Quantifying Human Trust in Controlling Robot Swarms: EEG-based Analysis and Classification

Jesus A. Orozco, Stergios Bachoumas, and Panagiotis Artemiadis*, *IEEE Senior Member*

***Abstract*— Trust is a cornerstone of human-to-human interaction. We build up or tear down relationships depending on our perceived trust levels. However, we tend to be cautious when trusting artificial systems such as robots or artificial intelligence. This article addresses the gap in determining levels of trust in scenarios where humans interact with robot swarms by directly providing command inputs. We use Electroencephalography (EEG) data of human subjects controlling robotics swarms via joystick commands and randomly inject enough noise to make their given task difficult. In doing so, we artificially create distrust in the system and use machine learning techniques to find EEG correlates of swarm trust. The results of this study suggest that, when directly interacting with a robot swarm, these neural correlates of trust in a robot swarm exist and are discernable via machine learning feature classification to an accuracy greater than 88%. This work shows great promise in establishing effective human-machine teaming in exploration, search and rescue operations, and defense. Successfully quantifying human trust levels is an essential building block to facilitate the adoption of robot swarms in real-world environments with human operators, as swarms can leverage these to adapt their behavior to maintain or regain a human's trust level in the team.**

I. Introduction

In an ever-demanding world, humans and robots must work efficiently together; humans also need to trust that the robotic agents will do what we expect them to do [1]. This is especially true now that we are developing fully autonomous systems, and we expect those systems to interact with or operate near humans [2]. Specifically to domains in search and rescue and area exploration, robot swarms are better equipped to deal with dynamic environments. They can be scaled up and are more manageable than a multi-robot system [3]. A swarm is a collection of robot agents enabled with basic control laws that exhibit emergent behavior in response to their collective interaction with the environment, similar to a flock of birds or school of fish [4]. In this environment, we need human operators to trust the robot swarms [5]. We can achieve trust if we provide reliable systems that can adapt seamlessly to varying levels of autonomy when we have a drop or gain in human trust [6].

Existing literature focuses on trust in human-swarm interaction through predictive mathematical models of trust [7–9]. Models of this type need to consider many variables (e.g., dependability, transparency, demographics, status information) that influence trust to predict trust adequately. We have seen implementations of predictive models to control swarm based on a Markov decision process [6], and more recently, a trust-aware control method based on a weighted mean subspace reduced algorithm [10]. While these provide impressive results, the models must encapsulate unforeseen events that might disrupt trust to work as intended.

There is, however, existing literature on human-swarm interaction that focuses on EEG correlates of trust. We recently found neural correlates of trust in human-swarm interaction using Electroencephalographic (EEG) signals [11] consistent with previous work that looks at trust in an investment game [12]. Specifically, we found features in the frontal and occipital regions consistent with previous work in addition to channels in the temporal region [11]. Our work applied machine learning techniques for feature selection and classification, with results over 98% classification accuracy. While the findings are desirable, in that paper, the human subject observes a virtual swarm traversing a computer screen and does not interact with the swarm.

This paper addresses the gap of describing human trust levels in continuously interacting with a robot swarm to complete a task using a joystick to provide velocity commands. Joystick control is implemented in previous works, too [13]. However, in that work, the joystick commands control the leader agents and provide steering commands that should influence the remaining agents through alignment, attraction, and repulsive laws. In this paper, we simplify our approach in simulation by allowing all agents to listen to the joystick command provided by the subject while only being influenced by a repulsion law and a wall-bouncing law we designed from [11]. To create a low-trust environment, we inject noise into the joystick commands, based on the idea that trust is defined as the expectation that an agent will behave as anticipated by the human [1]. Through the human-swarm interaction, EEG signals are recorded and processed through a classifier to distinguish between high- and low-trust cases. Our results with a simulated swarm show an 88% prediction accuracy in determining the interaction as low-trust or high-trust based on the EEG recordings alone. To validate what we saw in the simulation, we use real ground robots and apply a similar task to test on one subject. The results of the validation experiment show that we have a prediction accuracy of 93.96%. The promising results of this work can have a direct impact on real-world human-machine team applications, such as search and rescue operations, space exploration, and defense. These applications can benefit from

*This material is based upon work supported by the National Science Foundation under Grant No. #2014264, as well as by the U.S. Air Force Office of Scientific Research (AFOSR) award FA9550-18-1-0221.

Jesus A. Orozco, Stergios Bachoumas, and Panagiotis Artemiadis are with the Mechanical Engineering Department, at the University of Delaware, Newark, DE 19716, USA. jaon@udel.edu, stevbach@udel.edu, partem@udel.edu

*Corrresponding author: partem@udel.edu

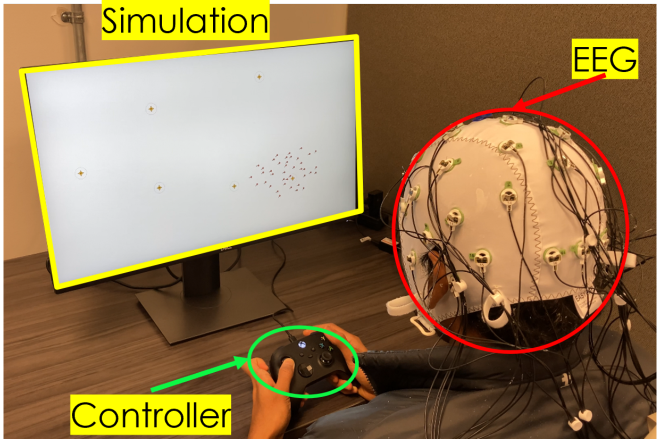

Fig. 1. Experimental setup. A subject is wearing an EEG cap while interacting with the virtual swarm via a video game controller on the computer screen.

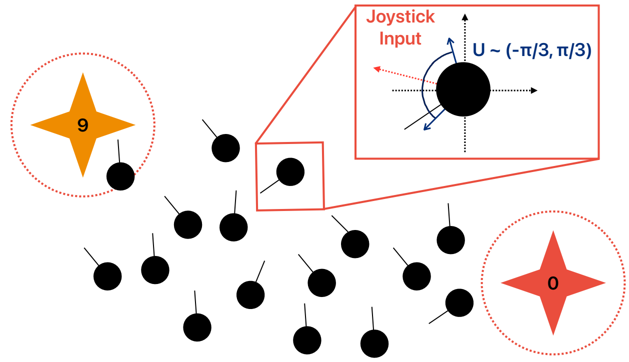

Fig. 2. Representative simulation layout. After an agent crosses the dashed circle barrier, the target number ticks down from 10, as seen on the upper left, until 10 agents have successfully crossed the barrier. Once this happens, the goal turns red, as seen in the bottom right. The zoomed-in portion shows how the uniformly distributed random noise is applied to the given joystick heading in red.

a more direct awareness of human trust levels of the robotic systems.

II. Methods

This paper expands on previous work that found differences in trust at the brain level when observing a virtual swarm [11], by allowing the subject to control the swarm via joystick commands. Control of the swarm enables us to explore neural correlates of trust when actively interacting with a swarm. The following subsections detail the data type and the swarm simulator used in this work. We then explore the EEG processing workflow, including the feature selection and feature classification used to predict trust levels. Finally, we detail our validation experiment using the physical environment with ground robots.

A. Experiment Setup

For this study, we collected EEG data and joystick commands as the human subject controlled the robot swarm. The 32-channel EEG data was collected and streamed via a wireless EEG device (LiveAmp 32, Brain Products GmbH, Gilching, Germany), with electrodes (actiCAP slim active electrodes, Brain Products GmbH, Gilching, Germany) attached to the subject via an EEG head cap (actiCAP snap, Brain Products, GmbH, Gilching, Germany). The electrode cap ensures electrode placement as detailed in the International 10-20 system [14]. The frequency of the data streamed to a computer is 500 Hz. A Microsoft Xbox One Wireless Controller was used as a joystick to control the swarm. The joystick data streamed at 30 Hz. Figure 1 shows the experiment configuration.

B. Task Description

Similar to [11], the experiment included 40 virtual non-holonomic agents displaced as discs with lines displaying their current heading, tasked with reaching six targets on a computer screen. The swarm spawns in two possible areas: on the bottom left or bottom right of the computer screen. The subject follows a path; the randomized spawn location determines the target path. The subject is to guide the swarm towards the targets in order, starting with the closest target. Completing a target requires at least 10 agents to cross the target boundary. Two visual cues let the subject know they were successful: one is a change in color of the yellow marker to red, and the other is a countdown starting from 10 and ticking down every time an agent crosses the boundary. The representation of the simulation is shown in Fig. 2. The second variant is described as a mirrored image of Fig. 2, with the swarm spawning on the bottom left of the screen. The subject has 25 seconds to complete all six targets, at which point the trial ends. A white screen persists for five seconds, and then another trial appears. Two conditions are being tested: high and low noise. The following section will describe in detail the two different conditions. The low and high noise trials are presented to the subject randomly. There are 20 trials per condition, of which 10 trials are left swarm spawn, and the remaining trials are right swarm spawn.

C. Swarm Simulator and Behavior

With its respective implementation, the simulator is mostly unchanged from [11]. A modification to allow joystick input via the XInput-Python 0.4.0 Python package and a boundary and repulsion force are used. The user must provide continuous input via joystick commands to control the swarm. When the user lets go of the joystick, the swarm stops. We obtain the unit vector, $\mathbf{h}$, from the joystick, indicating the x and y positions of the joystick command. From this, we obtain a heading angle, θ, to display a line on the agent showing the current agent heading. We injected noise in a uniform distribution of -60 to 60 degrees relative to the individual agent's current heading to test the low and high trust conditions, set via the noise flag, n_f, as 0 and 1, respectively. The joystick force, $\mathbf{j}_i$, for each agent i, $i = 1, 2, ..., n$, is defined as:

$$\mathbf{j}_i = \begin{cases} \mathbf{h} & \text{, if } n_f = 0 \\ \|\mathbf{h}\| \left(\cos(\alpha)\hat{i} + \sin(\alpha)\hat{j}\right) & \text{, if } n_f = 1 \end{cases} \quad (1)$$

$$\alpha = \theta + \mathrm{U} \tag{2}$$

where n is the swarm size, θ is the joystick heading, and $\mathrm{U} \sim (-\pi/3, \pi/3)$ is the random uniform distribution, from -60 to 60 degrees. Additionally, each agent is experiencing a repulsive force, which is defined as follows:

$$\mathbf{r}_i = \begin{cases} \displaystyle\sum_{j=1}^{n} \frac{\mathbf{p}_i - \mathbf{p}_j}{\|\mathbf{p}_i - \mathbf{p}_j\|^2}, \text{ if } \|\mathbf{p}_i - \mathbf{p}_j\| \leq s_r, \text{ and } i \neq j \\ \mathbf{0}, \text{ if } \|\mathbf{p}_i - \mathbf{p}_j\| > s_r \end{cases} \tag{3}$$

where $\mathbf{p}_i$ is the position of the agent i, $\mathbf{p}_j$ is the position of the j^{th} neighbor, s_r is the sensing radius for the repulsion vector, and $\|\cdot\|$ denotes the vector norm. We employed an inverse square law here to maximize the repulsion force when the neighboring agents are closer to the agent so as not to have all neighbors equally influence the resultant force. We also want to ensure that the agents do not displace past the perimeter of the computer screen. As such, the wall-bounce force, $\mathbf{w}_i$, is defined and implemented as in [11]. The resultant *directed* joystick-based behavior, $\mathbf{D}_i$, is defined as:

$$\mathbf{D}_i = \mathbf{j}_i + \mathbf{r}_i + \mathbf{w}_i. \tag{4}$$

The simulated agents are modeled as double integrators, similar to [15], where the agents are driven by resultant force in (4), which directly affects the two- dimensional acceleration of the agents.

D. Experiment Details

In order to participate in this study, all subjects must be between 21 and 50 years old, have no current or past history of visual impairment, have no known neurological issues, be proficient in English, and be in general good health. There were 11 subjects, 10 males and one female, with an average age of 26.4 ± 2.3 years. Due to a network connection issue experienced mid-experiment, we processed and analyzed the data for the first nine subjects. Additionally, subject 11 did not participate in the simulation experiment and only participated in the validation experiment with real robots. The protocol for this study was approved by the University of Delaware Institutional Review Board (IRB ID: 1487701-6), with informed consent given by all the participants.

Before the start of the experiment, we informed the subjects about what they would see on the screen. We task the subject to use the joystick to control the swarm towards each target, starting with the closest one from the swarm spawn site. We have a practice simulator for the subjects to familiarize themselves with the joystick control. All subjects were comfortable with the controls between 1-2 practice runs. We asked the subjects to limit eye blinks, to limit ocular artifacts, for each trial to the best of their abilities, but they could rest and prepare for the subsequent trial in the five-second white screen break.

E. Data Analysis

The data are processed and analyzed in MATLAB and the EEGLAB plugin [3]. We then generate features to train and test our prediction models. The workflow for data analysis is detailed in Fig. 3.

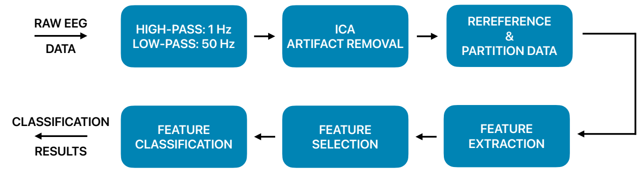

Fig. 3. EEG processing workflow.

1) EEG Data Processing: The complete EEG data, including the zero-stimulus white screen, is processed. We high-pass and low-pass filter the data with cutoff frequencies of 1 and 50 Hz, respectively. After filtering, we use EEGLAB to process the data further through Independent Component Analysis (ICA). We visually inspect and remove associated ocular artifacts. After this, the data will be re-referenced as an average of all the channels and split using the trial start and end markers to label the condition of each trial appropriately.

2) Feature Extraction, Selection, and Classification: First, the data is split into 1-second windows, with 50% overlap. Except for the 10 sub-band power spectrum values and the wavelet's energy and entropy for each waveband (0.5 Hz $< \delta <$ 4.0 Hz, 4 Hz $< \theta <$ 7.5 Hz, 8.0 Hz $< \alpha <$ 13.0 Hz, 14 Hz $< \beta <$ 26 Hz, 30 $< \gamma <$ 45.0 Hz), as described in [16], most features are calculated from the EEG Toolbox MATLAB plugin [17]. The features extracted from the toolbox are: Hjorth Parameters (i.e., Activity, Mobility, and Complexity), Band Power (i.e., Delta, Theta, Alpha, Beta, and Gamma), Normalized First and Second Distances, Power Ratio of Beta/Alpha, Log Root Sum of Sequential Variation (LRSSV), Renyi Entropy, Mean Energy, and Meant Teager Energy. The wavelet energies and entropies are calculated as in [18] for each waveband. The total feature vector per trial window, $\mathbf{X}$, is constructed in the following way:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1^T & \mathbf{X}_2^T & \cdots & \mathbf{X}_{32}^T \end{bmatrix}^T \tag{5}$$

where $\mathbf{X}_\ell \in \mathbb{R}^{35}$, $\ell = 1, 2, ..., 32$, is the feature vector for each of the 32 EEG channels, containing all the 35 features extracted, for a total of 1,120 features per second. To avoid interference from joystick motion, we removed the motor function-related EEG channels 8, 24, and 25 from further analysis, bringing our total number of features down to 1015 features per second.

The feature selection uses the MATLAB function *fscnca* based on the Neighborhood Component Analysis (NCA) learning method, and is implemented previously [11]. The feature classification employs an 80% - 20% training-testing split of approximately 1,920 windows per subject. We use a similar approach to [11] in that we use the k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), and Linear Discriminant Analysis (LDA) classifiers using *fitcknn*, *fitcsvm*, and *fitcdiscr* MATLAB functions, respectively. Additionally, the k-NN and SVM classifiers "OptimizeHyperparameters" flag is set to "auto" and the "HyperparameterOptimizationOptions" flag is set to do a "k-fold" of 5

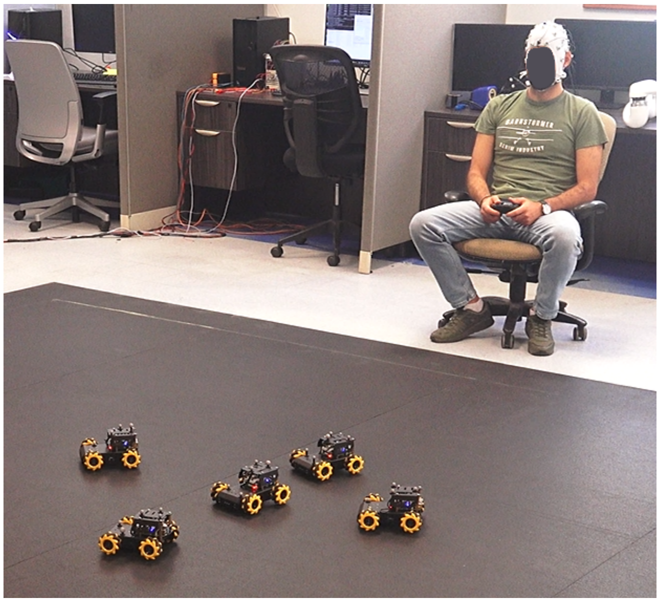

Fig. 4. Real-robot experiment. Five TurboPi robots are shown and they are controlled via joystick commands by the subject connected to an EEG system.

and "MaxObjectiveEvaluations" of 50 for both models. A separate classifier is trained for each subject.

F. Validation in Physical Environment

In addition to our study with the simulated swarm, we performed experiments with real robots in our lab to validate what we saw with a simulated swarm. The robot platform uses ROS2 Humble and Ubuntu 22.04. We switched from XInput to Pygame 2.6.0 for joystick commands due to XInput compatibility issues with Linux. For the swarm robots, we used five Hi Wonder Turbo Pi omnidirectional drive platforms designed for educational purposes, as they are inexpensive and simultaneously capable of handling the computation load of the implemented algorithms. Secondly, the position of each robot inside the arena is tracked using 8 Optitrack PrimeX 13 (OptiTrack, NaturalPoint, Inc.) cameras that provide position information at a 240 Hz frame rate. Finally, the data is synchronized and streamed using Optitack's Motive software to a central computer that handles the ROS2 communication with the robots. See Fig. 4 for a representation of the robot arena.

III. Results

We first show the simulation experiment results of all nine subjects and continue with the real-world validation experiment with one subject.

A. Simulation

We aim to discover if consistent neural correlates of trust exist when directly interacting with a swarm. We establish goal metrics for each subject, as shown in Table. I. TC is defined as the average target completion per trial for each subject. Time/Target is defined as the average time it took each subject to have 10 agents cross a target successfully.

TABLE I
Subject Trial Metrics

	Low Noise		High Noise	
Subject	**TC (%)**	**Time/Target (s)**	**TC (%)**	**Time/Target (s)**
1	93.33	1.54 ± 0.41	51.67	3.54 ± 1.52
2	87.50	1.42 ± 0.59	58.33	3.78 ± 2.16
3	95.00	1.53 ± 0.63	59.17	3.47 ± 1.40
4	66.67	2.20 ± 0.55	45.00	4.18 ± 1.35
5	90.00	1.79 ± 0.85	50.83	4.51 ± 2.56
6	87.50	1.63 ± 0.62	58.33	3.67 ± 1.50
7	98.33	1.42 ± 0.45	64.17	3.62 ± 1.36
8	92.50	1.68 ± 0.76	60.00	3.60 ± 1.74
9	98.33	1.43 ± 0.39	71.67	3.21 ± 1.21
Avg.	**89.91**	**1.61 ± 0.64**	**57.69**	**3.96 ± 1.70**

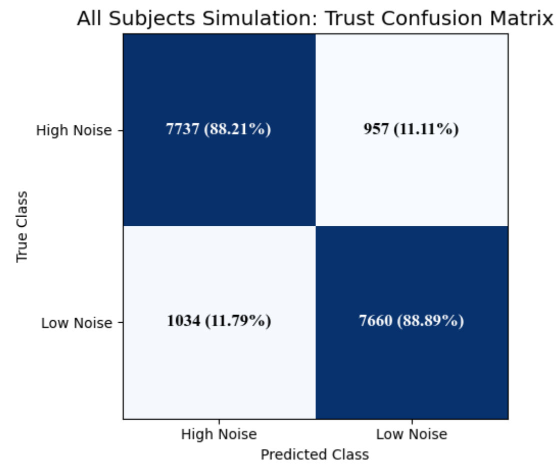

Fig. 5. Combined confusion matrix of all nine subjects using the low-noise and high-noise classes. The model was trained and tested for each subject, and the classification results were added together. Out of approximately 1,932 observations per subject, a total of 17,388 observations, the k-NN model correctly predicts 88.21% and 88.89% of the high-noise and low-noise conditions, respectively.

As it is seen, all subjects control the low-noise swarm better than the high-noise swarm. On average, we see a target completion of 89.91% for a low-noise swarm compared to a 57.69% target completion for a high-noise swarm. Even though Subject 4 appears as an outlier in controlling the low-noise swarm, they still fare better in target completion than when controlling the high-noise swarm.

Our classification results showed that the k-NN classifier outperformed the SVM and LDA classifiers in accuracy, sensitivity, precision, specificity, and F1 score; thus, we only report the k-NN results in this paper, as seen in Table II. Standard evaluation metrics, such as accuracy (ACC), sensitivity (SEN), precision (PRE), specificity (SPE), and F1 Score [19], are calculated, while our positive and negative labels in the confusion matrix (see Fig. 5) are high-noise and low-noise, respectively. Most of the subjects have between 87% and 96% prediction accuracy. The two outliers, subjects 7 and 8, are between 76% and 81% prediction accuracy. Figure 5 shows the confusion matrix of the prediction accuracy, combining the results for all subjects.

We seek to find if the 1105 features we extract per second

TABLE II

k-NN CLASSIFICATION RESULTS

Subject	ACC	SEN	PRE	SPE	F1 Score
1	0.9200	0.9169	0.9236	0.9231	0.9202
2	0.8802	0.8760	0.8860	0.8846	0.8810
3	0.9648	0.9638	0.9658	0.9657	0.9648
4	0.9398	0.9425	0.9366	0.9372	0.9395
5	0.9120	0.9208	0.9018	0.9035	0.9112
6	0.9027	0.8995	0.9069	0.9060	0.9032
7	0.8160	0.8056	0.8330	0.8271	0.8191
8	0.7644	0.7563	0.7805	0.7730	0.7682
9	0.8696	0.8656	0.8746	0.8736	0.8701
11*	0.9396	0.9456	0.9401	0.9330	0.9428

TABLE III

NCA FEATURES SELECTED

Subject	Features Selected	Channels	Feature Reduction (%)
1	24	16	97.64
2	18	14	98.23
3	28	22	97.24
4	10	10	99.01
5	14	13	98.62
6	16	14	98.42
7	213	29	79.01
8	238	29	76.55
9	17	10	98.33
11*	24	20	97.64

are meaningful in our feature classification. For this, we look at the features selected from the fold that provided the best prediction accuracy for each subject, and we note the number of channels needed for the best prediction accuracy. Except for subjects 7 and 8, the selected features are reduced to between 0.90-2.54% of the available features, a drastic reduction that allows for real-time implementation of the method. For subjects 7 and 8, the features are reduced to approximately 20.4% of the available features. The top 5 most prevalent features are the Log Root Sum of Sequential Variation (LRSSV) [20], Mean and Mean Teager Energy (ME, MTE) [21], Renyi Entropy (RE) [22], and Alpha Energy (AE) [18]. An important aspect of this study is determining whether we see similarities with similar trust research, albeit not necessarily on human-swarm interaction. For this, we highlight the most prevalent features topographically, as seen in Fig. 7.

B. Real-World Validation Experiment

We sought to validate the simulation results by conducting a real-world experiment focused on a similar task, albeit with fewer targets and spawning locations. In this validation phase, the subject (Subject 11) could not reach the other end of the arena in the allotted time for all 20 low-trust trials, but they did reach all 20 high-trust trials. We applied the same EEG processing and machine learning techniques from the simulation analysis in the validation data. As seen for Subject 11 in Table II and in Fig. 8, our classification results show a performance greater than 90% across all metrics. Additionally, the features selected and associated channels are in the family of those seen in simulation subjects, as seen in Table III.

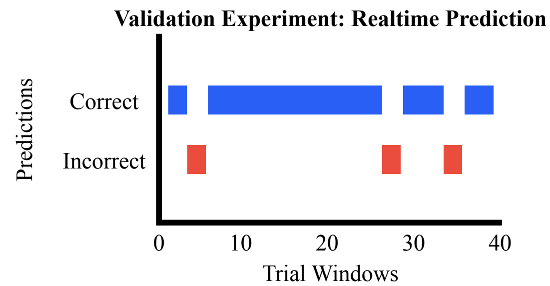

Fig. 6. Realtime prediction of a subject's trial. A classifier is used for successive windows of 1.0s with 0.5 s overlap, for a 20-second trial, leading to consistently accurate results and a prediction accuracy of 92.11%.

Finally, we explored the possibility of what a real-time application may look like. Thus, we trained on all trials but one, and after taking the trained k-NN model, we tested how this model would classify the trust level through a full 20-second trial. The results are shown in Fig. 6, where it is shown that the model consistently predicts the correct level of trust, except for a few instances.

IV. DISCUSSION

The results of this study confirm what we previously observed in a passive human-swarm interaction experiment: EEG correlates of swarm trust do exist when humans interact and control a robot swarm and are distinguishable via machine learning techniques, given the current command interface in simulation. This study provides more substantial evidence for its potential to impact future human-swarm teaming. The rest of the discussion focuses on the analysis of specific brain regions involved, shortcomings, and future applications of the findings in this study.

A. Brain Regions Related to Trust

Based on this study's simulation results, we confirm that the brain areas activated differently between low-noise and high-noise conditions, as depicted in Fig. 7, are consistent with brain regions responsible for trust identified in previous studies [11, 12]. The consistency is in the left frontal, right frontal, and occipital regions. We do see that subjects 7 and 8 show activity across all brain regions. We note that Table II shows the same subjects as outliers in classification prediction accuracy; however, Table I provides little insight into the differences observed as the metrics are consistent with the rest of the subjects.

B. EEG Features and Dimensionality Reduction

We noted in Table III that the NCA feature selection method used in [11] resulted in most subjects having a reduced feature space of at least 97%. Including all subjects, we have, on average, 64.22 ± 91.80 features selected for approximately 88% prediction accuracy. As previously discussed in section III.A, the top five features are: LRSSV, ME, MTE, RE, and AE. Based on the literature, features LSRSSV, ME, and MTE aid in classification of sleep stages

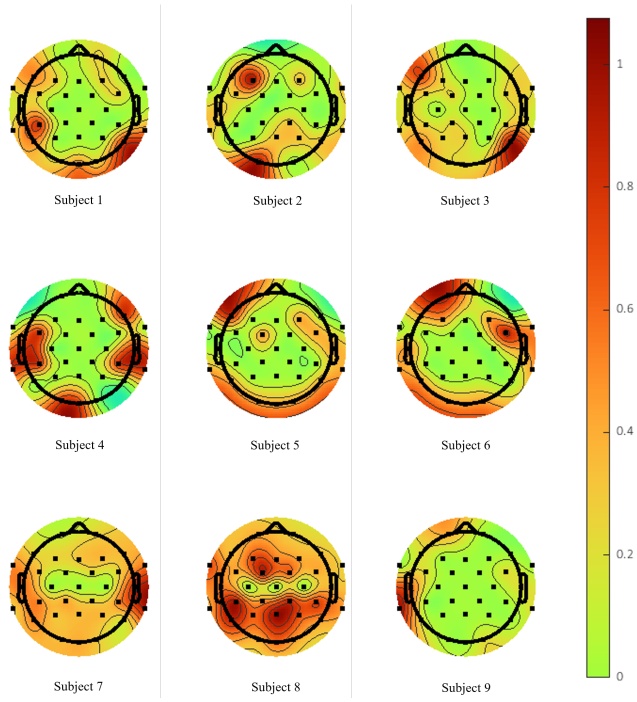

Fig. 7. Brain topographies of all nine subjects. The values are normalized per subject to the largest number of features used across all channels, where 1 represents the highest number of features used.

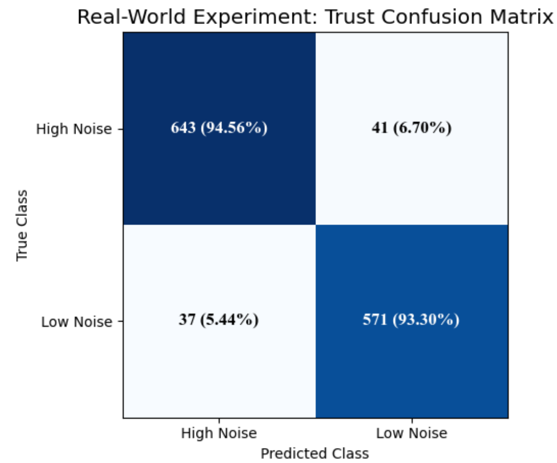

Fig. 8. Confusion matrix of Subject 11 that completed the validation experiment. Out of 1292 observations, the k-NN model correctly predicts 94.56% and 93.30% of the high-noise and low-noise conditions, respectively.

[20, 21], and RE for determining cry signals in infants [22]. Additionally, wavelet energy and entropy aid in emotion recognition [18]. For LRSSV, we see that the variations in the EEG experienced during the trials are distinguishable enough to provide a consistent neural correlate of trust across all subjects, with this feature being the highest ranked feature across all subjects. As far as we know, LRSSV, ME, MTE, and RE have not been used to study trust; however, features in the alpha wave have been used in trust recognition before [12, 23, 24].

C. Limitations and Future Applications

An EEG interface similar to the one used in this study might be considered too cumbersome to adapt to in real-world scenarios. However, technological advances in this domain will allow for a streamlined and reliable implementation in the near future. Moreover, this study shows that we can focus on recording only from specific regions, therefore allowing for a simpler setup.

This proposed work builds off our previous study, which found, with greater than 95% prediction accuracy, neural correlates of trust when observing different swarming behaviors. The current study focuses on another aspect of human-swarm interaction with the subject: providing commands to a swarm. Since this further expands on previous work, we focused on making the interface as simple as possible with direct continuous commands to the swarm. As such, future work could focus on providing additional control laws and intermittent joystick commands to free test subjects from continuous monitoring and commanding. With intermittent commands, we could observe the direct event that leads to a drop or a regaining of trust through analysis of EEG event potentials.

V. Conclusion

Given previously established neural correlates of trust in passive human-swarm interaction [11], we sought to determine if these neural correlates of trust remain when changing the swarm to listen to joystick commands. In this study, we developed the trust paradigm by focusing on changing the expected behavior of a swarm by injecting noise into the joystick command to elicit untrustworthiness. The results of this article show that we can identify low-trust and high-trust scenarios with a greater than 88% prediction accuracy by only using approximately 7% of the available features. This result is another building block for effective human-machine teaming with applications to many fields, such as search and rescue operations, space exploration, and defense. These applications would benefit from quantifying levels of human trust in robot swarms. This quantification can bridge the gap of safely deploying swarms in the real world, knowing they can adapt to a human operator's varying levels of trust.

REFERENCES

[1] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 46, no. 1, pp. 50–80, 2004.

[2] C. Weidemann, N. Mandischer, F. van Kerkom, B. Corves, M. Hüsing, T. Kraus, and C. Garus, "Literature review on recent trends and perspectives of collaborative robotics in work 4.0," *Robotics*, vol. 12, no. 3, 2023. [Online]. Available: https://www.mdpi.com/2218-6581/12/3/84

[3] R. D. Arnold, H. Yamaguchi, and T. Tanaka, "Search and rescue with autonomous flying robots through behavior-based cooperative intelligence," *Journal of International Humanitarian Action*, vol. 3, no. 1, pp. 1–18, 2018.

[4] C. W. Reynolds, "Flocks, herds and schools: A distributed behavioral model," *ACM SIGGRAPH Computer Graphics*, vol. 21, no. 4, pp. 25–34, 1987.

[5] J. Wilson, G. Chance, P. Winter, S. Lee, E. Milner, D. Abeywickrama, S. Windsor, J. Downer, K. Eder, J. Ives, *et al.*, "Trustworthy swarms," in *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*, 2023, pp. 1–11.

[6] C. Nam, P. Walker, H. Li, M. Lewis, and K. Sycara, "Models of trust in human control of swarms with varied levels of autonomy," *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 3, pp. 194–204, 2020.

[7] F. Ekman, M. Johansson, and J. Sochor, "Creating appropriate trust in automated vehicle systems: A framework for hmi design," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 1, pp. 95–101, 2018.

[8] S. Choo and C. S. Nam, "Detecting human trust calibration in automation: A convolutional neural network approach," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 4, pp. 774–783, 2022.

[9] W.-L. Hu, K. Akash, T. Reid, and N. Jain, "Computational modeling of the dynamics of human trust during human–machine interactions," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 6, pp. 485–497, 2019.

[10] Y. Guo, Y. Pang, J. Lyons, M. Lewis, K. Sycara, and R. Liu, "Trust-aware reflective control for fault-resilient dynamic task response in human–swarm cooperation," *AI*, vol. 5, no. 1, pp. 446–464, 2024.

[11] J. A. Orozco and P. Artemiadis, "Extracting human levels of trust in human–swarm interaction using EEG signals," *IEEE Transactions on Human-Machine Systems*, vol. 54, no. 2, pp. 182–191, 2024.

[12] M. Wang, A. Hussein, R. F. Rojas, K. Shafi, and H. A. Abbass, "EEG-based neural correlates of trust in human-autonomy interaction," in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2018, Conference Proceedings.

[13] P. Walker, S. Amirpour Amraii, M. Lewis, N. Chakraborty, and K. Sycara, "Control of swarms with multiple leader agents," in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2014, Conference Proceedings.

[14] G. H. Klem, "The ten-twenty electrode system of the international federation. the international federation of clinical neurophysiology," *Electroencephalogr. Clin. Neurophysiol. Suppl.*, vol. 52, pp. 3–6, 1999.

[15] A. Singh and P. Artemiadis, "Automatic identification of the leader in a swarm using an optimized clustering and probabilistic approach," in *2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2021, pp. 1–6.

[16] S. Sanei and J. A. Chambers, *EEG signal processing and machine learning*. John Wiley & Sons, 2021.

[17] J. Too, "EEG feature extraction toolbox," https://www.mathworks.com/matlabcentral/fileexchange/84235-eeg-feature-extraction-toolbox, 2020, [Online; accessed November 1, 2023].

[18] H. Candra, M. Yuwono, A. Handojoseno, R. Chai, S. Su, and H. T. Nguyen, "Recognizing emotions from EEG subbands using wavelet analysis," in *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2015, pp. 6030–6033.

[19] H. Dalianis, *Evaluation Metrics and Evaluation*. Cham: Springer International Publishing, 2018, pp. 45–53.

[20] P. Memar and F. Faradji, "A novel multi-class EEG-based sleep stage classification system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 1, pp. 84–95, 2017.

[21] B. Şen, M. Peker, A. Çavuşoğlu, and F. V. Çelebi, "A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms," *Journal of medical systems*, vol. 38, pp. 1–21, 2014.

[22] M. Hariharan, R. Sindhu, V. Vijean, H. Yazid, T. Nadarajaw, S. Yaacob, and K. Polat, "Improved binary dragonfly optimization algorithm and wavelet packet based non-linear features for infant cry classification," *Computer Methods and Programs in Biomedicine*, vol. 155, pp. 39–51, 2018.

[23] C. Blais, D. M. Ellis, K. M. Wingert, A. B. Cohen, and G. A. Brewer, "Alpha suppression over parietal electrode sites predicts decisions to trust," *Social neuroscience*, vol. 14, no. 2, pp. 226–235, 2019.

[24] S. Oh, Y. Seong, S. Yi, and S. Park, "Investigation of human trust by identifying stimulated brain regions using electroencephalogram," *ICT Express*, vol. 8, no. 3, pp. 363–370, 2022.