Analyzing EEG Brainwaves to Decode Attentional State for Faces and Scenes

Maryam Norouzi, Supti Biswas

Abstract*—Attention is the ability to facilitate processing perceptually salient information while blocking irrelevant information to an ongoing task. For example, visual attention is a complex phenomenon of searching for a target while filtering out competing stimuli. In the present study, we developed a new Brain-Computer Interface (BCI) platform to decode brainwave patterns during sustained attention in a participant. Scalp electroencephalography (EEG) signals using a wireless headset were collected in real-time during a visual attention task. In our experimental protocol, we primed participants to discriminate a sequence of composite images. Each image was a fair superimposition of a scene and a face image. The participants were asked to respond to the intended subcategory (e.g., indoor scenes) while withholding their responses for the irrelevant subcategories (e.g., outdoor scenes). We developed an individualized model using deep learning and machine learning techniques to decode the attentional state of the participant based on their brainwaves. Our model revealed the instantaneous attention toward face and scene categories. We conducted the experiment with four volunteer participants. We implemented three distinct models: MLP, CNN, and SVM, achieving average classification accuracies of approximately $91.\%$, 91.25%, and 23.75% respectively, across the four categories. The present work was an attempt to reveal the momentary level of sustained attention using EEG signals. The platform may have potential applications in visual attention evaluation and closed-loop brainwave regulation in the future.

Keywords: Glove-Based Exoskeleton, Soft Actuator, Liquid Crystal Elastomers (LCEs),

I. INTRODUCTION

ATTENTION is generally a core function in human cognition and perception. Attention is the ability to facilitate processing perceptually salient information while blocking irrelevant information to an ongoing task. The ability to focus attention and encode information are among the brain's most important cognitive and perceptive functions [1]. In visual attention, for example, the brain searches for a target while ignoring competing stimuli. Sustained attention refers to a cognitive capability to maintain focus during a task [2]. Deficits in attention are commonly seen in various brain disorders such as Alzheimer's disease (AD) and related dementia, Traumatic Brain Injuries (TBI), and Post-traumatic Stress Disorder (PTSD) [3]. Improvement in attentional states can boost perceptual and cognitive functions [4]. In recent years, electroencephalography (EEG) has become a useful low-cost tool for evaluating and training attention [5]. In recent studies, attention has been investigated as an inevitable phenomenon for working memory. Recently, neurofeedback-based training showed promising results in cognitive rehabilitation. Neurofeedback is a type of biofeedback in which neural activity is measured and presented through sensory channels (e.g. visual stimuli) to the participant in real-time to facilitate self-regulation of the targeted neural substrates that underlie a particular behavior or pathology.

In this study, we aim to develop a convenient and generalizable neurofeedback-based technology for attention training. In this article, our focus is on its open-loop brain attentional training technology that has been integrated with the use of an electroencephalography (EEG) headset and brain-computer interface platform. EEG time series are collected during the tasks using an 8-channel EEG headset. In this study, 4 cognitively healthy human subjects are recruited for an open-loop pre-evaluation, the first phase of a three-phase attentional experiment. In this work, we develop a neural network model to decode participants' attentional states based on their brainwaves.

II. MATERIALS AND METHODS

This section covers the materials and components of the integrated BCI platform. The experimental protocol used to collect the scalp EEG from multiple participants during a sustained attention task is also described. Subsequently, the applied techniques for decoding EEG data are explained in detail.

A. Development of the BCI Platform

The Brain-Computer Interface (BCI) platform (Fig. 1) consists of an EEG headset, a dual-monitor workstation (Fig. 2), and data analysis software programmed by Python. The protocol requires participants to wear an EEG cap with electrodes positioned according to the 10-20 system and sit in front of the designated subject monitor. A Graphic User Interface (GUI) was developed to allow a practitioner to conveniently administer the experimental protocol. We employ four image categories (Female, Male, Indoor, Outdoor) as stimuli as shown in Table I. Each stimulus lasts 1000 milliseconds and is presented with a 50% transparency for each composite image subcategory.

TABLE I
THE TASK-RELEVANT AND TASK-IRRELEVANT IMAGES FOR EACH BLOCK.

Block Number	Task-Relevant Image	Task-Irrelevant Image
1,2	Female	Male
3,4	Male	Female
5,6	Outdoor	Indoor
5,6	Indoor	Outdoor

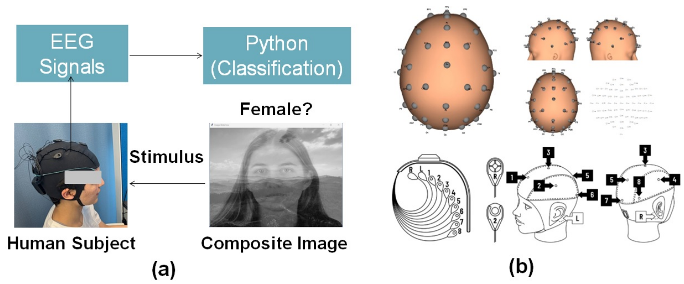
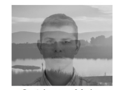


Fig. 1. The BCI platform a) A schematic of the EEG-based BCI system for decoding brainwaves during the pre-evaluation attention task. b) The electrode positions of the cap for the 8-electrode Unicorn system, specifically at Fz, C3, Cz, C4, Pz, PO7, Oz, and PO8, according to the 10/20 system.

TABLE II
A SAMPLE SEQUENCE OF TRIALS. A PARTICIPANT WAS EXPECTED TO RESPOND TO THE TASK-RELEVANT IMAGES BY CLICKING THE BUTTON WHILE IGNORING THE TASK-IRRELEVANT IMAGES (FROM LEFT TO RIGHT: TRIAL SEQUENCE).

Instruction (1000ms)	Composite Image (1000ms)	Composite Image (1000ms)	...	Composite Image (1000ms)
Indoor	 Indoor + Female	 Outdoor + Male	...	 Outdoor + Female
Expected Response	Relevant	Irrelevant	...	Irrelevant

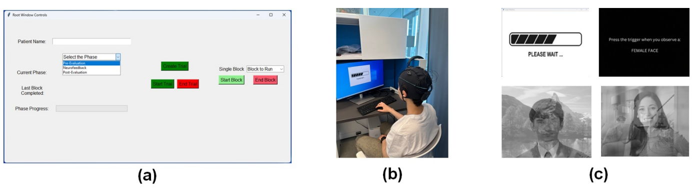


Fig. 2. Dual-monitor workstation a) Experimenter workstation. b) Subject c) Subject's workstation.

1) EEG Recording Device: EEG signals are acquired using a wireless headset called UNICORN HYBRID BLACK. The wearable EEG headset delivers high-quality EEG data from 8 Unicorn Hybrid EEG Electrodes, sampled with 24 Bits and 250 Hz per channel for many different BCI applications. The exact locations are labeled sequentially at Fz, C3, Cz, C4, Pz, PO7, Oz, and PO8, based on 10/20 international systems. The device collects brain signals and transmits the participant's brain signals to the PC with a Bluetooth connection.

2) Interface: We perform our data acquisition and analysis using Python. The visual stimulation and designed protocol are also controlled by it through a customized Graphic User Interface (GUI). For future works, the platform has also the capability to further adjust the transparency of image types in the superimposed image based on the attentional level of the instructed subcategory.

B. Experimental Protocol

Four subcategories of indoor/outdoor scene and male/female face images are chosen in this experiment as stimuli. Brightness and contrast for all face and scene images are adjusted so that the images on both face and scene categories have equalized and identical contrast. The images are all black and white with equal sizes, 800 × 1000 pixels. Face images are chosen to be neutral without any emotional expression. They are centered inside the composite image. Female faces have long hair and male ones have short hair. The indoor images are chosen from interior scenes. Outdoor images are natural landscapes. Our experimental protocol consisted of eight blocks of trials with a respite between blocks. Each block started with a five-second texture cue instructing the attended subcategory image, followed by 40 trials of image stimuli. The duration of each trial is set to one second. A trial includes a greyscale overlaid picture in which 50% of opacity is from the scene (indoor or outdoor) category and 50% is from the face (male or female) category. There is no repetition of face or scene images through each block of the experiment. This process helpes prevent any learning mechanism from happening to the participant. Participants are asked to identify whether the shown image contained the task-relevant image (e.g., an indoor image) by responding to each superimposed image. Participants are trained with a sequence of superimposed images of face and scene.

Four healthy participants, with a mean age of 30 years, voluntarily completed eight training blocks of the experiment, focusing on a monitor and responding via a trigger button. All participants had a normal or corrected-to-normal vision. They had no history of neurological or psychological disorder (based on self-report). All the subjects had an academic degree. The experimental protocol was approved by the Institutional Review Board at the University of Rhode Island (URI). All participants gave written consent to perform the experiment. The computerized task was provided by a PC with dual monitors. One monitor was viewed by the experimenter to control the experiment (Fig. 2a). The other monitor was positioned in front of the participants for the presentation of stimuli (Fig. 2b). The participants were asked to sit comfortably in a fixed chair focusing on a monitor and responding via a trigger button. They were also instructed to limit their excessive body movement. The participants were also asked to fixate their gaze on the middle of the screen while observing the stream of images. They were asked to push the keyboard key for each recognized relevant image and withhold their responses for irrelevant images (Fig. 3). The task-relevant subcategory images were distributed within each block in such a way that 90% of the composite images contained images from the task-relevant subcategory (e.g., indoor image) while the other half of composite images contained images from the task-irrelevant subcategory (e.g., outdoor image). Table II illustrates a sample sequence of composite images during a block and also the corresponding expected responses from participants. We alternated the task-relevant and task-irrelevant images among the four image subcategories as shown in Table I. Because of the difficulty in keeping constant sustained attention to the composite images during a block, we ran each block one time to prevent any fatigue from happening for the participant. The total time for the experiment was about 15 minutes per participant.

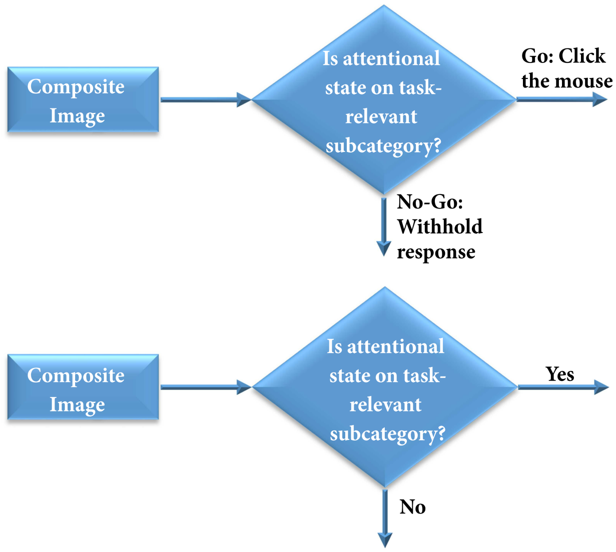


Fig. 3. Experimental Protocol for computing the behavioral response for a participant (top) and calculating the brainwave classification result for a participant (bottom).

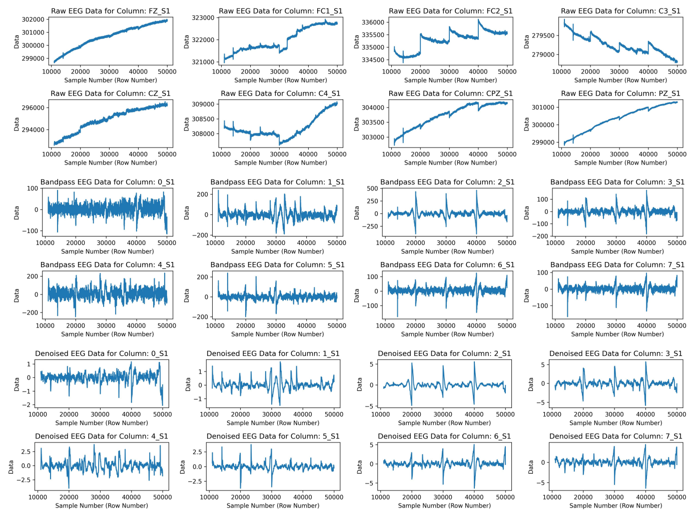


Fig. 4. Sequential preprocessing steps: from raw data to band-dassed and denoised data.

C. Signal pre-processing

In this study, we conducted a meticulous preprocessing of the EEG signals collected. Initially, we employed a band-pass filter to ensure only relevant frequencies were preserved, specifically within the range of 0.1 Hz to 30 Hz, a typical range for EEG data. This was executed using a Butterworth filter of the 5th order, with a sampling frequency set at 250 Hz. Following this, we performed a denoising process to eliminate any potential noise that could interfere with our data analysis. The denoising was conducted using the K-nearest neighbors regression algorithm, which was tailored individually for each EEG channel. The number of neighbors used in the regression was set to 50 for all channels, a value determined through prior optimization processes. Post band-pass filtering and denoising, the EEG signals were standardized using Z-scoring to provide each EEG channel equal weightage, irrespective of their original scale. Lastly, any linear trends in the EEG signals were removed using a detrending process. This comprehensive preprocessing pipeline ensured that our EEG data was primed for effective utilization in subsequent neural network models for the classification of attentional states.

1) Data Preparation and Processing Pipeline: In this study, a windowing technique is used to divide the continuous data, which was recorded at a rate of 250 Hz, into smaller, manageable sections. Each window encapsulated 250 data points, effectively representing one second of the recorded data. This process is performed for every second of data in the dataset, creating a series of 1-second snapshots of the entire recorded period. This dataset of windows is then shuffled to ensure that the order of the data does not introduce any bias into subsequent analyses, allowing for a fair and balanced exploration of the data. The next data processing pipeline involves several steps. Firstly, the data is reshaped to consider each 40-millisecond segment as one sample. This reshaping operation results in an array of 288 samples, each composed of 25 segments, with 10 data points and 8 channels in each segment. Next, the label array is similarly reshaped to match the structure of the data array. A squeeze operation is applied to flatten the label array into a simpler 1D structure, which is then used for oversampling. Oversampling is conducted to balance the dataset. The RandomOverSampler function from the imbalanced-learn library is used, which works by duplicating the minority class instances to equalize the number of instances across all classes. The resulting resampled dataset is then converted to the appropriate data types for further processing. The resampled dataset is then split into training and testing sets, with 90% of the data used for training and 10% reserved for testing. An additional untouched set is also separated out for validation purposes. Finally, the labels for the training, testing, and untouched sets are converted into a categorical format to be compatible with the requirements of the Keras library used for modeling. The categorical format represents the class labels as binary vectors, which is the required format for multi-class classification tasks in Keras. As well, to prepare the data for the CNN, the training and testing datasets are then reshaped into 4-dimensional arrays. The reshaping operation expands the dimensions of the data, adding two additional dimensions for the spatial (width and height) and channel dimensions required by CNNs. This allows the CNN to process the data as 2D images with single color channels, which is the standard input format for CNNs. The reshaping operation does not alter the underlying data but simply restructured it into a format compatible with the CNN architecture.

D. Classification Method

In the context of EEG signal processing, deep learning models have shown promising results for classification tasks, particularly in the area of brain-computer interfaces (BCIs) for rehabilitation and assistive technologies. In this study, we design and implement three models: Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), and Support

Vector Machine (SVM) for the classification of EEG signals. We experiment with different model architectures, activation functions, and hyperparameters to find the optimal configuration that yields the best classification performance. To evaluate the performance of the neural network model, we use a separate test dataset and compute metrics such as accuracy and precision. Additionally, we use visualization techniques such as confusion matrices and accuracy/loss plots to gain insights into the model's performance and potential areas for improvement.

1) Multilayer Perceptron: Our MLP model is constructed using the *Keras Sequential API*, with its architecture comprising an input layer, two hidden layers, and an output layer. The input layer is configured with 320 nodes corresponding to the 80-dimensional input. The hidden layers, each containing 800 nodes, are designed with the *'relu'* activation function, a popular choice for its ability to mitigate the vanishing gradient problem. To prevent overfitting, a dropout regularization technique is employed with a rate of 0.01, effectively ignoring random fractions of the nodes during training. Our output layer had 4 nodes, corresponding to the 4 classes of attentional states, and uses a *'softmax'* activation function for probabilistic classification. The model is compiled with the *'rmsprop'* optimizer due to its efficient handling of sparse gradients and online and non-stationary settings, and a learning rate of 0.0006. The training process is performed for 20 epochs with a batch size of 32. The model optimization is guided by the categorical cross-entropy loss function, while model performance is evaluated based on accuracy.

2) Convolutional Neural Network: The architecture of our CNN model consists of a single convolutional layer with 512 filters and a kernel size of $1x1$, which is applied to the input data with dimensions of $80x1x1$. This is then flattened to a $1D$ array for input to the subsequent fully connected layers. The model then includes three dense hidden layers, each containing 100 neurons and utilizing the Rectified Linear Unit (ReLU) activation function. The final layer is a dense layer with 4 neurons, corresponding to the number of classes in our classification problem, and uses a *softmax* activation function for output, thus providing the probability distribution over the four classes. The model uses the *Adam* optimizer with a custom learning rate of 0.0006 and is trained to minimize the categorical cross-entropy loss. The model's performance IS evaluated over 20 epochs with a batch size of 32, using separate validation data.

3) Support Vector Machine: We implement an SVM model for classification. Initially, the one-hot-encoded labels of the target variables are converted back to integer-encoded labels to meet the input requirements of the SVM model. Subsequently, an SVM model is constructed using a Radial basis function (rbf) kernel and one-vs-rest (ovr) decision function shape, which is typically used for multi-class classification tasks. The *'gamma'* parameter is set to *'auto'*, allowing the model to automatically adjust its fit to the data. The SVM model is trained using the preprocessed training set and subsequently makes predictions on the test set. The model's performance is evaluated based on its accuracy, i.e., the proportion of correct predictions made out of the total number of predictions. Furthermore, a detailed classification report is produced, providing additional performance metrics such as precision, recall, and F1-score for each class. For our SVM model, we employed the Hilbert transform for feature extraction to yield more informative input data. Specifically, we extracted the Hilbert envelope from the raw data, utilizing it as a unique input feature. This process allowed us to leverage the informative characteristics of the Hilbert envelope, enhancing the model's ability to distinguish between different classes and thus improving its overall performance in the classification task.

III. Results and Discussion

A. Result

This section illustrates the results of EEG decoding accuracy among sample datasets during a sustained attention task. We perform MLP, CNN, and SVM classification with the collected EEG signals dataset while participants were attending four classes of the scene (indoor/outdoor) and face (female/male) images. The scene (indoor/outdoor) accuracy indicates how accurately the model predicts the attentional state towards scene (indoor/outdoor) images whereas the face (female/male) accuracy indicates how accurately the model predicts the attentional state on face (female/male) images. On average, the accuracy of the MPL, CNN, and SVM models are around 91.75%, 91.25% 23.75%, respectively. Fig. **??** presents a comparative evaluation of accuracy and loss metrics for both MLP and CNN models, specifically for subjects 1 through 4. Similarly, Fig. **??** illustrates a comparative visualization of confusion matrix heatmaps for MLP, CNN, and SVM models, again focusing on subjects 1 to 4.

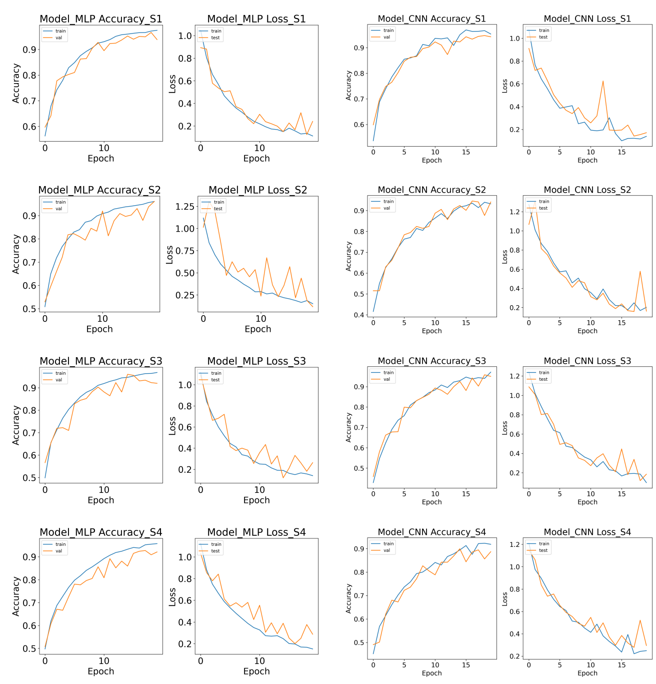


Fig. 5. Comparison of accuracy and loss metrics for MLP and CNN models across subjects 1-4

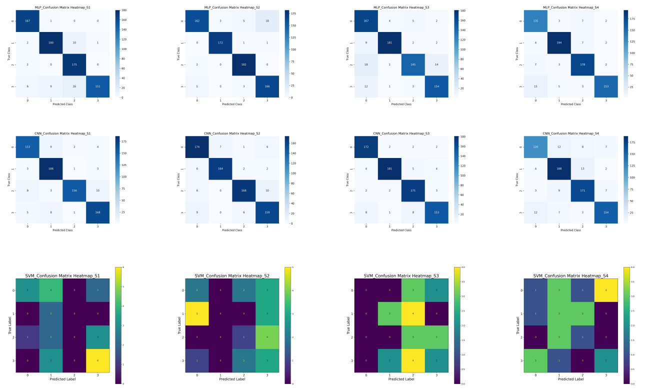


Fig. 6. Comparison of confusion matrix Heatmap for MLP, CNN and SVM models across subjects 1-4

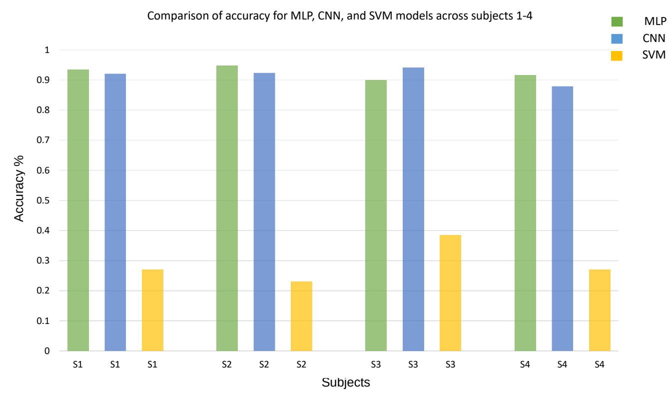


Fig. 7. Comparison of accuracy for MLP, CNN, and SVM models across subjects 1-4.

B. Discussion

Figure 7 provides a comparative view of the accuracy of MLP, CNN, and SVM models, specifically for subjects 1 through 4. Furthermore, Table III summarizes the comparative analysis of precision across different models for each class, encompassing all subjects. As demonstrated in Fig. 7 The MLP model achieves an approximate accuracy of 92%, slightly outperforming the CNN model, which stands at around 91%. This slight advantage of MLP over CNN is observed across most subjects, even when examining the precision of each class as a comparison metric. However, the accuracy and loss function plots (Fig. 5) reveal that the CNN model exhibits more smoothness, given the same number of epochs and learning rate, albeit at a slower speed than the MLP model. In contrast, the SVM model, configured with a linear kernel, significantly lags behind the aforementioned models in accuracy. Its inability to effectively predict the four categories of classes is evident, with precision results indicating that it struggles to predict any of the four classes with more than a 30%precision rate. In addition, both the MLP and CNN models demonstrated potential in EEG signal classification. As epochs progressed, the loss function decreased and accuracy improved in both (Fig. 5), achieving a solid final accuracy of around 90%, respectively, our model exceeded the 77% accuracy of a prior study [6].

TABLE III
COMPARATIVE EVALUATION OF PRECISION ACROSS DIFFERENT MODELS FOR EACH CLASS ACROSS ALL SUBJECTS

Subject Number	Precision %			
	Female	**Male**	**Indoor**	**Outdoor**
S1	MLP: 0.94 CNN: 0.90 SVM: 0.75	MLP: 0.94 CNN: 0.90 SVM: 0.18	MLP: 0.87 CNN: 0.97 SVM: 0.00	MLP: 0.99 CNN: 0.90 SVM: 0.50
S2	MLP: 0.94 CNN: 0.89 SVM: 0.00	MLP: 0.95 CNN: 0.95 SVM: 1.00	MLP: 0.95 CNN: 0.94 SVM: 0.00	MLP: 0.89 CNN: 0.89 SVM: 0.20
S3	MLP: 0.81 CNN: 0.94 SVM: 0.00	MLP: 0.96 CNN: 0.97 SVM: 0.30	MLP: 0.93 CNN: 0.91 SVM: 0.14	MLP: 0.89 CNN: 0.94 SVM: 0.40
S4	MLP: 0.83 CNN: 0.86 SVM: 0.50	MLP: 0.94 CNN: 0.37 SVM: 0.16	MLP: 0.91 CNN: 0.87 SVM: 0.20	MLP: 0.96 CNN: 0.90 SVM: 0.50

IV. CONCLUSION

IN this study, a new EEG-based BCI classification system is developed for evaluating attention during a visual discrimination task. The developed platform is able to collect EEG data in real time while presenting superimposed stimuli to a participant. A GUI is designed to give more flexibility and controllability to the practitioner for administering participants through the experiment. Four participants are recruited to test the feasibility of the system and to evaluate the viability of EEGbased classification of the participant's attentional state. EEG signals are collected from the whole brain and are sent to the computer for processing. For the purpose of data classification into four distinct categories, the labels 'Female', 'Male', 'Indoor', and 'Outdoor' are numerically encoded from 0 to 3, respectively. We develope and implement three types of models: MLP, CNN, and SVM. We utilize the Hilbert envelope for feature extraction for SVM. The average classification result between the four categories is around 91.75%, 91.25% 23.75%, for MLP, CNN, and SVM, respectively. The developed EEG-based BCI platform has the potential to be applied in real-time classification and neurofeedback tasks for diagnosing and training patients with attention deficits. In conclusion, our approach utilizing a deep neural network effectively streamlines EEG signal classification, offering significant potential for advancements in non-invasive brain signal analysis.

REFERENCES

[1] Mehdi Ordikhani-Seyedlar, Mikhail A Lebedev, Helge BD Sorensen, and Sadasivan Puthusserypady. Neurofeedback therapy for enhancing visual attention: state-of-the-art and challenges. *Frontiers in neuroscience*, 10:352, 2016.

[2] Megan T Debettencourt, Jonathan D Cohen, Ray F Lee, Kenneth A Norman, and Nicholas B Turk-Browne. Closed-loop training of attention with real-time brain imaging. *Nature neuroscience*, 18(3):470–475, 2015.

[3] Yang Jiang, Reza Abiri, and Xiaopeng Zhao. Tuning up the old brain with new tricks: attention training via neurofeedback. *Frontiers in aging neuroscience*, 9:52, 2017.

[4] Reza Abiri, Xiaopeng Zhao, and Yang Jiang. A real time eeg-based neurofeedback platform for attention training. In *Biomedical Engineering Society Annual Meeting (BMES 2016)*, 2016.

[5] Efthymios Angelakis, Stamatina Stathopoulou, Jennifer L Frymiare, Deborah L Green, Joel F Lubar, and John Kounios. Eeg neurofeedback: a brief overview and an example of peak alpha frequency training for cognitive enhancement in the elderly. *The clinical neuropsychologist*, 21(1):110–129, 2007.

[6] Reza Abiri, Soheil Borhani, Yang Jiang, and Xiaopeng Zhao. Decoding attentional state to faces and scenes using eeg brainwaves. *Complexity*, 2019, 2019.