World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:17, No:5, 2023

Electroencephalography-Based Intention Recognition and Consensus Assessment during Emergency Response

Siyao Zhu, Yifang Xu

***Abstract*—**After natural and man-made disasters, robots can bypass the danger, expedite the search, and acquire unprecedented situational awareness to design rescue plans. Brain-computer interface is a promising option to overcome the limitations of tedious manual control and operation of robots in the urgent search-and-rescue tasks. This study aims to test the feasibility of using electroencephalography (EEG) signals to decode human intentions and detect the level of consensus on robot-provided information. EEG signals were classified using machine-learning and deep-learning methods to discriminate search intentions and agreement perceptions. The results show that the average classification accuracy for intention recognition and consensus assessment is 67% and 72%, respectively, proving the potential of incorporating recognizable users' bioelectrical responses into advanced robot-assisted systems for emergency response.

***Keywords*—**Consensus assessment, electroencephalogram, EEG, emergency response, human-robot collaboration, intention recognition, search and rescue.

I. Introduction

ROBOTS are being developed and deployed to assist first responders in emergency response. For instance, snake robots can help rescue workers to search for victims trapped in collapsed structures [1], and unmanned aerial and ground vehicles can help law enforcement officers to look for suspects hidden around dark corners [2]. These robots can bypass the danger and expedite the search, and acquire unprecedented situational awareness for first responders to make decisions. However, robots will not replace humans during emergency response, because the tacit sensibilities, procedural knowledge, and inherent abilities to deal with unexpected situations retained by first responders remain difficult to quantify and program. The collaboration between robots and human partners has a great potential to improve team performance during emergency response.

However, it remains a challenge to communicate user intentions and perceptions to robots in a natural and convenient manner, which prevents fluent human-robot collaboration. The hands-free requirement from the first responders excludes the use of tedious manual control and operation. In unknown, unstructured, and obstructed environments, natural-language-based supervision is not amenable for first responders to formulate, and is difficult for robots to understand. Under certain circumstances such as searching for suspects, the use of language is prohibited. Hence, a new method is needed to naturally and rapidly capture human intentions and perceptions, and communicate them to robots for cognitive teaming. The emerging brain-computer interface (BCI) can be a promising option [3]. It has been demonstrated that EEG signals can be processed to decode human mental states [4]. Using BCI to capture and communicate human intentions and perceptions to robots will enable human-robot cognitive teaming. The robots can understand the situation-specific tactics that come with human intuition and expertise. For instance, by recognizing the search intentions of first responders, the robot can navigate to the region of interest, exploit its sensing capability to conduct the search, and communicate the acquired information back to first responders. In addition, the robot can sense the human consensus on the provided information and adjust its next step of work to facilitate trust formation and provide interaction accordingly.

Motivated by the above, the objective of this research is to test the feasibility of using EEG signals to recognize human intentions of using robots to search an area and assess human consensus on the robot-acquired information during emergency response. Two sets of experiments were conducted to collect EEG signals from multiple participants: one for intention recognition, and one for consensus assessment. Machine-learning and deep-learning methods were employed to train the classifiers. The contribution of this study is twofold. First, it delineates an application of BCI and provides an alternative to communicate user intentions and perceptions to robots during emergency response. The experiments and results provide preliminary insights on applying BCI in human-robot collaboration during emergency response. Second, a variety of models were trained and compared in this study to identify the best-performance model, which provides useful insights on selecting appropriate EEG data analysis methods.

II. Literature Review

A. EEG Overview

EEG is a non-invasive method. It uses multiple electrodes placed on the scalp to retrieve the electric signals over a period for electrodiagnosis of brain abnormalities and monitoring nervous systems [5]. Studies have been conducted to

S. Zhu is with the Department of Civil and Systems Engineering, Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: szhu34@jhu.edu).

Y. Xu is with the Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN 37996 USA (corresponding author, phone: 412-552-4447; e-mail: yxu79@vols.utk.edu).

demonstrate that cerebral signals can be observed and analyzed by the signals detected in different scalp regions. Research has concluded that different brain regions take charge of different functions. For example, the frontal lobes play important roles in cognitive functions such as attention and executive function [6]. The medial prefrontal cortex takes part in the arousal of memory and influences human decision making which is related to the user's experience [7]. The occipital lobe takes charge of visual processing [8], and corresponding signals will be detected when a person is exposed to a visual stimulus. The temporal lobe takes charge of long-term memory. The left temporal cortex is responsible for language recognition and processing [9]. The parietal lobe is an area processing various input sensory information including external sources and internal sources transmitted by muscles or other organs. This brain area is in charge of integrating perceived information and adjusting human behavior in certain circumstances or tasks [10]. Several areas of the parietal lobe are also involved in functions of language comprehension [11]. Some researchers focus on the single brain function by analyzing signals from specific brain regions. For example, when recognizing emotions and stress using EEG, researchers tend to analyze the difference in activation between the two cortical hemispheres because left frontal inactivation is proved to be relative to negative emotions and approach-related behaviors, while right frontal inactivation may result from positive emotion and withdraw behaviors [12].

The frequency of brain waves is also proved significant for decoding cortical electrical activity. Typically, the neural rhythms can be identified by dominant signal frequency. In general, the main frequencies of human brain waves are divided into delta (1 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 12 Hz), beta (12 – 25 Hz), and gamma band (above 25 Hz) [13]. The dominance of certain frequency bands in EEG signals reflects brain activities. For example, alpha waves are related to human sensory and memory and thus reflect mental activities. Alpha suppression indicates engagement and attention to perceived information and stimuli. Conversely, increased alpha power is an indication of a higher sense of relaxation and disengagement [14]. Beta wave is generated in posterior and frontal cerebral regions. Higher beta wave power is demonstrated to be correlated with focused attention and anxious thinking. Beta waves over the motor cortex correlate with body movement and muscle contractions. Typical research of beta waves mainly focused on motor control, reward feedback, and attention detection [15]-[17]. Gamma band has been proved to be involved in the human cognitive process. The increment of the gamma band reflects a higher level of attention to stimuli [18]. EEG can provide various information regarding cerebral activities triggered by external stimuli, such as emotion, cognitive load, thinking process, attention, and perception. EEG wearable devices are widely used in the recognition of emotional arousal and cognitive load measurement [19]. Existing studies mainly focus on different brain responses separately, such as analyzing emotional arousal and stress levels independently based on signals from specific brain areas which are dominant for the concerned mental state.

B. EEG-Based Intention Recognition

EEG has been proved to be efficient in intention understanding. Massive researches have been focused on the prediction of human intention for different movements. Lew et al. [20] demonstrated that readiness potential can detect self-paced reaching movement intention 500 ms before the actual onset. Wang and Makeig [21] found that event-related potentials (ERP) in the posterior parietal cortex (PPC) can be used to predict intended movement direction with a saccade-or-reach task experiment. Movement-related EEG potentials (MRPs) and event-related desynchronization (ERD) of the alpha wave were studied by Babiloni et al. [22] and were proved to be correlated with the preparation and execution of finger movements. Kim et al. [23] used ERP to detect driver emergency braking intention in different simulated driving environments. Ge et al. [24] used EEG-fNIRS bimodal system to understand the temporal-spatial features when participants do different actions. However, these researches mainly focus on predicting human intention prior to a movement. In our study, we propose to investigate the intention and identity of certain choices caused by the consideration of the current situation. Because it has been proved that the cortical activity of intended movement plays an important role in intention decoding, we expected to find the neural correlates of human intention when the intention is entire brain activity and cannot be affected by the tendency of making certain movements. Therefore, we minimized the need for limb movement when designing the experiment interface.

Oh [25] used a Support Vector Machine (SVM) model to classify EEG signals. The mission of the research is to classify the subject's intention of 'agreement' and 'disagreement' while reading the contents block. The features used in this research were frequency band powers of the brain signals from thirty electrodes. The results indicated that the classification performance was around 65%. Kang et al. [26] used an experiment that asked the subjects to look for a specific object in a series of images. The features used in this research are EEG data from 32 channels. Four frequency bands are decomposed by time-series signals to extract the EEG features. The researchers concluded that they acquired transitions of users' intentions when target stimuli appear and classified the transitions using the SVM model. They selected the five most significant pairs to test the classification model, and the classification accuracy was 63.6%. Wang et al. [27] performed an oddball task that asks participants to respond quickly and accurately when they noticed target stimuli occurred during observing a sequence of nontarget stimuli. The main feature of this research was the visual P2 of the ERP. This research concludes that compared with beautiful objects, the less beautiful objects cause negative emotional arousal, and generate a higher amplitude of P2. Zhang et al. [28] studied human purchase intention by setting up an experiment to ask the participants to browse luxury handbags with different brand logos. A five-point scale was used to illustrate customers' purchase intention. Features used in this research are three ERP components, i.e., N200, N400, and late positive potential (LPP). They conclude that consumers' purchase motivation for luxury

handbags is related to the satisfaction of social goals.

C. EEG-Based Perception Assessment

Researches have proved that EEG-measured error-related potentials (ErrPs) can be used in robotic control. ErrP signal can be generated in the brain if human users recognize an error that is generated either by the users themselves or someone else such as their intelligent machine partner [29]. Several BCIs have been established based on this theory. ErrP signals can be detected by the robot system, and the robot corrects its behavior to achieve EEG-based robotic control from human perspective [30]. Blais et al. [31] demonstrated that oscillatory activity in the alpha band can reflect the trust decisions by conducting a coin toss game with a trustworthy and untrustworthy person. Oh et al. [32] proved that beta waves are more powerful in situations of trust than mistrust using an experiment of word elicitation. Trust is related to low beta and mid beta waves, while mistrust is related to high beta waves. Researches also find that there is a relationship between neural signals and trust when analyzing human-machine interaction. Wang et al. [33] illustrated that the frontal and occipital areas of the brain are correlated with trust. Akash et al. [34] developed a classifier model for sensing human trust during HCI using EEG signals. Participants in their experiment were told to be a driver and an obstacle detection sensor would alert the participants of obstacles. Participants were asked to choose to trust or not trust the report based on previous experience with system feedback. In their experiment, the only stimulus was obstacle detected and the participants could not refer to the real driving scenario but just make decisions based on experience.

III. Method

A. Method Overview

Fig. 1 presents an overview of the EEG data analysis methods. After collecting EEG raw signals, the first step was signal preprocessing. A high-pass filter was applied to remove the signals with a frequency less than 1 Hz. Next, two types of algorithms were used for classification, the traditional machine learning algorithms, and deep learning algorithms. For machine learning algorithms, the data were further pre-processed via independent component analysis (ICA) using EEGLAB to remove artifacts [35]. The duration of each decision was divided into consecutive 1-second epochs with 0.5-second of overlapping. After signal preprocessing, the next part was feature extraction. Besides time-domain features and frequency-domain features extracted by Fast Fourier transform, discrete wavelet transform was applied to acquire features from decomposed frequency bands. After extracting all the features, the highly correlated features were removed, and the optimal feature combination and feature numbers were determined by feature recursive elimination and cross-validated selection method. Thereafter, five different machine learning classifiers were used to train the model. For deep learning algorithms, we trained the filtered data using different deep learning algorithms. Finally, a best-performance model for individual decision classification was selected based on the cross-validation score of each classifier.

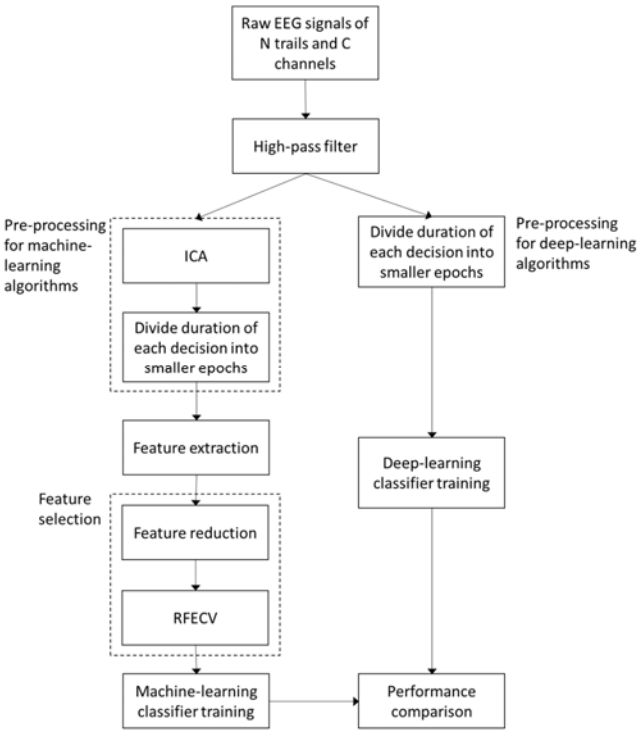

Fig. 1 Method Overview

B. Data Analysis

1) Feature Extraction

For EEG signals, two time-domain features and two frequency-domain features were extracted from each of the channels. The mean and variance of each trial were extracted as time-domain features, and the mean frequency and signal power of each trial were extracted as frequency-domain features. The mean frequency was defined as the mean frequency of the power spectrum. Therefore, 28 time-domain features (14 channels * 2 features) and 28 frequency-domain features were extracted for each epoch. The features are defined as follows.

(1) Mean value

$$\mu = \frac{1}{N}\sum_{i=0}^{N} x_i \tag{1}$$

(2) Variance

$$\sigma = \frac{1}{N-1}\sum_{i=0}^{N} |x_i - \mu|^2 \tag{2}$$

(3) Signal power

$$E = \sum_{i=0}^{N} |x_i|^2 \tag{3}$$

Besides the traditional time-domain and frequency-domain features, the spectral power features were extracted by applying Discrete Wavelet Transform (DWT) [36] for all 14 channels. Since the neural signal is a non-stationary signal, which means the statistics of the signals vary over time, DWT can provide the characteristics of both time and frequency information

simultaneously rather than just provide frequency information by Fourier Transform used in extracting frequency domain features. In this paper, a 5-level DWT decomposition filter was applied to the EEG signals of all channels to acquire frequency bands for each timestamp. The corresponding frequency bands and their band names for each DWT level are shown in Table I. Mean, variance, and signal power were calculated from each decomposed frequency band for each channel. In total, 210 features (5 frequency bands * 14 channels *3 features) were extracted.

TABLE I
FREQUENCY BANDS AND CLASSICAL BAND NAMES FOR DWT LEVELS

DWT Level	Frequency Range	Band Name
1	25-45 Hz	Gamma
2	16-25 Hz	High Beta
3	12-16 Hz	Low Beta
4	8-12 Hz	Alpha
5	4-8 Hz	Theta

2) Feature Selection

The feature set consisted of 266 features (28+28+210) extracted from each trial of each participant. The features were the potential variables for predicting the binary classes (e.g., search or not search, agree or disagree). Since the feature set was much large with limited samples, dimension reduction was applied to select a subset of the features to remove highly correlated features and nonrelevant features to improve accuracy for prediction. Principle Component Analysis (PCA) is a popular method for dimension reduction of EEG feature selection. However, after PCA processing, features are transformed into new components without the knowledge of what features are eliminated. In this paper, the correlation was computed between each of the two features and removed features that have a correlation value exceeding 0.8. The first procedure removed more than two-thirds of the total features that were highly correlated with reserved features. Secondly, feature elimination and cross-validated selection (RFECV) were used to find the best number of features. The Recursive Feature Elimination (RFE) method was applied as the algorithm to eliminate the least important feature in each iteration. The score was tested in each iteration based on the accuracy while performing Cross-Validation (CV), and the importance of each feature was obtained by Random Forest (RF) classification. The CV number was set to be four. The method was iterated with the least important feature being dropped in each iteration. The whole process was iterated 50 times to find the optimal feature combination with the best number of features.

3) Classifier Training

• Machine-Learning Methods

For classifier training, the generalization and prediction performance of each classifier was estimated by CV test and the accuracy of classifiers was demonstrated. CV checked the generalization performance when a classifier was applied to a new data set. To reduce variability and ensure the classifier was applied to general use, five-fold CV was performed in the dataset. The dataset was divided into five partitions, while each time one subset was performed as a testing set, and the other four subsets were performed as the training set. The fitness in prediction was measured by computing the mean score of the five CVs. The five-fold CV ensured a more unbiased result of model prediction performance. The classification algorithms considered in the paper are Logistic Regression, Naïve Bayes, Linear Support Vector Machine (SVM Linear), Linear Discriminant Analysis (LDA), and k-Nearest Neighbors (KNN) [37] classifiers, which were widely used in EEG signal classification [37]-[40].

• Deep-Learning Methods

The user intention and perception reflected by complex EEG signals can be decoded by different strategies and algorithms. The decoding process is transformed into a classification task by various decoding methods. Besides traditional machine learning methods described above, the deep learning network is also widely used in user intention and perception detection. The difference between a traditional machine learning approach and a deep learning approach is that the deep learning approach does not require manual work and human intervention to deal with feature processing. Convolutional Neural Network (CNN) can be applied to classify a series of spectrograms and find distinctive frequency-domain features automatically [41]. In recent years deep learning has shown promising results in decoding EEG signals due to the ability to understand raw and complex signals with massive data. The review made by Roy et al. demonstrated that about 40% of the studies used CNNs and 14% of the studies used Recurrent Neural Networks (RNNs). Most of the studies used raw time-series data to train the deep-learning data, and the accuracy was proved higher than traditional baselines in the studies [42].

In the paper, we applied different neural networks to the EEG data for decision-making classification in search-and-rescue situations. CNN, Depthwise Separable Convolution, Deep Convolutional Neural Networks (Deep ConvNet), Shallow Convolutional Neural Networks (Shallow ConvNet), and RNN were tested in the paper. Class weight was used to handle the imbalance data which is significantly varied among individuals. To reduce overfitting, the dropout rate was set to be 0.5. The data for each participant were divided into three datasets: training set, test set, and validation set. The models were fit and improved based on the training and validation set and evaluated by the test set. A high-pass FIR filter with 1 Hz lower cut-off frequency was applied to raw signals to remove eye artifacts. The same time window technology was applied for the data pre-processing of deep learning classifiers.

• ConvNet

The CNN model was developed by Gehr [43]. The multiple-channel time-series data was cut into short time frames. Fast Fourier Transform (FFT) was applied to transform the signals for each frame into three frequency bands, which are theta, alpha, and beta waves to generate a series of spectral topography maps of each channel, and then projected onto a 2D scalp map. The 2D images are inputs to ConvNet for feature learning and aggregation. For the network, the architecture

consists of two convolutional-max-pooling blocks and a softmax classification layer. The loss function was categorical cross-entropy and the optimizer was RMSProp.

- Depthwise Separable Convolution network

EEGNet was used as the Depthwise Separable Convolution network developed by Lawhernet al [44]. The architecture consists of a temporal convolution, a depthwise convolution, a separable convolution, and a pointwise convolution. The temporal convolution was applied to acquire the frequency map, and a depthwise convolution was connected to learn spatial filters based on the frequency filters. A separable convolution learned the temporal filters for each feature map. At the end of the architecture, a softmax classification layer was applied. The loss function was categorical cross-entropy and the optimizer was Adam.

- Deep ConvNet and Shallow ConvNet

A Deep ConvNet and a Shallow ConvNet developed by Schirrmeister et al. [45] were applied to the EEG data. The architecture of Deep ConvNet was four convolution-max-pooling blocks followed by three max-pooling blocks and a dense softmax classification layer. The first convolutional block contains a temporal convolution layer and a spatial convolution layer based on the electrodes. The Shallow ConvNet has a larger kernel size with shallow architecture compared with Deep ConvNet. Like the deep ConvNet, the architecture of the first two layers is the temporal convolution and a spatial convolution. A mean pooling layer and a dense softmax classification layer were followed by the two convolution layers. The loss function was categorical cross-entropy and the optimizer was Adam.

- RNN

A long short-term memory (LTSM) network was used. Two LSTM layers followed by a dense sigmoid classification layer were created. The loss function was binary cross-entropy and the optimizer was RMSProp.

IV. Experiments

Two experiments were designed to reflect the situations during emergency response. In the first experiment, we proposed to detect human intention about whether to search a rubble pile or not. In the second experiment, we simulated a robot partner to provide information to participants, and then assess human perception (i.e. consensus) on the information provided by the robot. The neural response of participants was recorded by Emotiv Epoc+ wearable EEG device with 14 channels (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4) and two reference sensors (P3 and P4). During the experiment, participants were asked to sit still throughout the experiment and keep silent to avoid noise impacts.

Example images shown to the participants in the two experiments are shown in Fig. 2. Experiment #1 was designed to collect EEG signals for recognizing search intentions. Images of different collapsed structures were presented to the participants. Among the images, 30 were taken from real disaster areas, and 30 were synthesized from three-dimensional (3D) models. The participants were instructed to decide whether they were willing to use a robot to search the disaster areas. The participants' answers, "Yes, proceed to search" and "No, leave to find another place", were recorded and associated with the EEG data. This experiment provided 60 decisions for the data from each participant. The decisions were then proceeded using the time window technology mentioned in the "Data Acquisition and Preprocessing" section above to enlarge the dataset.

Experiment #2 was designed to assess participant perception when an assistive robot provided information to the participants. In this experiment, there were 80 pictures of rubble piles captured in a simulated 3D disaster site shown to the participants. The participants were told that there is an assistive robot accompanying them, and the robot reported to them that there was a victim detected under the shown rubble piles. To detect the agreement perception, the participants were asked to decide if they agreed with the information provided by the robot. The design of experiment #2 was similar to experiment #1. However, to get enough sample size to build an individual model, each participant was asked to conduct 80 trials in total. This resulted in 80 decisions from each participant and the same time window technology was applied to enlarge the dataset. To avoid random selection caused by fatigue, the trials were divided into two groups with 40 trials each. Participants were asked to do the first group in the morning, and the second group in the afternoon for the same day. All the participants were asked to sit still and avoid head and body movements to reduce noise. To make the classification algorithm clearer, all the pictures shown to participants were rubbles with voids detected by robots.

V. Results

In this research, five machine learning classifiers and five deep learning classifiers were tested using two types of brain-related human behavior data. Because the EEG signals vary among individuals, the intra-subjective classifier was adopted for higher model performance. The result of experiment #1 illustrates the best-performance classifier when determining the search intention of each participant. The result of experiment #2 demonstrates the best-performance classifier when determining the agreement perception of each participant.

Fig. 3 shows some results of experiment #1. Fig. 3 (a) shows the distributions of the accuracy of each machine learning classification method and Fig. 3 (b) shows the accuracy distribution of each deep learning classification method for each participant (P1 to P7). The range of each bar in the figure indicates the maximum and minimum accuracy calculated by mean accuracy ± std. The two figures demonstrate that on average, the deep learning methods provide better accuracy than machine learning methods for all participants. Figs. 3 (c) and (d) show the ROC curve of the best-performance classifiers of machine learning and deep learning, respectively. Most of the curves are above the threshold except for some parts of the curves of both classifiers for participant 5. The detailed results of Fig. 3 are shown in Table II. Comparing the classifiers

separately, the best-performance classifier among machine learning methods was shown to be the SVM Linear classifier. There were five over seven participants achieved the best accuracy using the SVM linear classification method and the highest accuracy, 69.9 ± 8.4, was also achieved by the SVM linear classifier from participant 5. For deep learning methods, EEGNet stands out as the best-performance classifier. Five over seven participants achieved their best accuracy using the EEGNet classifier. Participant 4 and participant 7 reached a more than 70 mean accuracy by this classifier and the best accuracy among the seven participants was also achieved by using the EEGNet classifier, which is 73.3 ± 6.9 from participant 4. By comparing the detailed results between machine learning and deep learning classifiers for all the participants, it is concluded that for this experiment purpose, deep learning classifiers provide more accurate predictions.

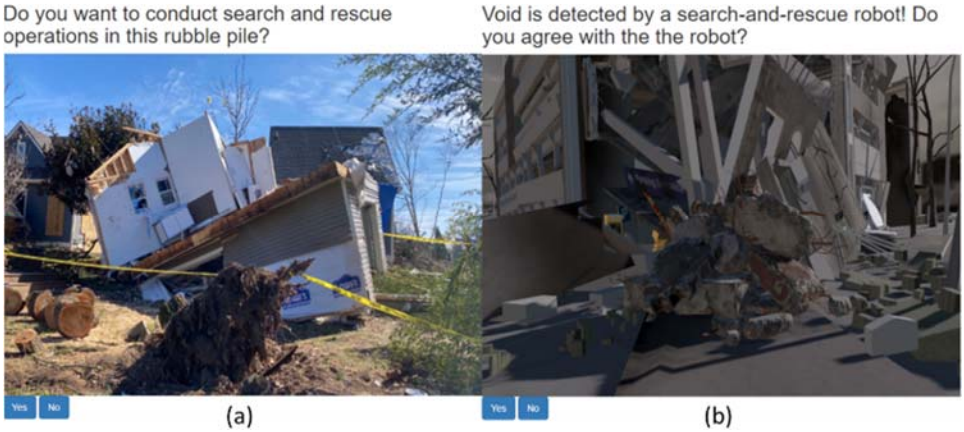

Fig. 2 Example Images Presented to Participants: (a) An example of a real image and questions of experiment #1; (b) An example of synthetic image and questions of experiment #2

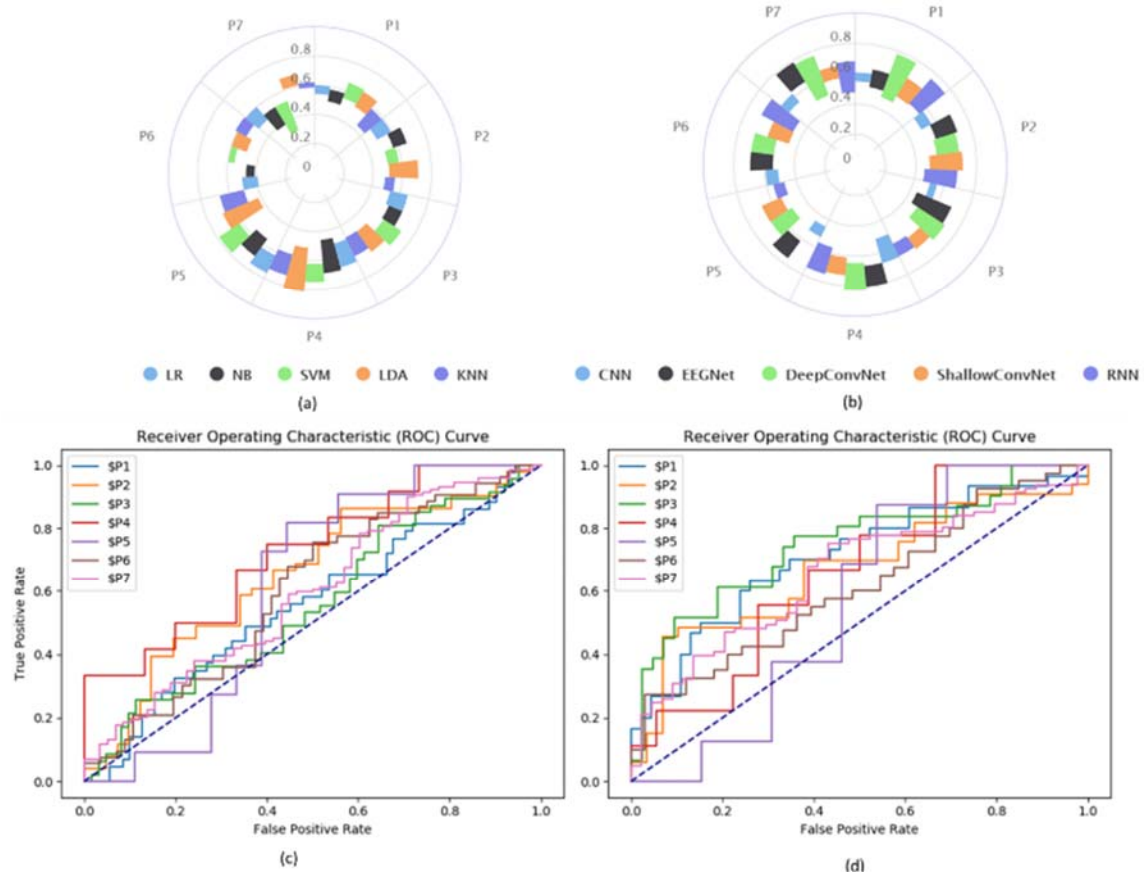

Fig. 3 Results of experiment #1: (a) Performance of machine-learning classifiers for individual intention prediction, (b) Performance of deep-learning classifiers for individual intention prediction; for both (a) and (b), the range is calculated by mean ± standard deviation, (c) ROC curve of the best machine learning classifier for each participant, (d) ROC curve of the best deep learning classifier for each participant; for both (c) and (d), the dashed dark blue line represents the threshold AUC = 0.5

TABLE II
COMPARISON OF BEST-PERFORMANCE MACHINE-LEARNING CLASSIFIER AND DEEP-LEARNING CLASSIFIER FOR EXPERIMENT #1

Participant	Best-performance Classifier	Mean Accuracy ± std	Min	Max	AUC
P1	RNN	65.0±10.9	54.1	75.9	0.71
P2	EEGNet	63.3±7.5	55.8	70.8	0.67
P3	SVM Linear	63.8±7.3	56.5	71.1	0.56
P4	EEGNet	73.3±7.0	66.3	80.3	0.66
P5	SVM Linear	69.9±8.4	61.5	78.3	0.61
P6	EEGNet	61.7±7.5	54.2	69.2	0.62
P7	EEGNet	71.7±7.5	64.2	79.2	0.66

Table III shows the overall best-performance classifier for each participant in experiment #1. From Table III, we can see that five over seven participants get the best accuracy from the deep learning classifier, and among them, four participants get the best accuracy from the EEGNet classifier. The highest accuracy among all the participants was 73.3 ± 7.0 and was achieved by participant 4 using the EEGNet classifier.

TABLE III
BEST-PERFORMANCE CLASSIFIER FOR EXPERIMENT #1

Participant	Classifier	Mean Accuracy ± Std	Min	Max	AUC
Best-Performance Machine-Learning Classifier					
P1	SVM Linear	60.8±5.6	55.2	66.4	0.55
P2	Naïve Bayes	61.4±4.9	56.5	66.3	0.65
P3	SVM Linear	63.8±7.3	56.6	71.1	0.56
P4	SVM Linear	68.1±6.2	61.8	74.3	0.72
P5	SVM Linear	69.9±8.4	61.5	78.3	0.61
P6	SVM Linear	57.6±2.1	55.5	59.6	0.60
P7	LDA	65.1±3.8	61.3	68.9	0.57
Best-Performance Deep-Learning Classifier					
P1	RNN	65.0±10.9	54.1	75.9	0.71
P2	EEGNet	63.3±7.4	55.9	70.8	0.67
P3	ShallowConvNet	63.3±4.5	58.8	67.9	0.75
P4	EEGNet	73.3±6.9	66.4	80.3	0.66
P5	EEGNet	68.3±6.9	61.4	75.3	0.57
P6	EEGNet	61.7±7.5	54.2	69.1	0.62
P7	EEGNet	71.7±7.5	64.2	79.1	0.66

We analyzed the receiver operating characteristic (ROC) and area under the ROC curve (AUC) for the best-performance machine-learning models and best-performance deep-learning models of the two experiments. ROC curve can be used to check the ability to distinguish classes. ROC is a probability curve with false positive rate as the x-axis and true positive rate as the y-axis. AUC value is between 0 and 1, and a higher AUC value indicates better model performance. For experiment #1, the best AUC was 0.71 from participant 1 using the deep learning method RNN. The AUCs for deep learning methods are all higher than those for machine learning methods (see Table II for details). Based on this result, it is concluded that for experiment #1, EEGNet is outstanding and can be considered as a good inter-subjective classifier for intension prediction purposes. However, there are three participants achieved better accuracy using other classifiers, the best model for intention prediction should be an intra-subjective model. In experiment #1, an average accuracy of 67% was achieved for the results of individual classifiers.

TABLE IV
COMPARISON OF BEST-PERFORMANCE MACHINE-LEARNING CLASSIFIER AND DEEP-LEARNING CLASSIFIER FOR EXPERIMENT #2

Participant	Best-performance Classifier	Mean Accuracy ± Std	Min	Max	AUC
P1	CNN	73.8 ± 3.6	70.2	77.4	0.62
P2	EEGNet	75.0 ± 7.7	67.3	82.7	0.56
P3	RNN	80.0 ± 2.8	77.2	82.8	0.83
P4	RNN	72.5 ± 5.6	66.9	78.1	0.64
P5	CNN	78.8 ± 3.4	75.4	82.2	0.57
P6	LDA	74.1 ± 5.0	79.1	69.1	0.76
P7	SVM	57.1 ± 2.2	54.9	59.3	0.44

TABLE V
BEST-PERFORMANCE CLASSIFIER FOR EXPERIMENT #2

Participant	Classifier	Mean Accuracy ± Std	Min	Max	AUC
Best-Performance Machine-Learning Classifier					
P1	LDA	68.9 ± 3.1	65.8	71.9	0.53
P2	LDA	68.6 ± 0.3	68.3	68.8	0.56
P3	SVM Linear	65.9 ± 0.3	65.6	66.2	0.62
P4	Logistic Regression	59.8 ± 9.6	50.2	69.4	0.54
P5	SVM Linear	69.3 ± 1.8	67.4	71.1	0.55
P6	SVM Linear	74.1 ± 5.0	69.1	79.1	0.76
P7	SVM Linear	57.1 ± 2.2	54.9	59.3	0.44
Best-Performance Deep-Learning Classifier					
P1	CNN	73.8 ± 3.6	70.2	77.4	0.62
P2	EEGNet	75.0 ± 7.7	67.3	82.7	0.56
P3	RNN	80.0 ± 2.8	77.2	82.8	0.83
P4	RNN	72.5 ± 5.6	66.9	78.1	0.64
P5	CNN	78.8 ± 3.4	75.4	82.2	0.57
P6	EEGNet	63.8 ± 13.6	50.2	77.3	0.61
P7	CNN	58.1 ± 7.2	50.9	65.3	0.49

Experiment #2 provided data to perform the analysis on agreement perception. The performance results of each classifier are shown in Fig. 4. Figs. 4 (a) and (b) indicate that the overall mean accuracy of deep learning classifiers is higher than the mean accuracy of machine learning classifiers. However, the standard deviations in the deep learning classifiers are much larger than those in machine learning methods. The results indicated an unstable accuracy when performing the deep learning classifiers for this experiment purpose. Figs. 4 (c) and (d) demonstrate the ROC curves of the best-performance classifiers of each participant. The curves in Fig. 4 (c) are for machine learning classifiers. Some of the curves are under the threshold of AUC = 0.5, which indicates a relatively bad performance that is even worse than random choices. Curves in Fig. 4 (d) show a much better result, which represents the best deep learning performance of each participant. Details of Fig.4 are listed in Table IV. Comparing the results separately, for the machine learning classifiers, the SVM Linear classifier achieved the best performance among the seven participants. Four over seven participants were found to reach the best mean accuracy using the SVM Linear classifier, while the highest accuracy of 74.1 ± 5.0 was also achieved by the classifier. For deep learning classifiers, in this experiment, there was no obvious outstanding classifier among the seven participants. However, there was a participant who achieved an accuracy of 80 ± 2.8, which was the highest

accuracy among all the participants, and was the only one who achieved an accuracy of more than 80. It is concluded that deep learning classifiers perform better than machine learning classifiers in most cases (see Tables IV and V for details).

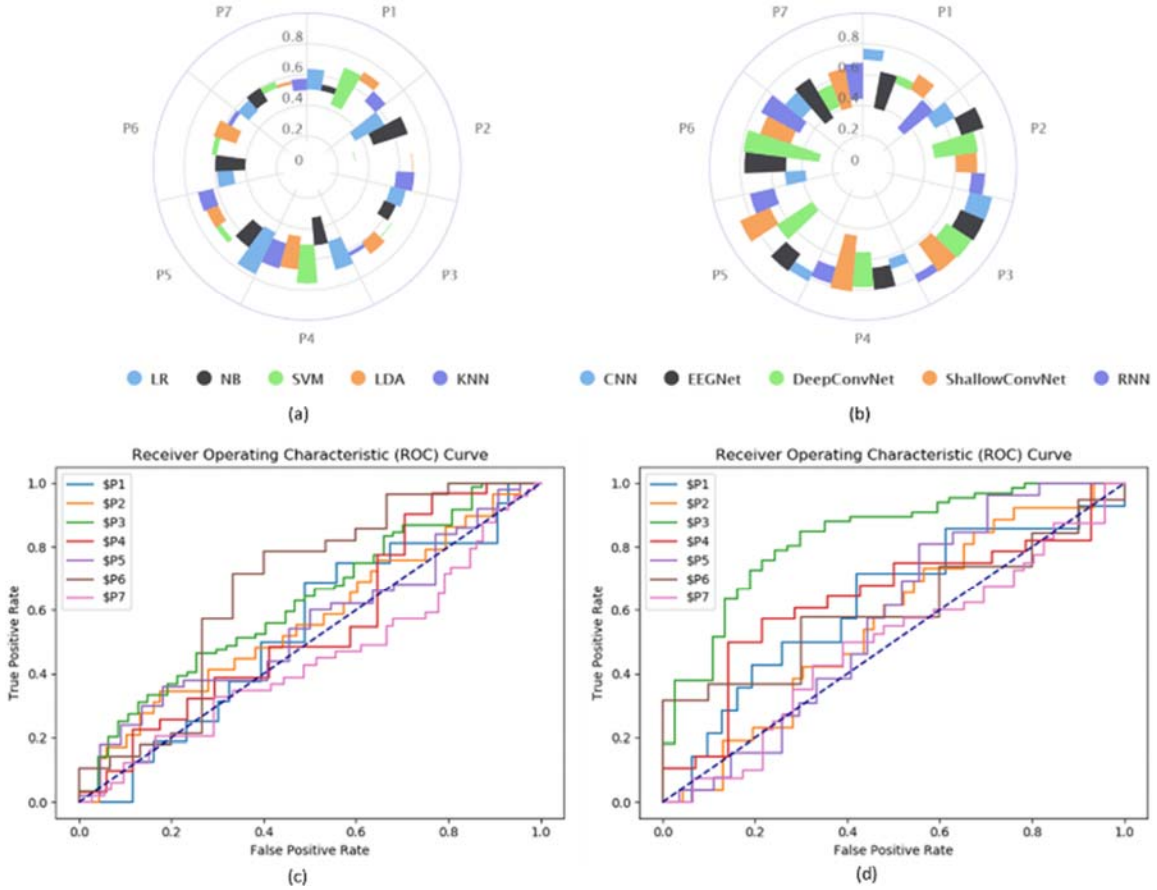

Fig. 4 Results of experiment #2: (a) Performance of machine-learning classifiers for individual intention prediction; (b) Performance of deep-learning classifiers for individual intention prediction. For both (a) and (b), the range is calculated by mean ± standard deviation; (c) ROC curve of the best machine learning classifier for each participant; (d) ROC curve of the best deep learning classifier for each participant. For both (c) and (d), the dashed dark blue line represents the threshold AUC = 0.5

In Table V, although there is no dominant classifier like EEGNet in experiment #1, the result indicated that generally, deep learning classifiers have a higher proportion of being the best classifier for a participant. There are five over seven participants achieved the best prediction performance using deep learning classifiers, among which CNN and RNN were both chosen twice as the best classifier. The highest mean accuracy was achieved by participant 3, which is 80.0 with a standard deviation of only 2.8. Participant 3 also got the highest AUC of 0.83. It is concluded that the deep learning classifiers achieve higher accuracy than the machine learning methods when comparing the accuracy across the participants. The variety of best classifiers among different participants in experiment #2 indicated the fact that the EEG signal is highly independent, thus the intra-subjective model for this kind of signal will perform better than the inter-subjective model. In experiment #2, the average accuracy was 72% for the individual classifiers.

VI. Discussion

This study proves the feasibility of using EEG signals to detect human intentions and the level of consensus on robot-provided information under emerging response situations. Instead of assessing intentions and perceptions that are triggered by specific events, this research directly treats the human intention as entire brain activity. Therefore, it provides a BCI as an alternative to communicate intentions and perceptions in human-robot collaboration during emergency response. It has the great potential to facilitate search and rescue when manual control and operation and natural language-based commands are not feasible. Moreover, the EEG-based consensus assessment proposed in this study allows the machine to understand human's agreement on robot-provided information, which is helpful to establish a BCI to evaluate human trust level and further improve trust in automation by adjusting machine behavior.

This research provides insights on selecting appropriate models for EEG-based intention perception recognition by comparing various machine learning and deep learning models.

The models were compared based on the AUC values, which are commonly used as a summary of the model skill. In general, deep-learning-based methods performed better for most participants than traditional machine-learning methods. It is found that EEG signals are individually different, and intra-subjective approaches would achieve a better result than inter-subjective approaches. This may be because that human intention and agreement perception is complex brain activity and almost all the cortical regions are involved in the process, which will cause significant differences among different people. Future studies can examine the impact of subject-specific information, such as age, gender, and level of experience on the classifier selection, and build suitable classifiers for groups of individuals that share the common characteristics.

Because of the limitation of current EEG sensing technology, a quiet and stable environment is required, and users were asked to stay steady to reduce artifacts. Movements of users may lead to massive noise and get unsatisfactory results. This hardware limitation may hinder the EEG-based application in the real search-and-rescue situation, where the disaster site is noisy and distracting, and the first responders and spontaneous volunteers need to walk around the site to observe the disaster site. However, the use of EEG signals for guiding robots during emergency response is still promising with teleoperation. If the first responders and volunteers cannot reach the disaster site in first-time, machines such as unmanned autonomous vehicles can be used to transmit the picture or recording to the first responders and allow them to teleoperate the search-and-rescue process. By detecting user intention to search certain areas, the on-site assistive intelligent machine could get more data for the interested area. By user perception assessment, the intelligent machine could understand the effect of current human-robot collaboration and adjust the provided information in time.

Although the accuracy is comparable with the state of the art in the area of implicit intention and perception prediction using EEG, it is still not reliable enough for practical application. The wrong prediction may lead to a massive decrease in human trust, especially for significant decisions in search and rescue. The assisted robots are expected to share responsibility with human partners so that guaranteed correct prediction is needed. The performance of prediction needs to be further improved in future research. In the experiment, only limited participants were involved. More participants could be involved in future research to make the result more robust.

VII. Conclusion

An EEG-based method is proposed to enhance the communication between human and robot in urban search and rescue environments by decoding brain signals. To better assist first responders and spontaneous volunteers, an intelligent robot should understand human intention and perception of the current situation. In this paper, we identified and predicted human intention in the first experiment, and studied how human subjects responded to robot advice in the second experiment. To decode human intention, EEG data were used to study the physiological response of subjects when they make decisions on their own. Individual classifiers are established to decode and predict subjects' intentions. To study human agreement perception during a human-robot collaboration task, the human response to robot advice was analyzed. Individual classifiers are developed to distinguish subjects' agreement perception of the robot. Different machine learning and deep learning algorithms were used to develop the intention prediction models and perception assessment models. For machine learning algorithms, significant features were extracted, and an optimal number of features was selected to develop the best model given the data. For the deep learning algorithm, frequency features were selected automatically based on the frequency and/or spatial characteristics. It is concluded that the features of frequency bands play a more important role than other features during the feature selection process. By comparing the performance of machine learning models, it is found that SVM Linear and LDA classifiers have higher mean accuracy. The comparison between the performance of machine learning classifiers and deep learning classifiers indicates that generally, deep learning models performed better than traditional machine-learning algorithms. It is concluded that human intention and agreement perception to assistive robots in search-and-rescue situations to some extent can be captured and predicted by EEG signals using various classification algorithms. However, the prediction accuracy varies among different individuals. The accuracy of a generalized classifier cannot be guaranteed since the psychological signals vary significantly among participants, which is in accordance with current findings. Individual classifiers would be more helpful for search-and-rescue situations.

References

[1] J. Ayers, J. L. Davis, and A. Rudolph, "Neurotechnology for biomimetic robots," p. 636, 2002.

[2] N. Pezeshkian, H. G. Nguyen, and A. Burmeister, "Unmanned ground vehicle radio relay deployment system for non-line-of-sight operations," *undefined*, Jan. 2007, doi: 10.21236/ADA475525.

[3] E. C. Leuthardt, G. Schalk, J. Roland, A. Rouse, and D. W. Moran, "Evolution of brain-computer interfaces: Going beyond classic motor physiology," *Neurosurg. Focus*, vol. 27, no. 1, 2009, doi: 10.3171/2009.4.FOCUS0979.

[4] R. Richer, N. Zhao, J. Amores, B. M. Eskofier, and J. A. Paradiso, "Real-time Mental State Recognition using a Wearable EEG," *undefined*, vol. 2018, pp. 5495–5498, Jul. 2018, doi: 10.1109/EMBC.2018.8513653.

[5] H. Adeli and S. Ghosh-Dastidar, "Automated EEG-based diagnosis of neurological disorders: inventing the future of neurology," p. 387, 2010.

[6] J. M. Fuster, "Frontal lobe and cognitive development," *J. Neurocytol.*, vol. 31, no. 3–5, pp. 373–385, Mar. 2002, doi: 10.1023/A:1024190429920.

[7] D. R. Euston, A. J. Gruber, and B. L. McNaughton, "The role of medial prefrontal cortex in memory and decision making," *Neuron*, vol. 76, no. 6, pp. 1057–1070, Dec. 2012, doi: 10.1016/J.NEURON.2012.12.002.

[8] A. Rehman and Y. Al Khalili, "Neuroanatomy, Occipital Lobe," *StatPearls*, Jul. 2021, Accessed: Nov. 11, 2021. (Online). Available: https://www.ncbi.nlm.nih.gov/books/NBK544320/.

[9] R. Lindenberg and L. Scheef, "Supramodal language comprehension: role of the left temporal lobe for listening and reading," *Neuropsychologia*, vol. 45, no. 10, pp. 2407–2415, 2007, doi: 10.1016/J.NEUROPSYCHOLOGIA.2007.02.008.

[10] L. Fogassi, P. F. Ferrari, B. Gesierich, S. Rozzi, F. Chersi, and G. Rizzolotti, "Neuroscience: Parietal lobe: From action organization to intention understanding," *Science (80-.).*, vol. 308, no. 5722, pp. 662–667, Apr. 2005, doi: 10.1126/SCIENCE.1106138/SUPPL_FILE/FOGASSI_SOM.PDF.

[11] D. Bzdok, G. Hartwigsen, A. Reid, A. R. Laird, P. T. Fox, and S. B. Eickhoff, "Left inferior parietal lobe engagement in social cognition and language," *Neurosci. Biobehav. Rev.*, vol. 68, pp. 319–334, Sep. 2016, doi: 10.1016/J.NEUBIOREV.2016.02.024.

[12] J. M. Spielberg, J. L. Stewart, R. L. Levin, G. A. Miller, and W. Heller, "Prefrontal Cortex, Emotion, and Approach/Withdrawal Motivation," *Soc. Personal. Psychol. Compass*, vol. 2, no. 1, pp. 135–153, Jan. 2008, doi: 10.1111/J.1751-9004.2007.00064.X.

[13] A. Maydeu-olivares, "Quantitative Methods in Psychology Quantitative Methods in Psychology," *Psychol. Bull.*, vol. 112, no. 1, pp. 155–159, 1992.

[14] M. Toscani, T. Marzi, S. Righi, M. P. Viggiano, and S. Baldassi, "Alpha waves: a neural signature of visual suppression," *Exp. brain Res.*, vol. 207, no. 3–4, pp. 213–219, Dec. 2010, doi: 10.1007/S00221-010-2444-7.

[15] M. Zaepffel, R. Trachel, B. E. Kilavik, and T. Brochier, "Modulations of EEG Beta Power during Planning and Execution of Grasping Movements," *PLoS One*, vol. 8, no. 3, p. e60060, Mar. 2013, doi: 10.1371/JOURNAL.PONE.0060060.

[16] Z. Yaple, M. Martinez-Saito, N. Novikov, D. Altukhov, A. Shestakova, and V. Klucharev, "Power of feedback-induced beta oscillations reflect omission of rewards: Evidence from an eeg gambling study," *Front. Neurosci.*, vol. 12, Oct. 2018, doi: 10.3389/FNINS.2018.00776/FULL.

[17] M. Gola, M. Magnuski, I. Szumska, and A. Wróbel, "EEG beta band activity is related to attention and attentional deficits in the visual performance of elderly subjects," *Int. J. Psychophysiol.*, vol. 89, no. 3, pp. 334–341, Sep. 2013, doi: 10.1016/J.IJPSYCHO.2013.05.007.

[18] X. Jia and A. Kohn, "Gamma Rhythms in the Brain," *PLoS Biol.*, vol. 9, no. 4, Apr. 2011, doi: 10.1371/JOURNAL.PBIO.1001045.

[19] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, "Emotion Assessment: Arousal Evaluation Using EEG's and Peripheral Physiological Signals," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4105 LNCS, pp. 530–537, 2006, doi: 10.1007/11848035_70.

[20] E. Lew, R. Chavarriaga, S. Silvoni, and J. del R. Millán, "Detection of self-paced reaching movement intention from EEGsignals," *Front. Neuroeng.*, vol. 5, no. July, Jul. 2012, doi: 10.3389/FNENG.2012.00013.

[21] Y. Wang and S. Makeig, "Predicting Intended Movement Direction Using EEG from Human Posterior Parietal Cortex," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5638 LNAI, pp. 437–446, 2009, doi: 10.1007/978-3-642-02812-0_52.

[22] C. Babiloni *et al.*, "Human movement-related potentials vs desynchronization of EEG alpha rhythm: a high-resolution EEG study," *Neuroimage*, vol. 10, no. 6, pp. 658–665, 1999, doi: 10.1006/NIMG.1999.0504.

[23] I. H. Kim, J. W. Kim, S. Haufe, and S. W. Lee, "Detection of braking intention in diverse situations during simulated driving based on EEG feature combination," *J. Neural Eng.*, vol. 12, no. 1, p. 016001, Nov. 2014, doi: 10.1088/1741-2560/12/1/016001.

[24] S. Ge *et al.*, "Temporal-Spatial Features of Intention Understanding Based on EEG-fNIRS Bimodal Measurement," *IEEE Access*, vol. 5, pp. 14245–14258, 2017, doi: 10.1109/ACCESS.2017.2723428.

[25] S.-H. Oh, "Subject Independent Classification of Implicit Intention Based on EEG Signals," *Int. J. Contents*, vol. 12, no. 3, pp. 12–16, Sep. 2016, doi: 10.5392/IJOC.2016.12.3.012.

[26] J. S. Kang, U. Park, V. Gonuguntla, K. C. Veluvolu, and M. Lee, "Human implicit intent recognition based on the phase synchrony of EEG signals," *Pattern Recognit. Lett.*, vol. 66, pp. 144–152, Nov. 2015, doi: 10.1016/J.PATREC.2015.06.013.

[27] X. Wang, Y. Huang, Q. Ma, and N. Li, "Event-related potential P2 correlates of implicit aesthetic experience," *Neuroreport*, vol. 23, no. 14, pp. 862–866, Oct. 2012, doi: 10.1097/WNR.0B013E3283587161.

[28] W. Zhang, J. Jin, A. Wang, Q. Ma, and H. Yu, "Consumers' Implicit Motivation of Purchasing Luxury Brands: An EEG Study," *Psychol. Res. Behav. Manag.*, vol. 12, p. 913, 2019, doi: 10.2147/PRBM.S215751.

[29] M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein, "ERP components on reaction errors and their functional significance: a tutorial," *Biol. Psychol.*, vol. 51, no. 2–3, pp. 87–107, Jan. 2000, doi: 10.1016/S0301-0511(99)00031-9.

[30] A. F. Salazar-Gomez, J. Delpreto, S. Gil, F. H. Guenther, and D. Rus, "Correcting robot mistakes in real time using EEG signals," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 6570–6577, Jul. 2017, doi: 10.1109/ICRA.2017.7989777.

[31] 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," *Soc. Neurosci.*, vol. 14, no. 2, pp. 226–235, Mar. 2019, doi: 10.1080/17470919.2018.1433717.

[32] S. Oh, Y. Seong, S. Yi, and S. Park, "Neurological Measurement of Human Trust in Automation Using Electroencephalogram," *Int. J. Fuzzy Log. Intell. Syst.*, vol. 20, no. 4, pp. 261–271, 2020, doi: 10.5391/IJFIS.2020.20.4.261.

[33] M. Wang, A. Hussein, R. F. Rojas, K. Shafi, and H. A. Abbass, "EEG-Based Neural Correlates of Trust in Human-Autonomy Interaction," *Proc. 2018 IEEE Symp. Ser. Comput. Intell. SSCI 2018*, pp. 350–357, Jan. 2019, doi: 10.1109/SSCI.2018.8628649.

[34] K. Akash, W.-L. Hu, N. Jain, and T. Reid, "A Classification Model for Sensing Human Trust in Machines Using EEG and GSR," *ACM Trans. Interact. Intell. Syst.*, vol. 8, no. 4, Mar. 2018, doi: 10.1145/3132743.

[35] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004, doi: 10.1016/J.JNEUMETH.2003.10.009.

[36] M. Akin, "Comparison of Wavelet Transform and FFT Methods in the Analysis of EEG Signals," *J. Med. Syst. 2002 263*, vol. 26, no. 3, pp. 241–247, Jun. 2002, doi: 10.1023/A:1015075101937.

[37] A. Yazdani, T. Ebrahimi, and U. Hoffmann, "Classification of EEG signals using Dempster Shafer theory and a K-nearest neighbor classifier," *2009 4th Int. IEEE/EMBS Conf. Neural Eng. NER '09*, pp. 327–330, 2009, doi: 10.1109/NER.2009.5109299.

[38] A. Subasi and E. Erçelebi, "Classification of EEG signals using neural network and logistic regression," *Comput. Methods Programs Biomed.*, vol. 78, no. 2, pp. 87–99, May 2005, doi: 10.1016/J.CMPB.2004.10.009.

[39] A. Sharmila and P. Geethanjali, "DWT Based Detection of Epileptic Seizure from EEG Signals Using Naive Bayes and k-NN Classifiers," *IEEE Access*, vol. 4, pp. 7716–7727, 2016, doi: 10.1109/ACCESS.2016.2585661.

[40] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 141–144, Jun. 2003, doi: 10.1109/TNSRE.2003.814441.

[41] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, 2019, doi: 10.1088/1741-2552/AB0AB5.

[42] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *J. Neural Eng.*, vol. 16, no. 5, p. 051001, Aug. 2019, doi: 10.1088/1741-2552/AB260C.

[43] T. Gehr, "Electroencephalogram (EEG) Classification of Known Vs Unknown Skills: Unpublished. Data from Jo Shattuck's Neural Substrates of Motor Learning the Effect of Real Time Haptic Feedback on Kinesthetic Awareness and Motor Skill Performance View project Electroencephalogram (EEG) Classification," 2017, Accessed: Nov. 12, 2021. (Online). Available: https://www.researchgate.net/publication/333853132.

[44] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces," *J. Neural Eng.*, vol. 15, no. 5, Nov. 2016, doi: 10.1088/1741-2552/aace8c.

[45] R. Schirrmeister, L. Gemein, K. Eggensperger, F. Hutter, and T. Ball, "Deep learning with convolutional neural networks for decoding and visualization of EEG pathology," *2017 IEEE Signal Process. Med. Biol. Symp. SPMB 2017 - Proc.*, vol. 2018-January, pp. 1–7, Aug. 2017, doi: 10.1109/SPMB.2017.8257015.