

SYSTEM AND METHOD FOR COGNITIVE ENHANCEMENT BASED ON ELECTROENCEPHALOGRAM (EEG) SIGNALS

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of priority to the Singapore application no. 10202260087R filed on 14 November 2022, the contents of which are hereby incorporated by reference in their entirety for all purposes.

TECHNICAL FIELD

[0002] This application relates to a system and method for modifying presentation of digital content on a display based on a user's electroencephalogram (EEG) signals. Through the dynamic manipulation of the digital content, the user's cognitive ability may be enhanced.

BACKGROUND

[0003] Interactions between humans and computers are normally performed through peripheral devices such as a keyboard, mouse, and display. This type of interaction inevitably involves a user's neuromuscular system. In place of such peripheral devices, it has been proposed that a Brain-Computer Interface (BCI) system be used to establish direct information exchange between the human brain and computer systems. A BCI system allows a user to bypass neuromuscular communications and instead allows the user to control systems based on the user's brain activity.

[0004] The human brain produces biological signals, including electrical patterns known as brainwaves, which can be detected and recorded through an electroencephalogram (EEG). These brainwaves can be quantified using EEG devices, which usually capture them in analog form. Subsequently, the brainwave data can be analyzed in their original analog format or converted into a digital representation for further analysis. Most of the existing BCI systems employ techniques that utilize recorded EEG signals as it is an easier and relatively cheaper brain signal acquisition technique as compared to other non-invasive methods such as functional Magnetic Resonance Imaging (fMRI), functional Near-infrared Spectroscopy (fNIRS) and Magnetoencephalography (MEG).

[0005] Through interactions with a BCI system, an individual would be able to auto-regulate their brain signals in response to real-time stimulus. Thus, by training the brain, it

would be possible for an individual to rewire their underlying neural circuits through the process of neuroplasticity. Hence, those skilled in the art are constantly looking for ways to enrich or enhance a person's cognitive ability by controlling a BCI system using the person's brain activity.

SUMMARY

[0006] In one aspect, the present application discloses a computer system configured to modify presentation of digital content on a display based on acquired raw electroencephalogram (EEG) signals. The computer system comprises a computer processor; and a non-transitory media readable by the processing unit, the media storing instructions that when executed by the processing unit, causes the processing unit to provide the raw EEG signals as inputs to a trained neural network and a trained multilayer perceptron (MLP) network. The processing unit is then directed to generate, using the trained neural network, cognitive scores, wherein the neural network was trained using EEG signals associated with sustained attention cognitive states and EEG signals associated with inattention cognitive states and to generate, using the trained MLP network, physiological scores for each time-window of the raw EEG signals, wherein the MLP network was trained using EEG signal features associated with alpha and beta bands. The processing unit is then further directed to provide the cognitive and the physiological scores to a display controller module and to modify, using the display controller module, the presentation of the digital content on the display based on the cognitive and the physiological scores.

[0007] In another embodiment of this aspect, the present application discloses that the raw EEG signals are acquired from a user using a mobile bio-sensor device communicatively connected to the computer system. The instructions to modify the presentation of the digital content on the display based on the cognitive and the physiological scores then additionally includes instructions for directing the processing unit to display, using the display controller module, a spatial-temporal activity on the display and to instruct, using the mobile bio-sensor device, the user to continuously engage in the displayed spatial-temporal activity until it is determined that averaged cognitive scores over a cluster of time-windows of the raw EEG signals have achieved a predetermined target.

[0008] In yet another embodiment of this aspect, the present application discloses that the raw EEG signals are acquired from a user using a mobile bio-sensor device communicatively

connected to the computer system. The instructions to modify the presentation of the digital content on the display based on the cognitive and the physiological scores then additionally includes instructions for directing the processing unit to display, using the display controller module, a spatial-temporal activity on the display and to instruct, using the mobile bio-sensor device, the user to continuously engage in the displayed spatial-temporal activity until it is determined that averaged physiological scores over a cluster of time-windows of the raw EEG signals have exceeded a predetermined threshold score.

[0009] In another aspect, the present application discloses a method to modify presentation of digital content on a display based on acquired raw electroencephalogram (EEG) signals, the method comprising the steps of receiving, using a machine learning module, the raw EEG signals and provided the received raw EEG signals as inputs to a trained neural network and a trained multilayer perceptron (MLP) network; generating, using the trained neural network, cognitive scores, wherein the neural network was trained using EEG signals associated with sustained attention cognitive states and EEG signals associated with inattention cognitive states; and generating, using the trained MLP network, physiological scores for each time-window of the raw EEG signals, wherein the MLP network was trained using EEG signal features associated with alpha and beta bands. The method then comprises the steps of receiving, using a display controller module, the cognitive and the physiological scores; and modifying, using the display controller module, the presentation of the digital content on the display based on the cognitive and the physiological scores.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] Various embodiments of the present disclosure are described below with reference to the following drawings:

Figure 1 illustrates a block diagram of a computer system and mobile device that may be used to implement an EEG-based cognitive enhancement system in accordance with embodiments of the present disclosure;

Figure 2 illustrates a block diagram representative of a processing system for performing embodiments of the present disclosure;

Figure 3a illustrates a set of stimuli for generating sustained attention EEG signals for training a neural network in accordance with embodiments of the present disclosure;

Figure 3b illustrates a timing protocol for the calibration of the collection of training datasets used to train the neural network in accordance with embodiments of the present disclosure;

Figure 4 illustrates a flowchart of a process for enhancing a user's cognitive ability based on the user's EEG signals in accordance with embodiments of the present disclosure;

Figure 5 illustrates block diagram showing the four main parts of a multi-scale convolutional neural network in accordance with embodiments of the present disclosure;

Figure 6 illustrates a location map of a cap with 32 channels in accordance with embodiments of the present disclosure;

Figure 7 illustrates a histogram plot showing the classification performance of the trained neural network;

Figure 8 illustrates a multilayer perceptron neural network structure in accordance with embodiments of the present disclosure; and

Figure 9 illustrates a flowchart of a process for modifying presentation of digital content on a display based on raw EEG signals in accordance with embodiments of the present disclosure.

DETAILED DESCRIPTION

[0011] The following detailed description is made with reference to the accompanying drawings, showing details and embodiments of the present disclosure for the purposes of illustration. Features that are described in the context of an embodiment may correspondingly be applicable to the same or similar features in the other embodiments, even if not explicitly described in these other embodiments. Additions and/or combinations and/or alternatives as described for a feature in the context of an embodiment may correspondingly be applicable to the same or similar feature in the other embodiments.

[0012] In the context of various embodiments, the articles “a”, “an” and “the” as used with regard to a feature or element include a reference to one or more of the features or elements.

[0013] In the context of various embodiments, the term “about” or “approximately” as applied to a numeric value encompasses the exact value and a reasonable variance as generally understood in the relevant technical field, e.g., within 10% of the specified value.

[0014] As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0015] As used herein, “comprising” means including, but not limited to, whatever follows the word “comprising”. Thus, use of the term “comprising” indicates that the listed elements are required or mandatory, but that other elements are optional and may or may not be present.

[0016] As used herein, “consisting of” means including, and limited to, whatever follows the phrase “consisting of”. Thus, use of the phrase “consisting of” indicates that the listed elements are required or mandatory, and that no other elements may be present.

[0017] As used herein, the terms "EEG signal" or “EEG data” generally refer to Electroencephalography (EEG), obtained from one or more subjects whereby these signals correspond to the subject’s brain activity.

[0018] Further, one skilled in the art will recognize that certain functional units in this description have been labelled as modules throughout the specification. The person skilled in the art will also recognize that a module may be implemented as circuits, logic chips or any sort of discrete component. Furthermore, one skilled in the art will also recognize that a module may be implemented in software which may then be executed by a variety of processor architectures. In embodiments of the disclosure, a module may also comprise computer instructions or executable code that may instruct a computer processor to carry out a sequence of events based on instructions received. The choice of the implementation of the modules is left as a design choice to a person skilled in the art and does not limit the scope of the claimed subject matter in any way.

[0019] Further, one skilled in the art will also recognize that the detailed workings, internal structures, techniques and training datasets of machine learning models, neural networks and anomaly detection models referred to in this disclosure have not been disclosed or shown in detail as such information is known to one skilled in the art and as such, has been omitted for brevity.

[0020] Electroencephalography (EEG) is a commonly employed method for directly measuring human brain activity. It involves the placement of multiple electrodes on the surface of the human head to collect EEG signals. With its high temporal resolution, EEG can capture dynamic changes in brain activity at a sub-second level. Leveraging machine learning and

signal processing techniques, a Brain-Computer Interface (BCI) system has the capability to discern human emotions using EEG data.

[0021] Existing BCI systems for cognitive enhancement typically employ repetitive, non-versatile and exhausting training environments. As the subject's mental capacity degrades due to tiredness or low energy levels, this causes the quality of the EEG signals to deteriorate thereby reducing the effectiveness of the cognitive enhancement system. Additionally, as EEG signals vary between subjects due to individual differences, the results obtained from signal processing and conventional machine learning techniques (that are used to extract temporal, spectral or spatial features from EEG signals) will degrade when these techniques are applied to different subjects.

[0022] Figure 1 illustrates a block diagram of a computer system and mobile device that may be used to implement an EEG-based cognitive enhancement system in accordance with embodiments of the present disclosure. Computer system 110 continuously monitors and receives EEG signals and human-computer interaction (HCI) signals 108 from mobile device 102, which is provided with a user or subject of computer system 110. Mobile device 102, which includes bio-sensor device 104 and human-computer interaction (HCI) devices 106, is configured to acquire EEG signals and HCI signals and subsequently communicate these signals to computer system 110. The communication of signals 108 to computer system 110 by mobile device 102 may be through wired or wireless data transmission means.

[0023] In particular, bio-sensor device 104 is typically a non-invasive device designed to capture and measure the electrical activity of a user's brain. In embodiments of the disclosure, bio-sensor device 104 may comprise, but is not limited to, multiple electrodes/channels, often embedded in a head piece such as a headset, placed strategically on the user's scalp to pick up electrical signals generated by neuronal activity. These electrodes are usually attached to a processing system that processes the electrical signals into EEG signals whereby the processed EEG signals may then be transmitted to computer system 110. In further embodiments of the disclosure, the EEG signals may be recorded from the subject using dry electrodes whereby the electrodes are positioned near the frontal lobe of the subject (to collect EEG signals from the frontal area) as it is this area which contains attention related information. Bio-sensor device

104 may then be connected to computer system 110 via Bluetooth such that data may be transmitted wirelessly between these two components.

[0024] HCI devices 106 comprise a broad spectrum of tools, interfaces and systems that allow users to interact with computing systems/devices. HCI devices 106 may comprise, but is not limited to, graphical user interfaces (GUIs), touchscreens, keyboards or touchpads, voice recognition systems, gesture control system, augmented and virtual reality interfaces, eye tracking systems and any other devices that facilitate data entry, information retrieval, communication between a user and a computing system, and allows for items in control system 110 to be controlled by the user. Signals acquired by HCI devices 106 may then be provided to computer system 110. In further embodiments of the disclosure, the orientation of a subject's eye gaze may be estimated using an eye tracker mounted on display 130 such that the eye tracker is able to capture the subject's gaze and communicate this information to computer system 110. The eye tracker may include projectors to create a pattern of near infrared light on the subject's eyes, a camera to record the pattern projected on the eyes and the gaze of the subject, and algorithms to generate co-ordinates of the subject's gaze to determine the orientation direction of the subject's eyes. All this information may then be provided to display controller module 120 to control the presentation of information on display unit 130.

[0025] As illustrated in Figure 1, computer system 110 comprises machine learning module 111, display controller module 120 and display unit 130 wherein machine learning module 111 includes a neural network module which includes a neural network 112 and a multilayer perceptron (MLP) network module which includes MLP network 114. In embodiments of the disclosure, display controller module 120 may be configured to modify and/or control digital content shown on display unit 130 based on data received from machine learning module 111 and/or mobile device 102. In embodiments of the disclosure, display controller module 120 may modify the presentation of digital content shown on display unit 130 by directly utilizing HCI signals acquired from mobile device 102.

[0026] Neural network 112 may comprise, but is not limited to, a multi-scale convolutional neural network having various convolutional layers and branches that simultaneously extract features and information from different scales of input data; and MLP network 114 may comprise, but is not limited to, a type of neural network that has an input layer, one or more

hidden layers, and an output layer whose artificial neurons are all interconnected, whereby each neuron in the network is linked to every neuron in adjacent layers with weighted connections. Neural network 112 may be trained using EEG signals associated with sustained attention physiological states and EEG signals associated with inattention physiological states, while MLP network 114 may be trained using EEG signals features associated with alpha and beta bands.

[0027] In embodiments of the disclosure, before the raw EEG signals are processed by trained neural network 112 and trained MLP network 114, the raw signals may be pre-processed by machine learning module 111. During the pro-processing of the raw EEG signals, a band-pass filter (from 0.3 to 40 Hz) is applied to remove low and high-frequency noises from the raw EEG signals. Artifacts are then removed from the raw EEG signals by subtracting smoothed references from the EEG signals whereby the smoothed references are obtained using a moving averaging technique involving multiple moving windows.

[0028] For example, let $z \in \mathbb{R}^L$ denote band-pass filtered EEG signals recorded from a single electrode where L is defined as the number of data samples. The windows to compute a set of moving averages are defined as l_i , $i \in [1, m]$ where m is defined as the number of moving average windows. The moving averages are then computed as $z_{(l_i)}[t] = \frac{1}{l_i} \sum_t^{t+l_i} z[t]$, where t denotes samples of the data vector, for each window length l_i . The moving averages are subtracted from the EEG signals and averaged as $x = \frac{1}{m} \sum_{i=1}^m z - z_{(l_i)}$ to obtain artifact removed EEG signals. This step is repeated for EEG signals acquired from all the electrodes, to obtain $X = [x_1, \dots, x_C]^T \in \mathbb{R}^{C \times L}$, where C is defined as the number of electrodes (channels).

[0029] Once trained, trained neural network 112 and trained MLP network 114 may then generate cognitive scores and physiological scores respectively based on acquired EEG signals from mobile device 102. It should be noted that the computation of the cognitive 116 and physiological 118 scores by the respective networks occurs concurrently. Cognitive 116 and physiological scores 118 are then provided to display controller module 120 which in turn is configured to modify data presented on display unit 130 based on the received cognitive 116 and physiological 118 scores.

[0030] In accordance with embodiments of the present disclosure, a block diagram representative of components of processing system 200 that may be provided within computer system 110, mobile device 102 or any of the modules provided within system 110 and/or mobile device 102 to carry out any of the functions described above is illustrated in Figure 2. One skilled in the art will recognize that the exact configuration of each processing system provided within these modules may be different and the exact configuration of processing system 200 may vary and the arrangement illustrated in Figure 2 is provided by way of example only.

[0031] In embodiments of the disclosure, processing system 200 may comprise controller 201 and user interface 202. User interface 202 is arranged to enable manual interactions between a user and the computing module as required and for this purpose includes the input/output components required for the user to enter instructions to provide updates to each of these modules. A person skilled in the art will recognize that components of user interface 202 may vary from embodiment to embodiment but will typically include one or more of display 240, keyboard 235 and optical device 236.

[0032] Controller 201 is in data communication with user interface 202 via bus 215 and includes memory 220, processor 205 mounted on a circuit board that processes instructions and data for performing the method of this embodiment, an operating system 206, an input/output (I/O) interface 230 for communicating with user interface 202 and a communications interface, in this embodiment in the form of a network card 250. Network card 250 may, for example, be utilized to send data from these modules via a wired or wireless network to other processing devices or to receive data via the wired or wireless network. Wireless networks that may be utilized by network card 250 include, but are not limited to, Wireless-Fidelity (Wi-Fi), Bluetooth, Near Field Communication (NFC), cellular networks, satellite networks, telecommunication networks, Wide Area Networks (WAN) and etc.

[0033] Memory 220 and operating system 206 are in data communication with CPU 205 via bus 210. The memory components include both volatile and non-volatile memory and more than one of each type of memory, including Random Access Memory (RAM) 223, Read Only Memory (ROM) 225 and a mass storage device 245, the last comprising one or more solid-state drives (SSDs). One skilled in the art will recognize that the memory components described

above comprise non-transitory computer-readable media and shall be taken to comprise all computer-readable media except for a transitory, propagating signal. Typically, the instructions are stored as program code in the memory components but can also be hardwired. Memory 220 may include a kernel and/or programming modules such as a software application that may be stored in either volatile or non-volatile memory.

[0034] Herein the term “processor” is used to refer generically to any device or component that can process such instructions and may include: a microprocessor, microcontroller, programmable logic device or other computational device. That is, processor 205 may be provided by any suitable logic circuitry for receiving inputs, processing them in accordance with instructions stored in memory and generating outputs (for example to the memory components or on display 240). In this embodiment, processor 205 may be a single core or multi-core processor with memory addressable space. In one example, processor 205 may be multi-core, comprising—for example—an 8 core CPU. In another example, it could be a cluster of CPU cores operating in parallel to accelerate computations.

[0035] In an embodiment of the disclosure, a Brain-Computer Interface (BCI) system comprising computer system 110 and mobile device 102 as previously illustrated in Figure 1 may be used to enhance a subject’s cognitive ability through an attention training process based on real-time feedback from the subject. In the proposed training system, a highly immersive attention training environment is displayed on display unit 130 whereby the environment is based on a galaxy-theme. The spatial-temporal environment shown on display 130 comprises an enticing multi-dimensional stereoscopic training environment to cultivate cognitive agility, spatial perception, and/or engagement.

[0036] In the displayed environment, the user may control the navigation of a virtual object, i.e., spaceship, using HCI signals from the user such as the user’s gaze direction and the user’s level of attention (i.e., cognitive scores) whereby a trained neural network 112 is used to compute the level of attention of a user based on the user’s EEG signals. It should be noted that the training system in this embodiment explicitly administers an imbedded relaxation routine and/or program during the attention training process, thus ensuring that the user is less fatigued and more motivated so that the user may be better engaged with the contents on the display for longer time. The engaging multi-dimensional galaxy environment shown on display 130 also

offers multiple levels of difficulty and multiplayer set-ups to cater for a wide range of training scenarios. Furthermore, the training system includes specific objectives aimed at enhancing sustained focus, selective attention, and the ability to resist distractions to foster a more engaging training experience whereby the training intensity is easily adjustable based on task performance and endurance, providing a flexible and personalized approach. The autonomous model used in the system also eliminates the need for individualized calibration exercises as it is not subject dependent.

[0037] In general, the proposed training system may be divided into two parts. In the initial stage, neural network 112 comprising a comprehensive Brain-Computer Interface (BCI) classification model is trained. During the training process, subjects are exposed to specially designed stimuli to evoke either heightened attention or relaxation, while their EEG data is simultaneously collected. To establish a subject-independent model that can be applied to any subject or user, two distinct cognitive tasks associated with sustained attention states and a lack of focus states (inattention) are employed. To elicit attention, a flanker's test with emoji-based stimuli is utilized, as depicted in Figure 3a. The first two rows 302 represent congruent stimuli set (all emojis are either sad or happy), while the other two rows 304 depict incongruent stimuli set (the middle emoji differs from the surrounding emojis). Subjects may respond to the stimuli using the left or right arrows on the keyboard, depending on whether the stimuli are congruent or incongruent, respectively, requiring focused attention. The subject's EEG signals are simultaneously collected and used to generate a training dataset of EEG signals associated with sustained attention cognitive states. For the inattention task, subjects are presented with a blank black screen and instructed to gaze around the border slowly to avoid fixating their attention on a specific point. The subject's EEG signals are simultaneously collected and used to generate a training dataset of EEG signals associated with inattention cognitive states.

[0038] In embodiments of the disclosure, the calibration data comprises EEG data from multiple sessions, e.g., 10 sessions, each consisting of six blocks, with each session lasting about three minutes and encompassing both training datasets (of EEG signals associated with sustained attention and inattention cognitive states) to generate multiple data recording trials. This is illustrated in Figure 3b. Neural network 112 is then trained using the recorded EEG signals from the two distinct training datasets so that trained neural network 112 may be used to detect the level of attention of the user. At the initial stage, MLP network 114 may be trained

using EEG signals features associated with alpha and beta bands. Trained MLP network 114 may then be used to detect a subject's level of fatigue based on the acquired EEG signals.

[0039] In the final stage of the proposed training system, the subject is presented with a highly immersive virtual display comprising a visual pathway embedded with spaceship control within a virtual solar system environment and is asked to control a spaceship to reach specified goals (visit selected planets in predefined order). Key processes performed by the system in the final stage of the proposed training system are illustrated in Figure 4. In this embodiment, the spaceship's speed of movement and movement direction may be controlled by the cognitive scores, or the attention level of the subject as decoded from the subject's EEG signals and also by using the subject's eye tracker outputs. In embodiments of the disclosure, the time duration for one session may be predetermined, e.g., 20 minutes, and a complete training cycle may comprise of multiple sessions. The system begins at step 402, by acquiring EEG signals and HCI signals from the subject.

[0040] At step 404, the cognitive scores of a subject or the subject's attention level may then be predicted by trained neural network 112 based on captured EEG signals. In embodiments of the disclosure, the cognitive scores may be computed at every predetermined period, e.g., 200 msec, for each EEG time segment, e.g., 4 seconds. The score generated by trained neural network 112 varies between '0' and '1', where a score of '0' represents a highly unfocused cognitive state and a score of '1' represents a highly focused cognitive state. In another embodiment of the disclosure, the score generated by trained neural network 112 may be presented as a probability score $S_{\text{attention}}$ which represents the probability of a cognitive attention state, where $S_{\text{attention}} = f(\text{seg}, \varphi)$ whereas seg indicates processed EEG segment and φ is the universal model parameters learnt from attention training process.

[0041] Concurrently at step 404, the subject's physiological scores or fatigue levels are continuously assessed using trained MLP network 114. In particular, trained MLP network 114 is configured to predict the subject's physiological scores or fatigue levels based on captured EEG signals. In the virtual display, the computed physiological scores or fatigue levels are monitored and mapped to the fuel level of spaceship.

[0042] Once the subject's physiological scores and cognitive scores have been computed, the system then determines at step 406, based on the computed physiological scores, if the subject is fatigued. If it is determined that the subject is fatigued, the subject will undergo a series of relaxation exercises at step 410. Conversely, if it is determined that the subject is not fatigued, the subject will then be directed to an attention training process at step 408.

[0043] The stimuli in the attention training process are designed to induce attention in the immersive environment and typically comprises a high dimensional spatial environment which interacts with the subject through the subject's cognitive scores (that are computed based on the subject's EEG signal), eye tracker signals and/or computer keyboard inputs (HCI signals). During the attention training process, four core attentional skills may be trained. The first attention skill comprises the subject's ability to stay focused on an important target without distraction (sustained attention), the second attention skill comprises the subject's ability to select and focus on particular input for further processing while simultaneously suppressing irrelevant or distracting information (selective attention), the third attention skill comprises the subject's ability to pay attention to two things at once (divided attention) and the fourth attention skill comprises the subject's ability to block potentially distracting information from the focus of attention or ability to control the subject's responses when conflicts are around (executive function). In embodiments of the disclosure, the objective of the immersive environment may be to visit multiple planets successively to distribute predefined target items as fast as possible, while avoiding all the obstacles and refilling fuel on timely basis on demand. The subject's performance may then be scored based on the amount of time taken to achieve the target, strategies adopted to accept/avoid objects and the subject's efficiency in gaining rewards and avoiding obstacles. Additionally, the subject will need to navigate the spaceship to the fuel station to top up fuel and this in turn initiates the relaxation process of the subject.

[0044] In embodiments of the disclosure, attention level (speed) and planet loci details (such as elevation and relative distance from other targets) may be displayed to help the subject have a deeper outlook of the training environment. The subject may then choose the difficulty level of the training, such as Basic, Hybrid or Intelligent. In this embodiment, a Basic level refers to a training sequence that uses eye tracker and keyboard strokes reaching only a few target locations without any hindrances on the way, a Hybrid level refers a training sequence that requires the subject to use EEG signals, keyboard and tracker with the target of reaching N

subsequent planets with a medium level of hurdles in between and an Intelligent level refers to a training sequence that integrates inputs from EEG signals and eye tracker, embedded with reward points and penalties, with the target of reaching more than N target locations while avoiding obstacles and enemies.

[0045] In an exemplary training sequence, N is defined as 4, and may be as high as 9 in higher levels of difficulty. Once the number of planets has been selected based on a difficulty level, the order at which consecutive targets are achieved may then be decided by the subject. The training environment is designed such that in the training framework, all planets are attached to an entrance checkpoint. The subject then has to go through the check points to touch the associated planet. Initially, before starting the spaceship, the time period left for each check point closure will be presented to the subjects and the subject has to use their eye gaze input to select the order of the interplanetary travel. The 4 target locations may be chosen based on the subject's attention level as estimated by trained neural network 112 and the subject's gaze and eye movements may be extracted from the eye tracker. As described above, it can be said that the digital display comprises a dynamically changing spatial-temporal display whereby the complexity of the spatial temporal display dynamically changes based on the EEG signals either to maintain certain attention levels or cognitive scores or to encourage the subject to achieve a predetermined target attention level or cognitive score.

[0046] Further, it should be noted that during the attention training process, trained MLP network 114 is continuously predicting the subject's physiological scores, or fatigue levels based on captured EEG signals. In embodiments of the disclosure, trained MLP network 114 generates physiological score values as an indicator of the subject's fatigue level, and these values may range from '0' to '1', where a '0' value indicates that the subject is not fatigued while a score of '1' indicates that the subject is highly fatigued, and these scores are computed every 200 msec for EEG signals that are captured in 4 sec time windows. One skilled in the art will recognize that these periods may be varied and are left as a design choice to one skilled in the art. Based on the fatigue level of the subject, the system will determine whether the relaxation process should be triggered.

[0047] If the relaxation process is to be triggered, the system then proceeds to step 410. At this step, the subject will undergo a series of relaxation exercises to restore the subject's mental

endurance and strength. In embodiments of the disclosure, the subject may be guided through a number of tasks such as mindful relaxation through audio-visual stimuli based deep breathing exercises, may be subjected to calming music and beautiful/calm images which all aim to refresh the cognitive skills and beneficial strategies of the subject so that when the subject undergoes the attention training process again, the subject would then be able to navigate through the processes faster. The length of the relaxation exercises may be fixed, e.g., 3 minutes or may be decided by the system. Additionally, the subject may select the specific exercise or combinations of exercises based on their requirements.

[0048] After the relaxation exercises have been completed, the system will repeat steps 402, 404 and 406 to determine if the subject is fatigued. If it is determined that the subject is not fatigued, the subject will then undergo the attention training process at step 408 as discussed above. Otherwise, if it is determined at step 406 that the subject is still fatigued, the subject will then undergo another set of relaxation exercises at step 410. In embodiments of the disclosure, regardless of whether the subject is determined to be “fatigued” or not, after step 410 has been repeated a number of times in succession, e.g., twice in succession, the subject will have to undergo the attention training process at step 408.

[0049] Steps 402 to 410 continuously repeat itself until a preset goal has been achieved, such as to navigate through a fixed number of planets while overcoming a number of obstacles, and the processes shown in Figure 4 will then end at step 412. In the proposed system, the goal selection is dynamically set based on performance indices achieved by the subject across ongoing sessions. For example, if the subject is taking significantly higher time period than the time allocated for a specific goal achievement (each goal is associated with a predefined time interval, for example, 3 min for Basic level training to reach 2 target locations), the system will recommend that the subject play the same level again. Also, if the subject finishes a specific level target within a time frame significantly lower time than the predefined time, the subject will be given an option to skip the very next level. This type of dynamic goal allocation based on subject’s performance indices is ideal as it optimizes the subject’s training and avoids training the subject unnecessarily. Factors such as sustainability of attention score, variation of fatigue score and frequency of requirement of relaxation mode activation can also control the goal selection to make the cognitive training more effective.

Neural Network

[0050] In embodiments of the disclosure, neural network 114 may include a multi-scale convolutional neural network that has a dynamic temporal layer configured to learn dynamic and frequency representations of all channels of EEG signals, an asymmetric spatial layer configured to compute global spatial kernels and hemisphere spatial kernels, a high-level fusion layer configured to fuse information from the global spatial and hemisphere spatial kernels to form fully connected layers that are activated by a SoftMax function. The detailed workings of this multi-scale convolutional neural network are described in this section.

[0051] EEG data may be viewed as a 2D time series, where one dimension represents the EEG channels (electrodes), and the other represents time, reflecting changes in brain activity over time. The spatial dimensions can show the patterns of brain activation in different functional areas (based on electrode placements). As illustrated in Figure 5, neural network utilizes modules comprising a dynamic temporal layer 502, an asymmetric spatial layer 504, a high-level fusion layer 506 to identify distinctive time-frequency-channel-specific EEG features associated with the user's emotional states and a classifier layer 508.

[0052] The dynamic temporal layer 502 employs multi-scale 1D convolutional kernels to capture more informative time-frequency representations and it does so by learning the dynamic temporal/frequency representations from the EEG data channel by channel. The asymmetric spatial layer 504 then leverages neuroscience findings, that indicate that brain activities in the right and left hemispheres are asymmetrically related to emotions or focused attention, to introduce a hemisphere kernel to model these asymmetries. It does so by obtaining the learning representations for each channel and applying them to learn the global spatial representations and the emotional or attentional asymmetry patterns using different scale convolutional kernels. The high-level fusion layer 506 then combines the learned representations from the hemisphere and global kernels or representations as obtained from the asymmetric spatial layer. The fused representations are then passed to classifier layer 508 having fully connected layers with SoftMax as the activation function

Dynamic Temporal Layer

[0053] The dynamic temporal layer consists of multi-scale 1D temporal kernels (T kernels). In order to enable the neural network to learn dynamic temporal representations, the lengths of the temporal kernels are set as the specific ratios of sampling rate f_s of the EEG signals. These ratios are defined as $\alpha^i \in \mathbb{R}$, where i is the level of the dynamic temporal layer and i will vary from 1 to L , if the dynamic temporal layer has L levels. Hence s_T^i , the size of T kernels in the i^{th} level, can be defined as:

$$s_T^i = (1, \alpha^i \cdot f_s), i \in [1, 2, 3] \quad \dots (1)$$

From the frequency perspective, the length of the T kernel is set as half the sampling rate, allowing for frequency information to be captured at 2 Hz and above. It is noted that activations related to attention/emotions are observed in Alpha (8-12 Hz), Beta (12-30 Hz), and Gamma (>30 Hz) bands. The temporal receptive field is then expanded by letting $L = 3$; $i = 1$ to 3, and $\alpha = 0.5$, as such, the ratio coefficients will become [0.5, 0.25, 0.125], learning diversified frequency representations.

[0054] It is hypothesized that the multiscale temporal kernels can enrich the learned dynamic frequency representations from EEG, providing more emotion related information. From the time perspective, multi-scale T kernels can capture long short-term temporal patterns and learn more diverse representations. The higher-level T kernel has a smaller ratio coefficient, which gives a shorter convolutional kernel length and vice versa. The long temporal kernel can learn long-term temporal and low-frequency diverse representations. The short kernel extracts short-term temporal and high-frequency representations. Let X denote EEG input samples where $X = [X^0, X^1, \dots, X^n]$, $X^n \in \mathbb{R}^{c \times l}$, where n is the number of EEG samples, c is the number of channels, and l is the length of each sample. The dynamic temporal representations can be generated by parallelly applying the multi-scale temporal kernels on the input EEG samples.

[0055] After *LeakyReLU*($\cdot$) activation function, the feature map is further down sampled by average pooling (AP). The reason for using average pooling is to reduce the effect of the noise as well as the feature dimension since EEG signals are of high dimensions with a low signal-noise ratio. Let $Z_{temporal}^i$ denote the output of the i -th level temporal kernel, $Z_{temporal}^i \in$

$\mathbb{R}^{n \times t \times c \times f_i}$, where n is the number of samples, t is the number of each level's T kernel, c is the number of channels, and f_i is the length of the feature after i -th level convolution operation. $Z_{temporal}^i$ is then defined as:

$$Z_{temporal}^i = AP(\phi_{L-ReLU}(ConvD(X, s_T^i))) \quad \dots(2)$$

where s_T^i is the T kernel size, X is the input EEG sample array, $Conv1D(\cdot)$ is the 1D convolution operation with the kernel size being s_T^i , step being (1,1), and $\Phi_{L-ReLU}(\cdot)$ is the LeakyReLU($\cdot$) activation function.

[0056] The output of each level's T kernel will be concatenated along the feature dimension. In order to reduce the internal covariate shift problems in neural networks, batch normalization was added after the dynamic temporal layer. Hence the final output of the dynamic temporal layer, $Z_T, Z_T \in \mathbb{R}^{n \times t \times c \times \sum f_i}$, is defined as:

$$Z_T = f_{bn}([Z_{temporal}^1, \dots, Z_{temporal}^i]), i \in [1, 2, 3] \quad \dots(3)$$

where f_{bn} is the batch normalization operation, and $[\cdot]$ stands for concatenation operation along the feature (f) dimension.

Asymmetric Spatial Layer

[0057] The asymmetric spatial layer has multi-scale 1D convolutional kernels whose sizes are related to the location of the EEG channels. There are two types of spatial kernels: global kernel and hemisphere kernel.

[0058] The global kernel has a size of $(c, 1)$, where c is the number of channels. Since the length of the kernel is the same as the channel dimension of the input EEG segment, it can learn the global spatial information. Additionally, the frontal area of brain emotional asymmetry is combined into the kernel design. The hemisphere kernel is used to extract the relations between the left and right hemispheres by sharing the convolutional kernels. The size of the hemisphere kernel is $(0.5 \cdot c, 1)$, and the step is $(0.5 \cdot c, 1)$, where c is the total number of channels. The hemisphere kernel is shared by two hemispheres without overlapping so that the asymmetric pattern can be extracted. The size of the spatial kernel s_S^j can be defined as:

$$s_S^j = (\delta^j \cdot c, 1), j \in [0,1] \quad \dots(4)$$

where $\delta = 0.5$ is the coefficient to control the ratio between the spatial kernel length and the total number of channels.

[0059] Let $\mathbf{Z}_{spatial}^j$ denote the output of the j -th type spatial kernel, $\mathbf{Z}_{spatial}^j \in \mathbb{R}^{n \times s \times c_j \times f}$, where n is the number of samples, s is the number of each type S kernel, c_j is the number of channels after j -th spatial convolution, and f is the length of the feature after each spatial convolution operation. $\mathbf{Z}_{spatial}^j$ may be defined as:

$$\mathbf{Z}_{spatial}^j = AP(\phi_{L-ReLU}(ConvD(\mathbf{Z}_T, s_S^j))) \quad \dots(5)$$

where s_S^j is the S kernel size, $\mathbf{Z}_T$ is the output of dynamic temporal layer, $ConvD(\cdot)$ is the 1D convolution operation with the kernel size being s_S^j , the step being (1,1) for the global kernels and $(0.5 \cdot c, 1)$ for the hemisphere kernels, and $\Phi_{L-ReLU}(\cdot)$ is the LeakyReLU($\cdot$) activation function.

[0060] In order to apply hemisphere kernels, the sequence of channels in the input EEG samples should be arranged in a particular way. The order of the channels should be $[channel_{left}, channel_{right}]$, where the $channel_{left}$ are the channels located in the left hemisphere, the $channel_{right}$ are the ones on the right hemisphere. The order for channels on each hemisphere should also be rearranged to make each kernel weight shared between pairs of symmetrically located electrodes on two hemispheres because the step of the hemisphere kernel is also $(0.5 \cdot c, 1)$. Figure 6 illustrates the electrode location map of a cap with 32 channels. As shown, the electrodes are divided into 3 groups: electrodes on the left hemisphere 602, electrodes on the right hemisphere 604, and the electrodes on the central line 606. The electrodes located on the central line of the head 606, F_z, C_z, P_z, and O_z, which cannot be paired with the left and right hemispheres, are removed to allow the neural network to learn the asymmetric pattern of left 602 and right 604 hemispheres better.

[0061] The final output of the asymmetric spatial layer, $\mathbf{Z}_S, \mathbf{Z}_S \in \mathbb{R}^{n \times s \times \sum c_j \times f}$ is defined as:

$$\mathbf{Z}_S = f_{bn}([\mathbf{Z}_{spatial}^0, \dots, \mathbf{Z}_{spatial}^j]), j \in [0,1] \quad \dots(6)$$

where f_{bn} is the batch normalization operation, and $[\cdot]$ stands for concatenation operation along the channel (c) dimension. The output of hemisphere kernel has a length of two in the spatial dimension, which refers to two hemispheres respectively. The output of global kernel is only a vector whose length in the channel dimension is one. After concatenation, the channel dimension is $\sum c_j = 3$.

High-level Fusion Layer

[0062] The high-level fusion layer is used to learn high-level spatial representations by fusing the learned information from global and hemispheres. Given the output of asymmetric spatial layer, $\mathbf{Z}_S \in \mathbb{R}^{n \times s \times 3 \times f}$, a 1D convolutional layer whose kernel size is (3, 1) is utilized to fuse the information along the spatial dimension. After LeakyReLU($\cdot$), average pooling, and batch normalization, a global average pooling layer (GAP) is added to overcome over-fitting and reduce the model size. The final learned global-hemisphere fusion representations will be generated by:

$$\mathbf{Z}_{fusion} = GAP(f_{bn}(AP(\Phi_{L-ReLU}(Conv1D(\mathbf{Z}_S, (3,1)))))) \quad \dots(7)$$

[0063] Finally, the latent representation of $\mathbf{Z}_{fusion}$ will be fed into fully connected layers. The final output layer is activated by the SoftMax function, $\Phi_{SoftMax}(\cdot)$. Hence the final output can be calculated by:

$$\mathbf{Output} = \Phi_{softmax}(\mathbf{W}' \Phi_{dp} \left(\Phi_{ReLU} \left(\mathbf{W} \left(\Gamma(\mathbf{Z}_{fusion}) \right) + b \right) \right) + b') \quad \dots(8)$$

where the $\Gamma(\cdot)$ is the squeeze operation, $\mathbf{W}$ and $\mathbf{W}'$ are the trainable weight matrix, b and b' are the bias terms. The structure of the entire neural network is shown in Table 1 below.

Model structure		Layers	Input	Output
Block1	3 branches (in parallel)	Conv2d, LK-ReLU, AP((1,8)) Kernel=15@(1, 64)	(-1, 1, 28, 512)	(-1, 15, 28, 56)
		Conv2d, LK-ReLU, AP((1,8)) Kernel=15@(1, 32)	(-1, 1, 28, 512)	(-1, 15, 28, 60)
		Conv2d, LK-ReLU, AP((1,8)) Kernel=15@(1, 16)	(-1, 1, 28, 512)	(-1, 15, 28, 62)
		Concatenate, BN		(-1, 15, 28, 178)
		Conv2d, LK-ReLU, AP((1,2)) Kernel=15@(28, 1)	(-1, 15, 28, 178)	(-1, 15, 1, 89)
Block2	2 branches (in parallel)	Conv2d, LK-ReLU, AP((1,2)) Kernel=15@(14, 1), Stride=(14, 1)	(-1, 15, 28, 178)	(-1, 15, 2, 89)
		Concatenate, BN		(-1, 15, 3, 89)
		Conv2d, LK-ReLU, AP((1,4)), BN, GAP Kernel=15@(3, 1)	(-1, 15, 3, 89)	(-1, 15, 1)
Block3		Flatten	(-1, 15, 1)	(-1, 15,)
		Linear(32), ReLU	(-1, 15,)	(-1, 32,)
		dropout(0.5)	(-1, 32,)	(-1, 32,)
		Linear(2)	(-1, 32,)	(-1, 2,)
Fully connected layers		softmax	(-1, 2,)	(-1, 2,)

TABLE 1

[0064] In Table 1, LK-ReLU represents the Leaky-ReLU activation function, AP represents the average pooling operation, BN stands for batch normalization, GAP represents the global average pooling, and ‘-1’ in the tensor size stands for the number of samples within one mini-batch. Additionally, the strides of CNNs are (1, 1) if not specified, and the one for pooling layers is the same as the pooling step.

[0065] When the neural network was trained using EEG signals associated with sustained attention cognitive states and EEG signals associated with inattention cognitive states as described in the previous sections, it was determined that the trained neural network was able to distinguish the attention and non-attention mental states of subjects, delivering a mean classification accuracy of 80% across 50 healthy subjects. This is illustrated in the histogram plotted in Figure 7.

Multi-Layer Perceptron (MLP) network

[0066] Figure 8 illustrates a multilayer perceptron neural network structure in accordance with embodiments of the present disclosure. A MLP network represents a generalization of the single layer perceptron where a single layer perceptron forms half-plane decision regions, while a multi-layer perceptron is able to form arbitrarily complex decision regions and may separate various input patterns.

[0067] In general, an MLP network comprises an input layer, a hidden layer, an output layer and a bias layer. The input layer comprises a layer of neurons that receives information from external sources and passes this information to the network for processing. The hidden layer comprises a layer of neurons that receive information from the input layer and processes them in a hidden way whereby it has no direct connection with other layers within the system. The output layer comprises a layer of neurons that receive processed information and send output signal out of the system. The bias layer acts on the neuron like an offset whereby the function of the bias is to provide a threshold for the activation of neurons and the bias input is connected to each of the hidden and output neurons in a network. The connection weights in the MLP network are then learnt from available training patterns and its performance is gradually improved over time by iteratively updating the weights in the network. In embodiments of the disclosure, the MLP network was trained using EEG signal features associated with alpha and beta bands. The training datasets used to train the MLP network may be obtained from Z. Cao, C.-H. Chuang, J.-K. King, and C.-T. Lin, "Multi-channel EEG recordings during a sustained-attention driving task," *Sci. Data*, vol. 6, no. 1, pp. 1–8, Apr. 2019.

[0068] A process for modifying presentation of digital content on a display based on raw EEG signals in accordance with embodiments of the present disclosure is illustrated in Figure 9 whereby process 900 may be carried out by modules contained within computer system 110.

[0069] Process 900 beings at step 902 with the system receiving the EEG signals and HCI signals. Process 900 then utilizes the trained neural network and the train MLP network to generate cognitive scores and physiological scores respectively. At step 906, the generated scores are then provided to a display controller module. Process 900 then utilizes the cognitive and/or the physiological scores to modify the presentation of digital content on the display. Process 900 then ends.

[0070] In embodiments of the disclosure, during the modification of the presentation of the digital content on the display based on the cognitive and the physiological scores, process 900 may display, using the display controller module, a first digital content on the display in response to a determination that averaged physiological scores over a cluster of time-windows of the raw EEG signals have exceeded a predetermined threshold score. In further embodiments, the first digital content may comprise a relaxation program.

[0071] In embodiments of the disclosure, during the modification of the presentation of the digital content on the display based on the cognitive and the physiological scores, process 900 may display, using the display controller module, a second digital content on the display in response to a determination that a first time-period has lapsed. In further embodiments, the second digital content comprises an attention training program.

[0072] In embodiments of the disclosure, before the cognitive scores are generated using the trained neural network, process 900 may filter the raw EEG signals provided to the trained neural network to remove low and high frequency noises; and may remove artifacts from the filtered EEG signals by subtracting smoothed references from the filtered EEG signals, wherein the smoothed references are obtained using a moving averaging technique.

[0073] In embodiments of the disclosure, the raw EEG signals may be acquired from a user using a mobile bio-sensor device communicatively connected to the computer system.

[0074] In embodiments of the disclosure, during the modification of the presentation of the digital content on the display based on the cognitive and the physiological scores, process 900 may display, using the display controller module, a dynamically changing spatial-temporal display on the display whereby a complexity of the spatial-temporal display dynamically changes to maintain the cognitive scores at a constant level.

[0075] In embodiments of the disclosure, where the raw EEG signals are acquired from a user using a mobile bio-sensor device communicatively connected to the computer system, during the modification of the presentation of the digital content on the display based on the cognitive and the physiological scores, process 900 may display using the display controller module, a spatial-temporal activity on the display; and instruct, using the mobile bio-sensor device, the user to continuously engage in the displayed spatial-temporal activity until it is determined that averaged cognitive scores over a cluster of time-windows of the raw EEG signals have achieved a predetermined target.

[0076] In embodiments of the disclosure, where the raw EEG signals are acquired from a user using a mobile bio-sensor device communicatively connected to the computer system, during the modification of the presentation of the digital content on the display based on the cognitive and the physiological scores, process 900 may display using the display controller

module, a spatial-temporal activity on the display; and instruct, using the mobile bio-sensor device, the user to continuously engage in the displayed spatial-temporal activity until it is determined that averaged physiological scores over a cluster of time-windows of the raw EEG signals have exceeded a predetermined threshold score.

[0077] In embodiments of the disclosure, process 900 may obtain, using the mobile bio-sensor device, human-computer interaction (HCI) signals from the user; provide the HCI signals to the display controller module; and modify, using the display controller module, the presentation of the digital content on the display based on the cognitive and the physiological scores and the HCI signals.

Example of an Attention Training System

[0078] An exemplary cognitive training environment/ program/ game is disclosed in this section. In this example, the program comprises a third person camera view 3D scene. The speed of the ship in this program is calculated using an acceleration curve where the higher the attention score, the higher the curve acceleration curve of the spaceship. The real time 4-channel EEG signal segment is transformed into corresponding attention scores predicted by the trained neural network and the score varies from 0 to 1, representing highly unfocused state to highly focused state respectively.

[0079] The nose direction of the ship is dependent on the Gaze Point, the coordinate point at which the user is looking. If the spaceship has to divert directions towards the left, right, up or down, the user's eye gaze has to be shifted from a central point towards the left, the right, upwards and downwards respectively.

[0080] The ship is also provided with a fuel bar which indicates the fuel level, and the indicator value changes according to the fuel level. As the ship travels, its fuel reduces. To provide the user with extra fuel, the ship has to be navigated through a green ring like stone structure and when the user enters this ring, the active scenes being displayed will be paused and a cutscene of instructions managed by the entire relaxation portion of the training will then be displayed. The platform is designed to be capable of directed relaxation to aid users in reducing mental fatigue. During relaxation exercise period, the subject will perform deep breathing which activates subject's brain functioning for further program experience.

[0081] Distractions are provided in the training theme, such as asteroid belts, rocks and fog that exist in the visual space of the player's spaceship. If the player's spaceship happens to hit the asteroid or immovable space objects, the spaceship will be destroyed. Hence extreme care has to be taken to avoid them as efficiently as possible. The training program is designed such that the player has to use his sustained attention, selective attention and divided attention to expel the distractors as fast as possible.

[0082] For example, if a rock appears on the path travelling towards the ship, the player has to improve his focused attention such that ship moves faster and escapes from the line of sight. If two asteroids come towards the ship at different speeds, the player has to decide which asteroid should be destroyed first based on its speed. In other words, the fastest moving asteroid would have to be eliminated first followed by the other. The action sequence taken by the player and the time amount of time taken will be evaluated by the program to track the player's level of divided attention. If the program determines that the player's sustained attention level is significantly above a certain threshold value α , the player's spaceship will gain a better shield that may directly destroy any asteroids/rocks, thereby boosting the power play of the spaceship.

[0083] During the gameplay, a counter is used to display the amount of time of the entire mission, e.g., the counter displays the minutes and seconds of the game time. However, this counter will be inactivated when relaxation mode is activated by the player and this counter will resume when the player returns to the training program. Data for each mission is taken from the counter and stored in a separate output file. During the goal directed navigation, fundamental skills related to executive function include proficiency in adaptable thinking, planning, self-monitoring, self-control, time management, and organization will also be circumlocutorily trained. A 2D map for planning the interplanetary travel mission and checking the travel status, and higher dimension perspective view for training may also be made available for subject's information.

[0084] Numerous other changes, substitutions, variations, and modifications may be ascertained by the skilled in the art and it is intended that the present application encompass all such changes, substitutions, variations, and modifications as falling within the scope of the appended claims.

CLAIMS:

1. A computer system configured to modify presentation of digital content on a display based on acquired raw electroencephalogram (EEG) signals, the computer system comprising:
a computer processor; and
a non-transitory media readable by the processing unit, the media storing instructions that when executed by the processing unit, causes the processing unit to:
provide the raw EEG signals as inputs to a trained neural network and a trained multilayer perceptron (MLP) network;
generate, using the trained neural network, cognitive scores, wherein the neural network was trained using EEG signals associated with sustained attention cognitive states and EEG signals associated with inattention cognitive states;
generate, using the trained MLP network, physiological scores for each time-window of the raw EEG signals, wherein the MLP network was trained using EEG signal features associated with alpha and beta bands;
provide the cognitive and the physiological scores to a display controller module;
and
modify, using the display controller module, the presentation of the digital content on the display based on the cognitive and the physiological scores.
2. The computer system according to claim 1, whereby the instructions to modify the presentation of the digital content on the display based on the cognitive and the physiological scores further comprises instructions for directing the processing unit to:
display, using the display controller module, a first digital content on the display in response to a determination that averaged physiological scores over a cluster of time-windows of the raw EEG signals have exceeded a predetermined threshold score.
3. The computer system according to claim 2, wherein the first digital content comprises a relaxation program.
4. The computer system according to any one of claims 1 to 3, whereby the instructions to modify the presentation of the digital content on the display based on the cognitive and the physiological scores further comprises instructions for directing the processing unit to:

display, using the display controller module, a second digital content on the display in response to a determination that a first time-period has lapsed.

5. The computer system according to claim 4, wherein the second digital content comprises an attention training program.
6. The computer system according to claim 1, whereby before the instructions to generate cognitive scores using the trained neural network, the non-transitory media further comprises instructions for directing the processing unit to:
 - filter the raw EEG signals provided to the trained neural network to remove low and high frequency noises; and
 - remove artifacts from the filtered EEG signals by subtracting smoothed references from the filtered EEG signals, wherein the smoothed references are obtained using a moving averaging technique.
7. The computer system according to claim 1, whereby the raw EEG signals are acquired from a user using a mobile bio-sensor device communicatively connected to the computer system.
8. The computer system according to any one of claims 1 to 7, whereby the instructions to modify the presentation of the digital content on the display based on the cognitive and the physiological scores further comprises instructions for directing the processing unit to:
 - display, using the display controller module, a dynamically changing spatial-temporal display on the display whereby a complexity of the spatial-temporal display dynamically changes to maintain the cognitive scores at a constant level.
9. The computer system according to claim 1, whereby the raw EEG signals are acquired from a user using a mobile bio-sensor device communicatively connected to the computer system, and whereby the instructions to modify the presentation of the digital content on the display based on the cognitive and the physiological scores further comprises instructions for directing the processing unit to:
 - display, using the display controller module, a spatial-temporal activity on the display; and

instruct, using the mobile bio-sensor device, the user to continuously engage in the displayed spatial-temporal activity until it is determined that averaged cognitive scores over a cluster of time-windows of the raw EEG signals have achieved a predetermined target.

10. The computer system according to claim 1, whereby the raw EEG signals are acquired from a user using a mobile bio-sensor device communicatively connected to the computer system, and whereby the instructions to modify the presentation of the digital content on the display based on the cognitive and the physiological scores further comprises instructions for directing the processing unit to:

display, using the display controller module, a spatial-temporal activity on the display; and

instruct, using the mobile bio-sensor device, the user to continuously engage in the displayed spatial-temporal activity until it is determined that averaged physiological scores over a cluster of time-windows of the raw EEG signals have exceeded a predetermined threshold score.

11. The computer system according to claim 7 whereby the non-transitory media further comprises instructions for directing the processor to:

obtain, using the mobile bio-sensor device, human-computer interaction (HCI) signals from the user;

provide the HCI signals to the display controller module; and

modify, using the display controller module, the presentation of the digital content on the display based on the cognitive and the physiological scores and the HCI signals.

12. The computer system according to claim 1 wherein the neural network comprises a multi-scale convolutional neural network that includes a dynamic temporal layer configured to learn dynamic and frequency representations of all channels of EEG signals, an asymmetric spatial layer configured to compute global spatial kernels and hemisphere spatial kernels, a high-level fusion layer configured to fuse information from the global spatial and hemisphere spatial kernels to form fully connected layers that are activated by a softmax function.

13. A method to modify presentation of digital content on a display based on acquired raw electroencephalogram (EEG) signals, the method comprising:

receiving, using a machine learning module, the raw EEG signals and provided the received raw EEG signals as inputs to a trained neural network and a trained multilayer perceptron (MLP) network;

generating, using the trained neural network, cognitive scores, wherein the neural network was trained using EEG signals associated with sustained attention cognitive states and EEG signals associated with inattention cognitive states;

generating, using the trained MLP network, physiological scores for each time-window of the raw EEG signals, wherein the MLP network was trained using EEG signal features associated with alpha and beta bands;

receiving, using a display controller module, the cognitive and the physiological scores; and

modifying, using the display controller module, the presentation of the digital content on the display based on the cognitive and the physiological scores.

14. The method according to claim 13, whereby the modifying of the presentation of the digital content on the display based on the cognitive and the physiological scores further comprises the steps of:

displaying, using the display controller module, a first digital content on the display in response to a determination that averaged physiological scores over a cluster of time-windows of the raw EEG signals have exceeded a predetermined threshold score.

15. The method according to claim 14, wherein the first digital content comprises a relaxation program.

16. The method according to any one of claims 13 to 15, whereby the modifying of the presentation of the digital content on the display based on the cognitive and the physiological scores further comprises the steps of:

displaying, using the display controller module, a second digital content on the display in response to a determination that a first time-period has lapsed.

17. The method according to claim 16, wherein the second digital content comprises an attention training program.

18. The method according to claim 13, whereby before the step of generating the cognitive scores using the trained neural network, the method further comprises the steps of:
- filtering, using the machine learning module, the raw EEG signals provided to the trained neural network to remove low and high frequency noises; and
 - removing artifacts from the filtered EEG signals by subtracting smoothed references from the filtered EEG signals, wherein the smoothed references are obtained using a moving averaging technique.
19. The method according to claim 13, whereby the raw EEG signals are acquired from a user using a mobile bio-sensor device communicatively connected to the computer system.
20. The method according to any one of claims 13 to 19, whereby the modifying of the presentation of the digital content on the display based on the cognitive and the physiological scores further comprises the steps of:
- displaying, using the display controller module, a dynamically changing spatial-temporal display on the display whereby a complexity of the spatial-temporal display dynamically changes to maintain the cognitive scores at a constant level.
21. The method according to claim 13, whereby the raw EEG signals are acquired from a user using a mobile bio-sensor device communicatively connected to the computer system, and whereby the modifying the presentation of the digital content on the display based on the cognitive and the physiological scores further comprises the steps of:
- displaying, using the display controller module, a spatial-temporal activity on the display; and
 - instructing, using the mobile bio-sensor device, the user to continuously engage in the displayed spatial-temporal activity until it is determined that averaged cognitive scores over a cluster of time-windows of the raw EEG signals have achieved a predetermined target.
22. The method according to claim 13, whereby the raw EEG signals are acquired from a user using a mobile bio-sensor device communicatively connected to the computer system, and whereby the modifying of the presentation of the digital content on the display based on the cognitive and the physiological scores further comprises the steps of:

displaying, using the display controller module, a spatial-temporal activity on the display; and

instructing, using the mobile bio-sensor device, the user to continuously engage in the displayed spatial-temporal activity until it is determined that averaged physiological scores over a cluster of time-windows of the raw EEG signals have exceeded a predetermined threshold score.

23. The method according to claim 19 whereby the method further comprises the steps of:

obtaining, using the mobile bio-sensor device, human-computer interaction (HCI) signals from the user;

providing the HCI signals to the display controller module; and

modifying, using the display controller module, the presentation of the digital content on the display based on the cognitive and the physiological scores and the HCI signals.

24. The method according to claim 13 wherein the neural network comprises a multi-scale convolutional neural network that includes a dynamic temporal layer configured to learn dynamic and frequency representations of all channels of EEG signals, an asymmetric spatial layer configured to compute global spatial kernels and hemisphere spatial kernels, a high-level fusion layer configured to fuse information from the global spatial and hemisphere spatial kernels to form fully connected layers that are activated by a softmax function.