studies have shown that secondary task performance, when the secondary task is assigned within the same context as the primary task, will decrease with high cognitive load (Sweller, 1988). In our context, this amounts to accuracy in estimating the number of fish on the virtual underwater video. Accuracy in estimating the number of fish during the secondary task was calculated based on participant responses that indicated a range for the number of fish (e.g. 6–8, 9–11). Assuming that a response indicated that all values within a range were equally probable, the error in estimating the number of

fish was calculated as

$$E_s = \frac{1}{3} \sum_i |N_f - N_i|, \quad (2)$$

where N_f is the true number of fish shown on the screen and N_i are the three values within a selected range, so that for example, if the participant selected 6–8, $N_i = \{6, 7, 8\}$, and $|\cdot|$ denotes absolute value. In this measure, a higher value of E_s indicated lower accuracy.

2.6.3. Selecting an independently validated measure of cognitive load

The EEG-based measure of cognitive load for object classification was determined by maximizing the combined strength of association with respect to primary task reaction time and secondary task accuracy. Specifically, we calculated different measures of cognitive load by varying the type of measurement (three types: TAR, reduction in α power, and weighted difference of α and θ power), the temporal location of the baseline measurement (two types: single baseline recorded prior to all trials or a different baseline considered prior to each trial), and the weighting of the electrodes (thirteen different combinations discussed next); these measures were then correlated with reaction time and secondary task accuracy E_s . The measure that maximized the product of the R^2 values with respect to the primary task reaction time and E_s was selected as cognitive load.

The three different types of measurements that were considered included the: (i) TAR (Cabañero et al., 2019), (ii) reduction in α power band (α -diff) (Klimesch, 1999), and (iii) difference in θ and α powers weighted by difference in mean frequency ($\theta - \alpha$ -diff) (E. W. Anderson et al., 2011). Out of these only the TAR measure used select channels and did not require weighting; the latter two involved weighting individual channels along the scalp.

The TAR measure of cognitive load for a particular trial t is defined as (Trammell et al., 2017)

$$TAR(t) = \frac{I_t^{FZ}(\theta)}{I_t^{PZ}(\alpha)} \approx \frac{I_t^{F3}(\theta) + I_t^{F4}(\theta)}{I_t^{P7}(\alpha) + I_t^{P8}(\alpha)}, \quad (3)$$

where $I_t^{FZ}(\theta)$, for example, denotes the spectral power within the θ band at the FZ electrode. Since the FZ electrode was not available on our headset, the same was approximated by adding the spectral power in $F3$ and $F4$ electrodes. The spectral power was calculated using the fast-fourier transform (FFT). For all cognitive load calculations we considered θ band frequencies to lie within 4–9.5 Hz and α frequencies to lie within 9.5–11.5 Hz corresponding to young healthy adults (Klimesch, 1999).

The α -diff, or reduction in α power as a way to measure cognitive load is based on the phenomena that α waves desynchronize with increasing mental load (Klimesch, 1999). Because desynchronization should lead to reduced power, cognitive load based on reduction in α power for a channel k when comparing test to baseline measurements and is determined as (Klimesch, 1999)

$$\Delta I^k(\alpha) = I_b^k(\alpha) - I_t^k(\alpha), \quad (4)$$

where $I_b^k(\alpha)$ denotes the power in the α band during a baseline task on the k -th electrode, and subscript t denotes the same during a trial. The net cognitive load is calculated as a weighted combination of reductions in individual channels in a 14-channel headset as

$$L_\alpha(t) = \sum_{k=1}^{14} w_k \Delta I^k(\alpha), \quad (5)$$

where w_k is the weighting for k -th electrode.

The $\theta - \alpha$ -diff, or difference in α and θ powers weighted by the difference in mean frequency is motivated by the observation that desynchronization in α frequencies may be accompanied by synchronization of θ rhythms thus leading to an increase in power of frequencies within the θ band (E. W. Anderson et al., 2011). The resulting measure of cognitive load is weighted by the shift in mean frequency within a particular band. The mean (or gravity) frequency of ω (θ or α) band at k -th electrode is calculated as (E. W. Anderson et al., 2011)

$$\overline{f(\omega)}^k = \frac{\sum_{i=0}^{n-1} I^{k,i}(\omega) f^{k,i}(\omega)}{\sum_{i=0}^{n-1} I^{k,i}(\omega)}, \quad (6)$$

where n , equal to one more than half the sample size, is the number of bins in individual frequency bands, $f^{k,i}(\omega)$ is the frequency in bin i and $I^{k,i}(\omega)$ is the energy density computed in the bin i . The frequency shift of a wave band is $\Delta \overline{f(\omega)}^k = \overline{f_t(\omega)}^k - \overline{f_b(\omega)}^k$, where the subscripts t and b as before represent trial and baseline measurements. The net cognitive load is then calculated as weighted combination of shift in energy density over all the electrodes as

$$L_{\theta-\alpha} = \sum_{k=1}^{14} w_k \left(\Delta I^k(\theta) \Delta \overline{f(\theta)}^k - \Delta I^k(\alpha) \Delta \overline{f(\alpha)}^k \right), \quad (7)$$

where w_k is the weighting parameter for k -th electrode.

2.7. Weighting of cognitive load for each electrode

To explore the different electrode-weighting strategies, EEG data from various channels was weighted according to a two dimensional Gaussian distribution centered on four different locations on the scalp determined along the azimuth and latitude (Böcker, van Avermaete, & van den Berg-Lenssen, 1994). The four selected locations were symmetrical across left and right hemisphere. For each location the Gaussian distribution was varied in terms of three values of standard deviations (0.1, 0.5, 1.0) resulting in a total of twelve different weighting combinations; a thirteenth combination corresponded to uniform weighting in all electrodes. Because there is research suggesting that the frontal cortex is associated with working memory (E. W. Anderson et al., 2011; Braver et al., 1997), the weighting was only applied to frontal electrodes thus ignoring temporal and parietal regions. Figure 4 shows the weighting combinations evaluated as part of this study.

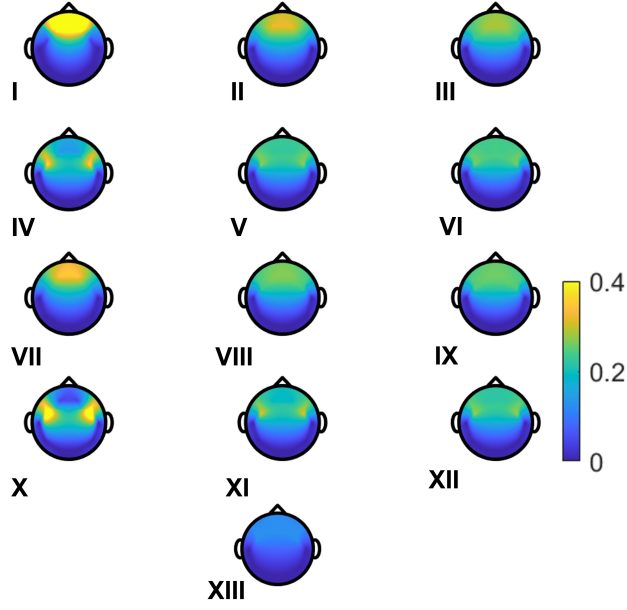


Figure 4. Cognitive load from different channels were weighted according to thirteen different combinations shown on topographic images of the scalp. Combinations I–XII corresponded to Gaussian distributions for frontal electrodes centered at different locations on the scalp with increasing standard deviations. Combination XIII corresponded to uniform weighting across all frontal electrodes.

2.8. Statistical analysis

Cognitive load from each measure for different weightings and baseline measurements wherever possible was linearly correlated with z-scored values of primary task reaction time and secondary task accuracy. For each correlation, an estimate of the slope, the R^2 value capturing the strength of association, and a p value denoting the significance of fit were calculated. The measure of cognitive load that maximized the product of the strength of association with reaction time and secondary task accuracy was then used to model cognitive load and evaluate it as a function of experimental conditions.

The selected measure of cognitive load was then modeled as a generalized function of turbidity of the environment (categorical with two levels), number of fish (categorical with three levels), camera distance (categorical with two levels), and fish speed (categorical with two levels), along with their combinations (Zuur & Ieno, 2016; McCullagh & Nelder, 2019); fish species and participant number were treated as a random intercept. All two-way interaction effects were tested so that a normally distributed generalized linear model (GLM) with an identity link function was set up as

$$\begin{aligned}
 \text{Cognitive load} &\sim \text{Gaussian}(\mu) \\
 \mu &= \text{Turbidity} + \text{NumberFish} + \text{FishSpeed} \\
 &\quad + \text{CameraDistance} + \text{Turbidity} \times \text{NumberFish} \\
 &\quad + \text{Turbidity} \times \text{FishSpeed} + \text{Turbidity} \times \text{CameraDistance} \\
 &\quad + \dots \\
 &\quad + \text{FishSpeed} \times \text{CameraDistance}
 \end{aligned} \tag{8}$$

Different GLM models were created and evaluated with increasing number of predictor variables (experimental conditions) to determine the best fit in terms of residual error as well as the number of predictors. Only models with two-way interactions were tested so that they could be interpreted easily. Each model was compared to a previous model using with fewer number of predictors in terms of the deviance, and significance of the deviance. Additionally, for each model, the corrected Akaike Information Criterion (AICc) was calculated to determine which models could be used for further analysis. The model that was significantly different with respect to a comparison with another having fewer predictors and had the lowest AICc was selected to analyze the dependence of cognitive load on experimental conditions.

3. Results

3.1. Mental workload and quality of underwater environment

Table 1. Average response to post experiment questionnaires. Scale for each questionnaire is shown in parentheses.

NASA TLX questionnaire (0–20)		Response
How mentally demanding was the task?		10.818 ± 6.231
How physically demanding was the task?		2.818 ± 3.647
How hurried or rushed was the pace of the task?		10.273 ± 6.017
How successful were you in accomplishing what you were asked to do?		11.182 ± 3.762
How hard did you have to work to accomplish your level of performance?		9.091 ± 4.975
How insecure discouraged irritated stressed and annoyed were you?		6.045 ± 5.296
Fish familiarity questionnaire (0–6)		
How familiar are with Round Goby?		1.32 ± 1.91
How familiar are with Yellow Perch?		1.91 ± 2.09
How familiar are with Eurasian Ruffe?		0.63 ± 1.29
Virtual video questionnaire (0–6)		
How would you describe your past experience with virtual reality?		2.36 ± 1.94
How natural did you find the virtual environment?		2.45 ± 1.56

Table 1 shows the average responses to survey questions posed to participants after the experiment. Participants found the classification tasks to be somewhat mentally demanding, but not physically demanding. They felt that the pace of the tasks was somewhat rushed and that they were generally successful in accomplishing the task. Participants did not feel discouraged or stressed while performing the tasks.

Regarding fish familiarity, participants were generally unfamiliar with the fish. With respect to the virtual environment, participants had less than average experience with virtual reality and found the virtual underwater environment to be slightly less than somewhat natural.

3.2. Independently validated measure of cognitive load

Cognitive load based on TAR did not correlate with secondary task accuracy (estimate = 0.002, $p = 0.96$, $R^2 = 0$) or reaction time (estimate = 0.023, $p = 0.65$, $R^2 = 0$).

When the baseline for the first trial was considered as the baseline for all tasks (single baseline for all trials), and cognitive load was calculated using $\theta - \alpha$ -diff, none of the thirteen weighting combinations were found to be significantly correlated with primary

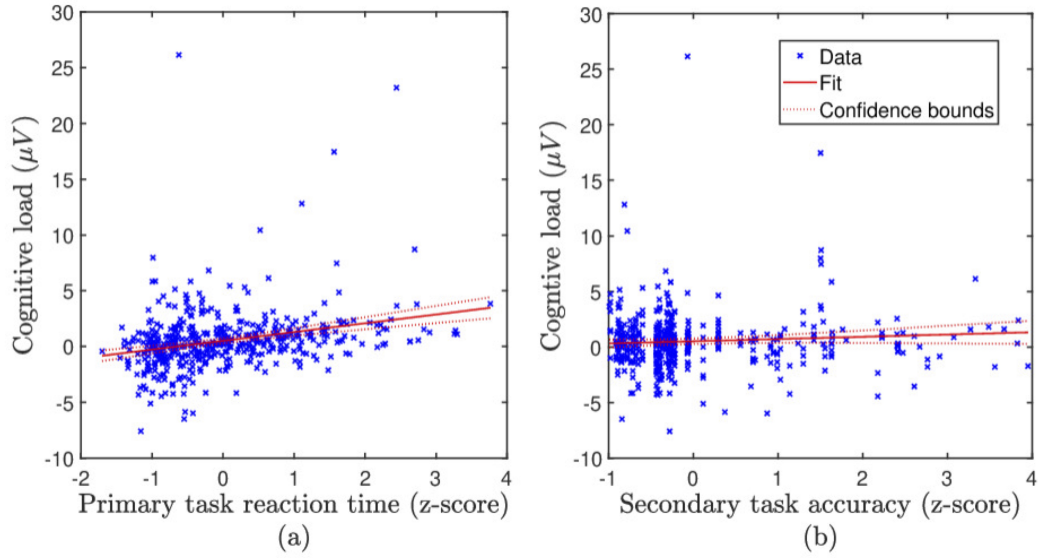


Figure 5. Cognitive load computed using α -diff as a linear function of z-scored reaction time (a) and secondary task accuracy (b) for combination X.

task reaction time or secondary task accuracy (Table S1 in supplementary document). On the other hand, when cognitive load was calculated using α -diff, combination I maximized the strength of association and was positively correlated with primary task reaction time (estimate= 0.679, $p < 0.001$, $R^2 = 0.071$) but was not correlated with secondary task accuracy (estimate= 0.027, $p = 0.814$, $R^2 = 0$).

With baseline recorded prior to every trial, and cognitive load calculated using $\theta - \alpha$ -diff, again, combination X was found to be negatively correlated with primary task reaction time (estimate= -0.080 , $p = 0.041$, $R^2 = 0.009$), but not with secondary task accuracy (estimate = 0.033 , $p = 0.403$, $R^2 = 0.001$, Table S2 in supplementary document). Contrastingly, when cognitive load was computed using α -diff, combination X maximized the strength of association and was positively correlated with primary task reaction time (estimate= 0.786 , $p < 0.001$, $R^2 = 0.0786$) as well as with secondary task accuracy albeit without significance with the latter (estimate= 0.202 , $p = 0.118$, $R^2 = 0.005$). The sensitivity of this correlation to the threshold on absolute EEG amplitude is shown in table S3 in supplementary document. Figure 5 shows the linear fits for this combination and measure of cognitive load to reaction time and secondary task accuracy.

3.3. Effect of environment and fish motion on cognitive load

Table 2 shows the GLMs fit to the cognitive load calculated using baseline recorded prior to every trial, combination X and α -diff with increasing number of experimental conditions (predictors). Each successive model is assigned a number and is compared to another model with fewer predictors. In this respect, model 6 was found to be statistically significant based on its p -value being less than 0.05 and the lowest AICc value. In particular, model 6, which tested interaction between Turbidity and Fish-Speed had a significantly low deviance when compared to model 4, which tested main effects only. Another model (model 9) tested interaction between NumberFish and

Table 2. Generalized linear models with sequentially increasing number of predictors. Each model is compared with respect to another with fewer number of predictors to calculate deviance and AICc.

Model #	Model	Compared to	Df	Residual Df	Residual Deviance	Deviance	AICc	p
0	1							
1	Turbidity	0	1	479	3658.7	NA	2341.12	NA
2	Turbidity + NumberFish	1	2	478	3658	0.7	2343.06	0.77
3	Turbidity + NumberFish + FishSpeed	2	1	475	3653.9	4.1	2346.59	0.76
4	Turbidity + NumberFish + FishSpeed + CameraDistance	3	1	474	3653.4	0.5	2348.58	0.81
5	FishSpeed + CameraDistance + Turbidity*NumberFish	4	2	472	3652.2	1.2	2350.49	0.69
6	NumberFish + CameraDistance + Turbidity + FishSpeed + Turbidity * FishSpeed	4	1	471	3626.6	25.6	2351.26	0.19
7	Turbidity + CameraDistance + NumberFish + FishSpeed + NumberFish*FishSpeed	4	2	469	3597.4	54.8	2348.35	0.04
8	NumberFish + FishSpeed + Turbidity + CameraDistance + Turbidity*CameraDistance	4	1	468	3591.8	60.4	2353.87	0.69
9	Turbidity + FishSpeed + NumberFish + CameraDistance + NumberFish * CameraDistance	4	2	466	3574.5	77.7	2350.36	0.14
10	Turbidity + NumberFish + FishSpeed + CameraDistance + FishSpeed*CameraDistance	4	1	465	3529.6	122.6	2348.49	0.048
					3511	141.2	2349.64	0.09

Table 3. Model that satisfied the criteria for being statistically significant and lowest AICc

Model 6			
Predictors	Estimate	Standard Error	p
Intercept	0.5429	0.33666	0.108
NumberFish10	0.18278	0.30999	0.556
NumberFish12	0.18678	0.30492	0.541
CameraDistance(Far)	0.09599	0.25276	0.704
Turbidity(High)	-0.4497	0.36115	0.214
FishSpeed(Fast)	-0.4603	0.35934	0.201
Turbidity(High):FishSpeed(High)	1.03165	0.50573	0.042

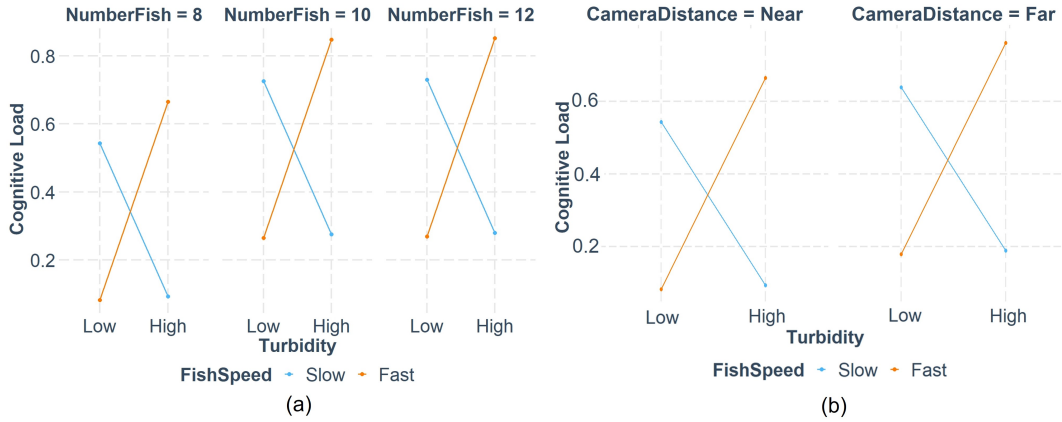


Figure 6. Interaction between Turbidity and FishSpeed for different NumberFish (a) and CameraDistance (b).

CameraDistance also had significantly low deviance than model 4 but it had a slightly higher AICc and a much larger deviance than model 6. Model 6 was ultimately selected for detailed analysis.

Table 3 shows the GLM results for models 6, where we find the interaction between Turbidity and FishSpeed to be significant. Figure 6a investigates this interaction with respect to NumberFish and reveals that high Turbidity leads to higher cognitive load when fish were moving faster for all group sizes. Figure 6b reveals the same effect irrespective of the CameraDistance.

4. Discussion

Measurement of cognitive load reliably in real time presents an opportunity to design efficient human computer interaction strategies where the assigned tasks are responsive to mental workload. Object classification is one such type of task that is often seen in citizen science studies, where volunteers are asked to identify features of interest in field videos, and this work allows interpreting human workload during such tasks. While brain activity through EEG data has often been used as a non-invasive way to measure real-time cognitive load, a single agreed-upon approach to determine which measure to use is not available. In this study we design an experiment that included recording the reaction time and performance in a contextually relevant secondary task to independently validate a measure of cognitive load in moving object classification. The experimental setup was designed to replicate a fish species identification task in virtual underwater environment authored to systematically vary multiple environmental parameters. The measure of cognitive load that correlated with reaction time and secondary task performance was then used to investigate dependence of cognitive load on environmental parameters including turbidity, camera distance, fish speed, and fish group size.

Given that the experiment was designed to be performed sitting at a computer desk, it was not perceived as physically demanding, however, participants felt that the classification tasks were moderately demanding in terms of mental effort and pace. Because the NASA-TLX questionnaire was presented at the end of the experiment, it lacked the granularity needed to make a real-time inference about mental workload. At the same time, these responses indicate that the experiment design was adequate in terms of number and type of classification tasks to be performed and did not pose significant mental workload.

Majority of the participants were unfamiliar with the fish species or their impact on the environment. In this context, the experiment served to increase their awareness of the presence of these fish species in the Great lakes. Participants found the virtual environment to be less than somewhat natural indicating that additional features must be added for believability. While for this study, we kept the environment simplistic to focus on the object classification task, additional features that could be added include color patterns on the fish body, speed dependent tail movement, and a greater variety of plants and aquatic organisms in the environment.

The measure of cognitive load that achieved the strongest association with primary task reaction time and secondary task performance was calculated by considering baseline recorded prior to each trial and strongly weighting the frontal electrodes $FC5$ and $FC6$. Baseline prior to every trial is likely required in situations where the task difficulty changes nonlinearly with time. The location of electrodes that provided the validated measure is supported by studies suggesting that prefrontal cortex in humans brain play role in cognitive load (Nissim et al., 2017; Sweller, 1988). While this measure significantly correlated with reaction time it failed to do so with secondary task accuracy. It is likely that the effect of secondary task on mental workload was diluted due to the relatively few number of fish which participants could also count manually instead of estimating. Indeed, the experimenter had noted a few instances where participants were seen counting the number of fish on the screen.

The α -diff measure of cognitive load was more robust as an estimate of cognitive load compared to $\theta - \alpha$ -diff. While the desynchronization in α power and synchronization in θ power as the brain switches from rest to trial state is well accepted (Klimesch, 1999; Gómez-Ramírez, Freedman, Mateos, Perez Velazquez, & Valiante, 2017; Solomon

et al., 2017), an accompanied increase in mean frequency is related to the speed of processing information (Klimesch, 1999); this implies that depending on how the task is handled, the change in mean frequency can lead to positive or negative cognitive load, an aspect we observe in our calculations as well.

Within our object classification experiments, no single factor was found to be significant enough to explain cognitive load; instead cognitive load was found to depend on a combination of factors. The lack of main effects is likely because of the relatively low resolution for some of the factors. For example, factors such as NumberFish had similar values (8,10, and 12) to avoid guesswork during secondary task, and CameraDistance where the close distance was constrained by keeping a majority of the fish group in sight.

The model that satisfied the selection criteria showed higher cognitive load as environmental factors made it more challenging to identify the fish species albeit with some unexpected trends. For example, participants experienced higher cognitive load with higher turbidity when fish were moving faster. On the other hand, cognitive load dropped with higher turbidity when fish were moving slow. While it is possible that because the fish are moving in three-dimensions, higher turbidity caused individual fish to disappear more often leading to a higher mental effort in recognizing the fish species, this does not explain the drop in cognitive load when fish move slowly. Instead, it is likely that slower fish automatically maintained a tighter spatial group. At high turbidity levels the resulting cohesive group could have been perceived at a stronger contrast to the background making them easier to identify.

The study was limited in terms of the realism of the virtual environment and the range of variables it could test. The environment is already being improved as part of ongoing work where fish now have detailed color patterns on their body, species specific tail movement, and the virtual environment resembles the real counterpart to a higher extent (Bhattacharya & Butail, 2022). A first-person mobile robot perspective within the virtual environment provides a larger and more realistic range of environmental variables to change within a task, enabling cognitively responsive human-robot interaction strategies.

In conclusion, results from this study can be used in enhancing the human-computer or robot interaction experience in classification tasks—an open problem in human-machine interaction (Vinciarelli et al., 2015). The validated measure of cognitive load can complement post-experiment surveys in human computer interaction and aid in identification of instances of high cognitive load in long-drawn mentally intensive tasks. With respect to invasive species classification that is common in citizen science, the measure of cognitive load with commercial-grade EEG headset can provide an interactive interface that adapts task allocation to human cognitive load. Although not performed, the measure can be classified in real-time enable cognitively responsive applications. The dependence of cognitive load on turbidity in the environment and object speed are expected and can form the first steps to differentiating tasks with respect to mental workload. The study also has implications in citizen science and invasive species identification by using the underwater virtual environment to test the usability of different types of interfaces and tasks prior to actual deployment.

5. Acknowledgement

This research was supported by National Science Foundation grant #IIS-2033918.

6. Declaration of Competing Interest

The authors declare they have no financial interests.

References

- Anderson, E. W., Potter, K. C., Matzen, L. E., Shepherd, J. F., Preston, G. A., & Silva, C. T. (2011, June). A user study of visualization effectiveness using eeg and cognitive load. In *Computer Graphics Forum* (Vol. 30).
- Anderson, J. R., Reder, L. M., & Lebiere, C. (1996, June). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, 30(3), 221–256.
- Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review*, 22(4), 425–438.
- Ariji, Y., Yanashita, Y., Kutsuna, S., Muramatsu, C., Fukuda, M., Kise, Y., ... Ariji, E. (2019). Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 128(4), 424–430.
- Bales, G., & Kong, Z. (2022). Neurophysiological and behavioral differences in human-multiagent tasks: An EEG network perspective. *ACM Transactions on Human-Robot Interaction*.
- Bhattacharya, A., & Butail, S. (2022). Designing a virtual reality testbed for direct human-swarm interaction in aquatic species monitoring. In *Proceedings of the Modeling, Estimation and Control Conference*. New Jersey, USA.
- Böcker, K. B. E., van Avermaete, J. A. G., & van den Berg-Lenssen, M. (1994). The international 10–20 system revisited: Cartesian and spherical co-ordinates. *Brain Topography*, 6(3), 231–235.
- Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage*, 5(1), 49–62.
- Brünken, R., Seufert, T., & Paas, F. (2010). Measuring Cognitive Load. In J. L. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive Load Theory* (pp. 181–202). Cambridge: Cambridge University Press.
- Cabañero, L., Hervás, R., González, I., Fontecha, J., Mondéjar, T., & Bravo, J. (2019). Analysis of cognitive load using eeg when interacting with mobile devices. In *Proceedings of the international conference on ubiquitous computing and ambient intelligence*. MDPI.
- Cao, A., Chintamani, K. K., Pandya, A. K., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior Research Methods*, 41(1), 113–117.
- Couzin, I. D., Krause, J., James, R., Ruxton, G. D., & Franks, N. R. (2002). Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology*, 218(1), 1–11.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11(1), 19–23.
- Gómez-Ramírez, J., Freedman, S., Mateos, D., Perez Velazquez, J. L., & Valiente, T. A. (2017). Exploring the alpha desynchronization hypothesis in resting state

- networks with intracranial electroencephalography and wiring cost estimates. *Scientific Reports*, 7(1), 1–11.
- Kiiski, H., Bennett, M., Rueda-Delgado, L. M., Farina, F. R., Knight, R., Boyle, R., . . . others (2020). EEG spectral power, but not theta/beta ratio, is a neuromarker for adult ADHD. *European Journal of Neuroscience*, 51(10), 2095–2109.
- King, A. J., George, A., Buckle, D. J., Novak, P. A., & Fulton, C. J. (2018). Efficacy of remote underwater video cameras for monitoring tropical wetland fishes. *Hydrobiologia*, 807(1), 145–164.
- Klauer, K. C., & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of Experimental Psychology: General*, 133(3), 355–381.
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, 29(2-3), 169–195.
- Langenkämper, D., Simon-Lledó, E., Hosking, B., Jones, D. O., & Nattkemper, T. W. (2019). On the impact of citizen science-derived data quality on deep learning based classification in marine images. *PloS One*, 14(6).
- Laut, J., Henry, E., Nov, O., & Porfiri, M. (2013). Development of a mechatronics-based citizen science platform for aquatic environmental monitoring. *IEEE/ASME Transactions on Mechatronics*, 19(5), 1541–1551.
- Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, 19(1), 577–621.
- Mallet, D., & Pelletier, D. (2014). Underwater video techniques for observing coastal marine biodiversity: a review of sixty years of publications (1952–2012). *Fisheries Research*, 154, 44–62.
- McCullagh, P., & Nelder, J. A. (2019). *Generalized linear models*. Routledge.
- Nissim, N. R., O’Shea, A. M., Bryant, V., Porges, E. C., Cohen, R., & Woods, A. J. (2017). Frontal structural neural correlates of working memory performance in older adults. *Frontiers in Aging Neuroscience*, 8.
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003, March). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38(1), 63–71.
- Palermo, E., Laut, J., Nov, O., Cappa, P., & Porfiri, M. (2017). Spatial memory training in a citizen science context. *Computers in Human Behavior*, 73, 38–46.
- Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, 3(11), 1199–1204.
- Rodin, C. D., de Lima, L. N., de Alcantara Andrade, F. A., Haddad, D. B., Johansen, T. A., & Storvold, R. (2018). Object classification in thermal images using convolutional neural networks for search and rescue missions with unmanned aerial systems. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*.
- Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., & Harvey, E. (2016). Fish species classification in unconstrained underwater environments based on deep learning. *Limnology and Oceanography: Methods*, 14(9), 570–585.
- Schmidt, M., Kanda, P., Basile, L., da Silva Lopes, H. F., Baratho, R., Demario, J., . . . others (2013). Index of alpha/theta ratio of the electroencephalogram: a new marker for alzheimer’s disease. *Frontiers in Aging Neuroscience*, 5, 60.
- Solomon, E. A., Kragel, J. E., Sperling, M. R., Sharan, A., Worrell, G., Kucewicz, M., . . . Jobst, B. C. (2017). Widespread theta synchrony and high-frequency

- desynchronization underlies enhanced cognition. *Nature Communications*, 8(1), 1–14.
- Starr, J., Schweik, C. M., Bush, N., Fletcher, L., Finn, J., Fish, J., & Barger, C. T. (2014). Lights, Camera... Citizen Science: Assessing the Effectiveness of Smartphone-Based Video Training in Invasive Plant Identification. *PLoS One*, 9(11).
- Suzuki-Ohno, Y., Westfechtel, T., Yokoyama, J., Ohno, K., Nakashizuka, T., Kawata, M., & Okatani, T. (2022). Deep learning increases the availability of organism photographs taken by citizens in citizen science programs. *Scientific Reports*, 12(1), 1–10.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
- Tong, S., & Thankor, N. V. (2009). *Quantitative EEG Analysis methods and clinical applications*. Artech House.
- Trammell, J. P., MacRae, P. G., Davis, G., Bergstedt, D., & Anderson, A. E. (2017). The relationship of cognitive performance and the theta-alpha power ratio is age-dependent: an eeg study of short term memory and reasoning during task and resting-state in healthy young and old adults. *Frontiers in Aging Neuroscience*, 9, 364.
- Trouille, L., Lintott, C. J., & Fortson, L. F. (2019). Citizen science frontiers: Efficiency, engagement, and serendipitous discovery with human-machine systems. *Proceedings of the National Academy of Sciences*, 116(6), 1902–1909.
- Tytell, E. D., Borazjani, I., Sotiropoulos, F., Baker, T. V., Anderson, E. J., & Lauder, G. V. (2010). Disentangling the functional roles of morphology and motion in the swimming of fish. *Integrative and Comparative Biology*, 50(6), 1140–1154.
- Urigüen, J. A., & Garcia-Zapirain, B. (2015). EEG artifact removal—state-of-the-art and guidelines. *Journal of Neural Engineering*, 12(3).
- USGS. (2022). *Nonindigenous Aquatic Species*. <https://nas.er.usgs.gov/taxgroup/fish/default.aspx>.
- Vinciarelli, A., Esposito, A., André, E., Bonin, F., Chetouani, M., Cohn, J. F., ... others (2015). Open challenges in modelling, analysis and synthesis of human behaviour in human-human and human-machine interactions. *Cognitive Computation*, 7(4), 397–413.
- Wick, M. J., Angradi, T. R., Pawlowski, M. B., Bolgrien, D., Debbout, R., Launspach, J., & Nord, M. (2020). Deep Lake Explorer: A web application for crowdsourcing the classification of benthic underwater video from the Laurentian Great Lakes. *Journal of Great Lakes Research*, 46(5), 1469–1478.
- Zarjam, P., Epps, J., & Lovell, N. H. (2015). Beyond subjective self-rating: EEG signal classification of cognitive workload. *IEEE Transactions on Autonomous Mental Development*, 7(4), 301–310.
- Zhang, L., Li, S. Z., Yuan, X., & Xiang, S. (2007). Real-time object classification in video surveillance based on appearance learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Zuur, A. F., & Ieno, E. N. (2016). A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution*, 7(6), 636–645.

7. About the Authors

Arunim Bhattacharya (arunimbhattacharya1@gmail.com) is a PhD student in the Department of Mechanical Engineering at Northern Illinois University. He received a Bachelor's degree in Mechanical Engineering from Kalinga Institute of Industrial Technology, India and Master's from Illinois Institute of Technology, Chicago. His research interests include virtual reality and human-swarm interaction.

Sachit Butail (sbutail@niu.edu) is an Associate Professor of Mechanical Engineering at Northern Illinois University. He received his Ph.D. in Aerospace Engineering from University of Maryland, College Park, in 2012. His research interests include dynamical systems and controls, robotics, collective behavior, pattern recognition, and bioinspired autonomy.