Research Article

Using AI-Based Classification Techniques to Process EEG Data Collected during the Visual Short-Term Memory Assessment

Milos Antonijevic [iD],[1] **Miodrag Zivkovic,**[1] **Sladjana Arsic,**[2] **and Aleksandar Jevremovic**[1]

[1]*Informatics and Computing Department, Singidunum University, Belgrade, Serbia*
[2]*Department Cupria, Academy of Educational Medical Professional Studies, Krusevac, Serbia*

Correspondence should be addressed to Milos Antonijevic; mantonijevic@singidunum.ac.rs

Visual short-term memory (VSTM) is defined as the ability to remember a small amount of visual information, such as colors and shapes, during a short period of time. VSTM is a part of short-term memory, which can hold information up to 30 seconds. In this paper, we present the results of research where we classified the data gathered by using an electroencephalogram (EEG) during a VSTM experiment. The experiment was performed with 12 participants that were required to remember as many details as possible from the two images, displayed for 1 minute. The first assessment was done in an isolated environment, while the second assessment was done in front of the other participants, in order to increase the stress of the examinee. The classification of the EEG data was done by using four algorithms: Naive Bayes, support vector, KNN, and random forest. The results obtained show that AI-based classification could be successfully used in the proposed way, since we were able to correctly classify the order of the images presented 90.12% of the time and type of the displayed image 90.51% of the time.

1. Introduction

Visual short-term memory (VSTM) is defined as the ability to remember a small amount of visual information, such as colors and shapes, during a short period of time [1]. There are many different tests designed to determine properties of VSTM, such as the capacity of VSTM, the time the subject is able to retain remembered information, and the influence of different external factors. VSTM is a part of short-term memory (STM). The information that is kept in VSTM can be processed further through working memory, it can be converted to long-term memory, or it can simply be forgotten. Short-term memory has two main characteristics: limited capacity and limited time.

The capacity of the short-term memory is limited to seven elements (plus or minus two elements) [2]. Short-term memory capacity is almost constant in a way that different people can remember more or less the same number of elements. Element remembering skills also depend on other factors, such as the length of the words, feeling associated with the stimulus, and other personal factors. STM can hold information up to 30 seconds. However, this information can be remembered if it is repetitive or sense is given to it.

Working memory or operative memory is a set of processes that allows us to keep and manipulate temporary data and perform complex cognitive activities. Working memory is a type of STM. Visual or audio material that is received by the brain is stored temporarily, but it is actively manipulated. Both processes, storing and manipulation, are integrated through consciously directed attention [3].

The multicomponent model of working memory was introduced by Baddeley and Hitch [4, 5]. The latest version of the model [1] consists of three systems, which include components for keeping and processing information. The first system is a central executive system, which acts as a monitoring system, and it is responsible for directing attention to relevant information. It is also responsible for the coordination of other "slave" subsystems and organization of activities needed to perform some action. The second system is the phonological loop, which acts as a slave subsystem. It is responsible for the management and storing of verbal and written material in memory. The third system

of interest is the visuospatial sketchpad, which is responsible for the management and storing of visual and spatial information. The visuospatial sketchpad is responsible for VSTM.

VSTM enables storing of received visual information and also later usage of remembered visual information. VSTM is very important for normal functioning of cognitive abilities and for performing everyday activities. Any damage to VSTM may reduce the amount of information and time that a person is able to retain it.

Neuropsychological assessment enables the testing and assessment of VSTM. The most frequently used practices include the classic direct and indirect numbers tests from Wechsler's scale; the NEPSY test by Korkman, Kirk, and Kemp (from 1998); continuous performance test (CPT); memory malingering test (TOMM); visual organization test (VOT); test of variables of attention (TOVA); and Tower of London test. These tests measure not only visual short-term memory but also short-term memory, reaction speed, working memory, visual scanning, perception of the environment, remembering of the context, naming, distinguishing, and speed of data processing [6].

VISMEM is a very important and frequently used test, and it was created using the classic TOMM [7]. The test consists of showing an image to the subject for a limited period of time, typically around 60–70 seconds. During this time, the subject has a task to look at the image and memorize the context and remember as many details from the image as possible. After the given time expires, the image is removed, and the subject gives answers to different questions about the image [8].

The main goal of this research is to determine the possible correlation between a participant's emotional state while doing the visual short-term working memory test and the test results he/she achieved. In other words, we tried to determine if it is possible to predict the result of the testing and to which extent, based on the measured emotional states of the subjects. The secondary objective is to determine the influence of external factors, more precisely the audience that is present during the testing on the participant's emotional state and achieved test results. We expected that stress would be increased in the case where the audience is present and when the actions and results of participants are transparent and visible to the audience.

We developed a custom Web application, which was used to perform the experiment. In the first step, the application shows a certain image (Image A or Image B) for a limited time of 60 seconds (during this interval, the subject is trying to remember as many details as possible). After the given time expires, the image is removed, and the application automatically diverts subjects to the questions related to the image content.

Every participant in the experiment was measured twice. The first assessment (first image) was done in an isolated environment, while the second assessment (second image) was done in front of the other participants, in order to increase the stress of the examinee. Data about the mental state of the participant was gathered during the duration of the experiment in both cases.

In order to get insight into the mental and emotional states of the participants, we used an electroencephalography (EEG) device, which recorded EEG states during the experiment in a time synchronized manner. We developed an EEG device application for processing raw signals called *MyEmotivator* to extract six emotional states: interest, excitement, engagement, stress, relaxation, and focus. Data about actions of the subject from the Web application and mental states of the subject recorded by the EEG device were synchronized by using the Human-Computer Interaction Monitoring and Analytics Platform (HCI-MAP) [9], with an expected error rate of <0.001 s.

The data collected during the experiment (EEG data and data about the actions within the application—image presented, image hidden, question presented, etc.) were analyzed by using four classification algorithms: Naive Bayes, support vector, K-nearest neighbors (KNN), and random forest. In this paper, we classified data using three different classes: (1) order of displaying the image (image displayed first or second (i.e., with or without the audience)), (2) type of the image (Image A or Image B), and (3) correctness of the answer (image, correct, incorrect).

There were several conclusions that could be drawn from the classification results. First, the classification results showed that, by using the emotional states of the participants and question duration interval, we could determine with the best accuracy of 90.12% if the participant was in the presence of the audience when answering the question. Second, results obtained from quantitative analysis of the emotional states of the participants enabled us to determine with the accuracy of 90.51% which image (Image A or Image B, regardless if it was displayed first or second in a row) the participant was viewing before answering the question. However, we did not achieve any significant classification results regarding the correctness of the answer.

2. Related Work

The EMOTIV EPOC+ device was used in the study. The device is capable of isolating a P300 low-voltage signal (2–5 μV), which is considered to be associated with a stimulus evaluation or categorization [10, 11]. Its low power as compared to an EEG means that the device can distinguish this signal from the background noise that occurs during the measurement. The overall conclusion is that this device can be used as a reliable brain-computer interface (BCI).

In a previous study [12], the authors investigated the emotional states of students, which was represented in the form of frustration and excitement that occurred as a result of feedback information gathered from intelligent tutoring systems (ITSs). By analyzing the obtained data, the authors developed a system that enabled student emotions to be anticipated and the feedback information to be modified accordingly.

Similarly, emotional reactions to different visual stimuli were examined in two independent works [13, 14]. The first of these studied the ability to recognize EEG patterns in a state of relaxation and while imagining two different types of pictures (faces and houses), and an accuracy of 48% was

achieved. In the second work, the authors developed a method of interpreting EEG values in order to discriminate between mental patterns when participants observed pleasant and unpleasant pictures as compared to neutral content.

The goal of the experiment described in the work by Esfahani and Sundararajan [15] was the detection of the level of pleasure. The authors also tested a method for correcting the robot's behavior in order to increase the pleasure level. Although the experiment was carried out on a small number of participants (four males), a correct classification was achieved in an average of 79.2% of cases.

In one of our papers, we described some steps towards applying artificial intelligence and EEG signals for the improvement of electronic assessments [16]. The first analysis pointed to the possibility of using certain question types in electronic tests in order to influence the psychological states of students during assessments. For example, by inserting "funny" questions with one obvious correct answer, this system can decrease the stress of the students. Furthermore, the most interesting questions are the easy ones, while the focus of the student can be increased by using "impossible" questions with no correct answers.

If a situation is a very important one for the person, there is rarely just one sentiment or one tendency for action or behavior. Usually, there are multiple emotions, either happening in parallel or sequentially one after another. Therefore, it is possible that stress and relaxation, although adverse feelings, reach the same or similar levels in one moment. When the subject observes an image with multiple details, it is required to engage both mentally and emotionally in order to remember as many details as possible. As a result, stress is increasing in this situation. This kind of stress is useful, as it helps in achieving goals and increases efficiency during the engagement; it is called eustress [17].

According to one study [18], emotional images with complex scenes have a different effect on the visual short-term memory as compared to neutral images with objects. Stress is always present, but it has a higher level in the case of complex emotional images. The difference in visual complexity of the observed image also affects remembering efficiency—complex emotional images are remembered better and with more details than neutral images. Emotional excitement increases the activity of the amygdala and hippocampus, which in turn can increase remembering efficiency.

Another study [19] dealt with research on how visual information presented in different time intervals affects the precision and reliability of reproduction during VSTM evaluation. Results have shown that the best reproduction efficiency is after the first evaluation of VSTM. The study has also shown that, in the case of applying two or more VSTM tests, reproduction efficiency is better if the evaluation is performed in short time intervals as compared to longer time intervals (several hours or more).

In one study [20], the authors measured two cognitive skills, focused attention and working memory, using a wearable EEG device. By training several different classifiers to predict three levels (low, medium, and high) of mentioned skills, they were able to obtain an accuracy of 84% and 81% for the focused attention and working memory, respectively.

3. Applications and Sensors

This research of human-computer interaction was based on using an EEG device sensor. The EEG is a noninvasive method of tracking changes in electrical voltage of brain neurons during a defined time interval. Besides medical applications (e.g., epilepsy diagnostics), collected data can be used in the research of human brain reactions to specific events that occur during some defined time frame. The EMOTIV EPOC + device [21] was used for measuring the variability of emotional characteristics of the subjects, depending on the changes in the surrounding environment. The EMOTIV EPOC+ is a wireless EEG device with 14 channels designed for measuring the activities of the cerebral cortex. Access to the raw EEG data makes this device applicable in developing BCI applications. The manufacturer has developed the algorithms for extracting values of six emotional states (interest, engagement, excitement, stress, relaxation, and focus) from raw EEG data that we used in this research [22].

Although there are some similar solutions (e.g., Lab Streaming Layer https://github.com/sccn/labstreaminglayer) for collecting and synchronizing data from different sensor devices and client applications, in this research, we used the HCI-MAP [9]. We developed a separate application for each sensor with the possibility to send data to the platform over the HTTP(S) protocol and the HCI-MAP API. Each application was implemented as a Web application that uses the same interface for sending data (Figure 1).

One of the main challenges when using multiple sensors is an aggregation of collected data [23]. In the case of complex experiments that are conducted in a distributed environment, time synchronization of different sensors is critical in order to have valid measurements and correct fusion of collected data on the remote server. Some of the reasons for shifting data processing to a remote server(s) are not only the existence of more participants in experiments but also the need for fast processing of collected data and returning results to the main application in the form of generated feedback.

3.1. VSTM Application. VSTM application has been implemented as a modern interactive Web application. Technologies used include HTML5, CSS, and JavaScript. The application contains three main sections: initialization screen, a screen with the image to be memorized, and questions. The initialization screen is displayed while the VSTM application is loading, and during this process, time synchronization with the HCI-MAP server takes place. Time synchronization is crucial for aggregation of application data with data gathered from sensors. The test's main screen with the image is given in Figure 2.

The subject had limited time to remember as many details from the image as possible. After the available time expired, the image was removed and questions were shown. The screen with an example question is shown in Figure 3.

Questions had been selected to exercise VSTM and to verify how many details the subject had remembered from the image. The subject had limited time to answer the questions, which is shown at the top of the screen. It was required
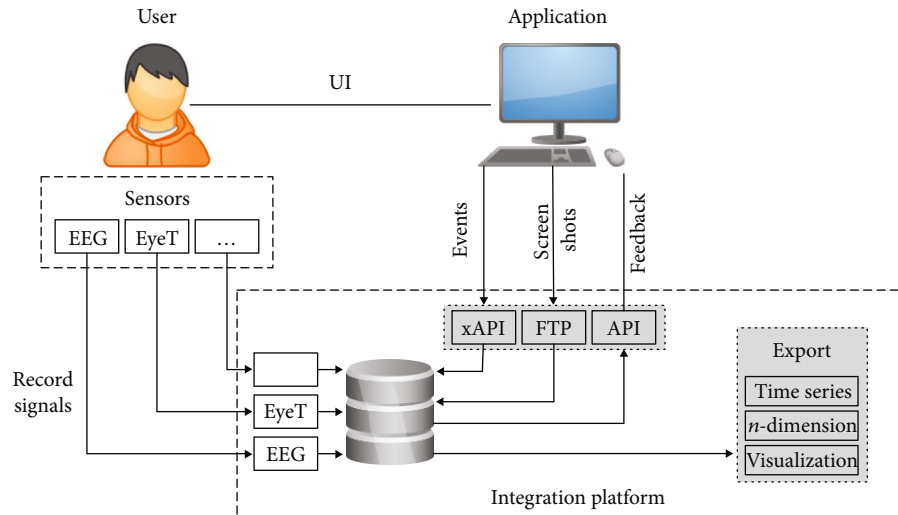
Figure 1: The basic architecture of HCI-MAP application.

(a)
(b)

Figure 2: The main screen of VSTM application, with Image A and Image B.
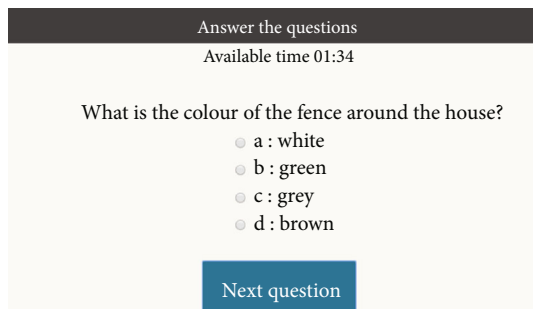

Figure 3: Screen with a question in VSTM application.

that the subject had to answer the question before moving to the next one (by clicking on the "Next Question" button), and it was not possible to go back to the previous question.

The initial test results were displayed immediately after the test was finished, in one of two possible ways. After the subject answered all the questions, or if the timer expired, results were calculated and the number of correct answers was shown to the user. The whole session was recorded on the HCI-MAP and could be exported as a CSV file for a more detailed quantitative/qualitative analysis. Each session had several types of events that were triggered by the Web application and recorded the session onto the HCI-MAP. These events include the following:

(i) *image_presented*: event is triggered when the image is shown to the subject

(ii) *image_timeout*: event is triggered if the available time expires

(iii) *image_removed*: event is triggered when the image is not visible to the subject anymore

(iv) *quiz_started*: event is triggered when the quiz has been started

(v) *question_shown*: event is triggered when a new question is shown to the subject. Since there is no option to go back to the previous question, this event means that the user has answered the question and went to the next one

(vi) *quiz_completed*: event is triggered when the subject answers all available questions

After the *quiz_completed* event, the complete list of the subject's answers was sent to the HCI-MAP. With the data received from sensors, it was possible to do further analysis. For example, by collecting eye tracking data, it was possible to create a heat map, which showed regions of the image where the user spent most of the time looking in red color. It is possible to further analyze the heat map and to see the

correlation between where the user was looking and correct/-wrong answer ratio.

3.2. MyEmotivator. *MyEmotivator* application was developed with a goal to record and display six emotional states (interest, engagement, excitement, stress, relaxation, and focus) in real time by using the EMOTIV EPOC+ interface (Figure 4.). The application was used to measure the variability of emotional characteristics in subjects, depending on the changes in the environment. Our goal was to obtain experimental data to get insight and perform later analysis of the impact that different external factors had on the emotional states of subjects.

The application supports two methods for saving data: locally on a measuring device and remotely by sending collected data to the server (Figure 4, Section 4). Selection of the desired EPOC+ device would initiate connecting the device with the application, which usually takes between 1 and 2 seconds in regular conditions. Three levels of signal quality are defined: no signal, bad signal, and good signal. Based on the signal levels, connectors would be marked with different colors (red, orange, or green) on the main screen (Figure 4, Section 2).

The default sampling frequency of each signal was two times in a second (500 ms). Each of the six sampled values (interest, engagement, excitement, stress, relaxation, and focus) was in the range 0–100, where 100 is a maximum level of emotion for a given user and 0 is the theoretical minimum.

3.3. HCI-MAP. The HCI-MAP was used for synchronization of gathered data from client applications and various sensors, data aggregation and processing in real time, and returning of obtained results in suitable formats for further analysis by computer or interpretation by humans (Figure 5). Currently, the supported sensors are EEG, eye tracking, facial emotion recognition, and mouse tracking sensors. However, connections to other sensors can be easily implemented through the open platform interface. Besides the sensors, using the same interface, a platform can receive information from user applications. Data can be exported not only from the platform as a time series (in CVS format) but also as more complex reports (e.g., recording of the user interface with eye position visualization in the form of a heat map).

One of the main challenges is the time synchronization of collected data. For this purpose, a network time protocol (NTP) was developed [24], which enables time synchronization of data collected from different sources in an environment where there is a possible latency in data transfer. When using the HCI-MAP, all sensors are time synchronized with the possible error in the range from -0.5 to +0.5 ms.

The HCI-MAP uses the TCP/IP network stack, and it communicates with sensors by using the HTTP(S) protocol. Communication with some sensors is direct. For example, software that monitors the mouse cursor position is realized as an application that runs on the client's computer. On the other hand, some sensors communicate with the platform indirectly. For example, software that delivers data from the EEG sensor to the platform runs on an Android tablet, which
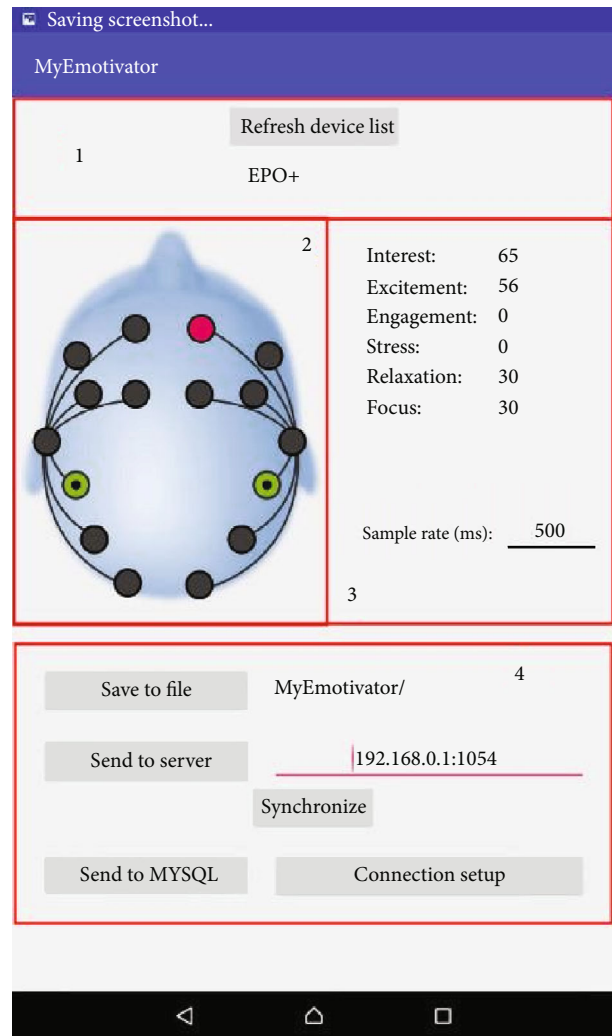

FIGURE 4: The main screen of "MyEmotivator" application.

communicates with the EEG device itself using the Bluetooth protocol (Figure 6).

In general, it is possible to connect any device to the platform, as long as it has support for TCP/IP (i.e., HTTP(S) protocol). Hosting of the server part of the HCI-MAP can be done either in the local network or in the cloud environment available on the Internet. However, despite the possibility of using different and multiple sensors, in this research, we collected the data (events and states) only from the VSTM Web application, the EEG device, eye tracker, and mouse, while only the EEG data (sampled by the VSTM application events) was used in the analysis. The inclusion of the data from other sensors could potentially significantly improve the experiment but requires significant modifications in the experimental design and in the data analysis.

4. Materials and Methods

Twelve subjects, divided into two groups, took part in the experiment. Each subject was presented with two images during two recording sessions, without taking a break. The
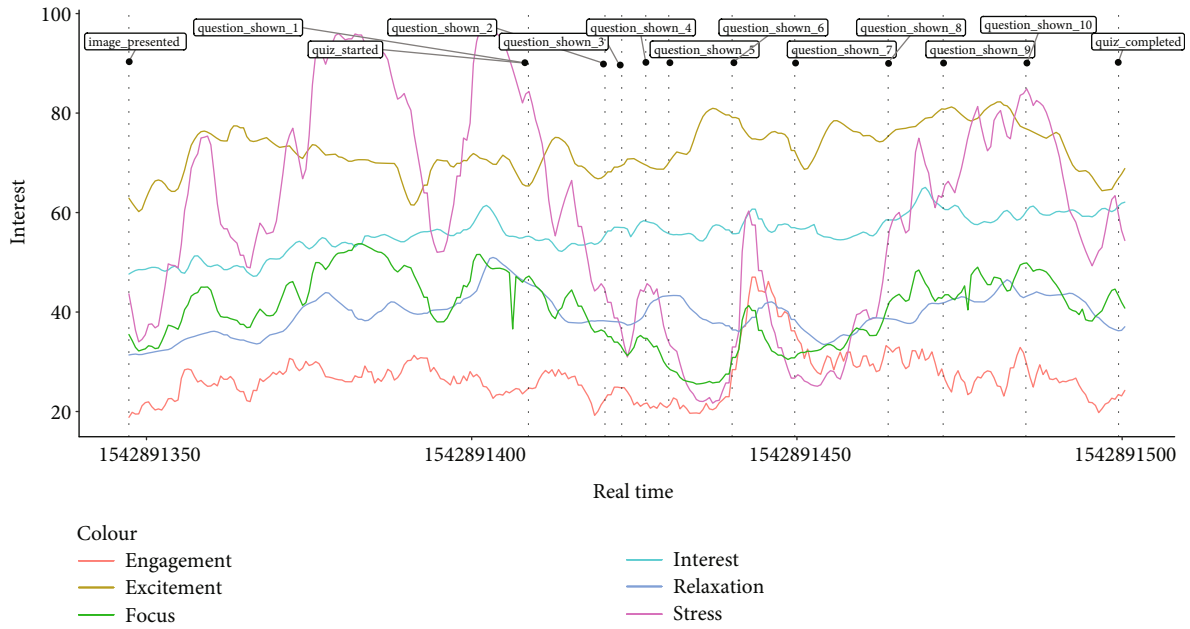
FIGURE 5: Example of the visualized EEG data collected by using the HCI-MAP.
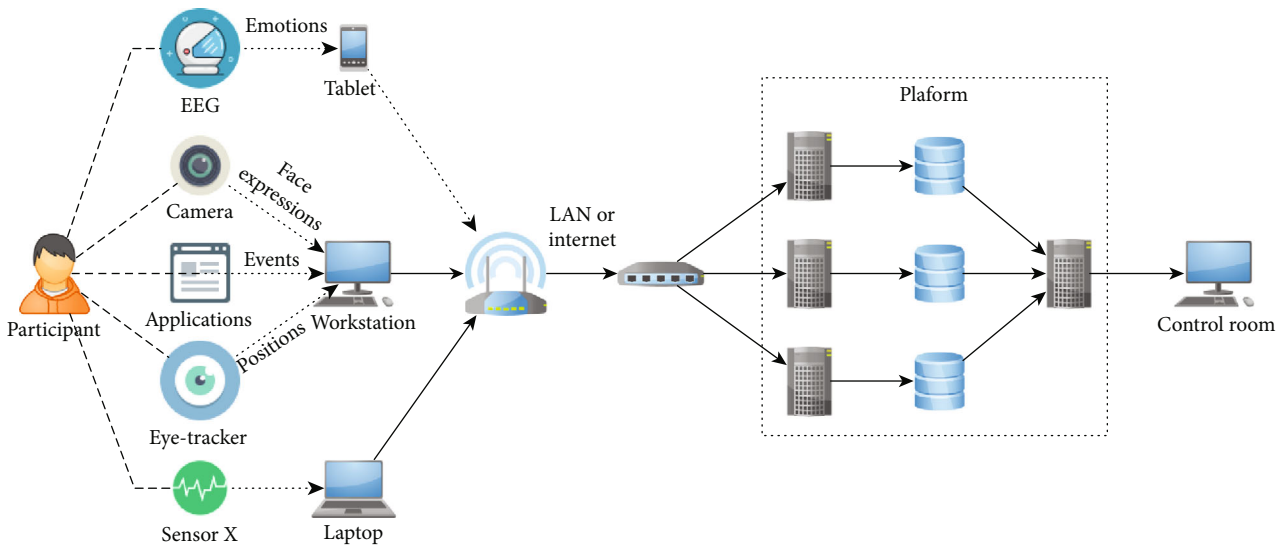

FIGURE 6: The network architecture of HCI-MAP.
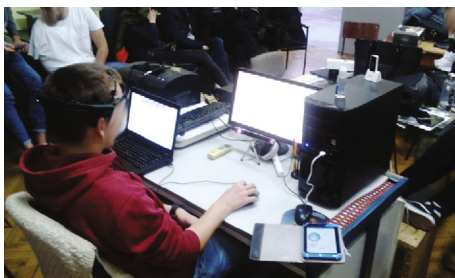
FIGURE 7: Participants wearing EEG equipment during the experiment.

first image was shown in an isolated environment, and the second image was shown in front of the audience. The first group of subjects was presented with Image A as the first image and Image B as the second image. The second group of subjects was presented with the reversed order of images (Image B first and Image A second). Subjects wore the EEG device on their head during the test, and the complete session was recorded with MyEmotivator software. Test setup is shown in Figure 7.

At the beginning of the testing, the subject was presented with an image for 1 minute, with adequate instructions to remember as many details as possible. After that, the subject had 2 minutes to answer 10 questions about image details.

After completion of the test, the number of correct answers was shown to the subject.

We obtained 11 complete measurements. For each user, we had two sets of data, one with measurements related to the first presented image and another related to the second presented image. In the end, we had 23 valid datasets for classification—11 full datasets (for both images) and one set containing partial data from participant number 12 (while watching Image A, second in a row).

Gathered data was organized by participant number, question number, image type (picture A or picture B), image order of viewing (image viewed first or second), and correctness of the answer, using the following features.

4.1. Average Values of Six Emotional States. Gathered values of measured emotional states from EEG signals are absolute, varying in range from 0 to 100. We used three features (*minimal*, *maximal*, and *average* values) for each emotional state. Therefore, we had a total of 18 features (interest min, interest max, interest avg, engagement min, engagement max, engagement avg, etc.). The values of these features were determined for each time interval between two consecutive *question_shown* events (including the first *image_presented* and the last *quiz_completed* event) for a given user session. The features were calculated for each participant, image, and question. For example, we had the minimal, maximal, and average values of each emotional state for participant number 1, while answering question number 1, when Image A was presented as the first picture without the audience.

4.2. The Time Duration for Answering the Questions or Viewing the Image. By using the developed HCI-MAP, we were able to achieve the required synchronization precision of all sensors and application data (<5 ms error margin). The triggered events sent from the VSTM application were recorded in milliseconds, which enabled us to get the exact time period that the participant spent answering every question and viewing the image. There were a total of 11 values for this feature for each user session (10 questions and 1 image display period).

4.3. Normalized Average Value of Six Emotional States. Different personality traits of the human entail great differences in emotional reactions to external stimuli that are reflected through the variance in maximal and minimal values of measured emotional states. For that reason, we introduced a new, relative feature: the *normalized average value* of the emotional state. This feature describes the relation between each average value of emotion and maximal emotion value of the whole session. It is calculated with the formula:

$$\text{normalized value of the emotion} = \frac{\text{average emotion value}}{\text{maximal session emotion value}}. \tag{1}$$

The feature value ranges from 0 to 1. By using it in the classification, in some cases, we increased the percent of successful classifications by approximately 10%.

In our experiment, we made a couple of hypotheses. The first is related to the presence of the audience during the test. We expected that the presence of the audience would represent a distraction for the participants that would be manifested with a greater number of wrong answers and noticeable differences in levels of stress and focus, which can be detected in our quantitative analysis. On the other hand, we expected that the presence of the audience would increase the stress of the participants, and this could be detected by analysis of EEG data. Also, we expected that concentration would be higher while the participants observed the image as compared to when answering the questions, and for the questions where a participant does not know the answer, the stress would be increased.

5. Results and Discussion

5.1. Classification Results. In this paper, we classified data using three different classes: (1) *order of displaying the image* (image displayed first or second (i.e., with or without the audience)), (2) *type of the image* (Image A or Image B), and (3) *correctness of the answer* (image, correct or incorrect). The classification was done using four algorithms: Naive Bayes, support vector, KNN, and random forest. In this paper, only the best two results are presented.

5.1.1. Order of Displaying the Image. Each classifier was presented with two classes: image displayed first in a row and image displayed second. The classification was performed with and without (using only EEG data) the *time duration* feature as a classification attribute.

If the *time duration* feature was used, the best results were achieved using the *K-nearest neighbors classifier*, with $k = 1$, and Euclidean distance function as a search algorithm (Table 1). From a total of 253 instances, the correct classification was done for 228 of them, which is an accuracy of *90.12%* for cross-validation (Table 1(a)) and *86.05%* when training the classifier with 66% of the available data (Table 1(b)). We performed the classification with a k value of *2* (84.19% cross-validation; 86.05% with 66% training set), $k = 3$ (85.38% cross-validation; 86.05% with 66% training set), and $k = 7$ (81.82% cross-validation; 86.05% with 66% training set).

With the *time duration* feature, the second best result was achieved using the *random forest classification algorithm*, with number of iterations set to 100 and unlimited tree depth. In this case, the percentage of correctly classified instances was 85.38% for cross-validation and 86.05% with 66% training set (Table 2).

When not using the *time duration* feature as an attribute in classification algorithms, the best results were achieved using the *K-nearest neighbors classifier*, with $k = 1$, and Euclidean distance function as a search algorithm. Using only gathered emotional state values calculated from EEG data, we had 226 (of 253; *89.33%*) correctly classified instances with cross-validation (Table 3(a)) and 74 (of 86; *86.05%*) correctly classified instances when using 66% of data for training and the rest for testing the classifier (Table 3(b)).

TABLE 1: Order of displaying the image, K-nearest neighbors classifier ($k = 1$) results with time duration feature.

First	Second	<– Classified as
	(a) Cross-validation	
106	15	First
10	122	Second
	(b) Training set with 66% of data	
33	3	First
9	41	Second

TABLE 2: Order of displaying the image, random forest classifier results with time duration feature.

First	Second	<– Classified as
	(a) Cross-validation	
105	16	First
21	111	Second
	(b) Training set with 66% of data	
33	3	First
9	41	Second

TABLE 3: Order of displaying the image, K-nearest neighbors classifier ($k = 1$) results without time duration feature (with EEG data features only).

First	Second	<– Classified as
	(a) Cross-validation	
105	16	First
11	121	Second
	(b) Training set with 66% of data	
33	3	First
9	41	Second

The second best result using emotional state features only was achieved using the *random forest classification algorithm*, with number of iterations set to 100 and unlimited tree depth. With cross-validation, the obtained classification accuracy was 84.98%, and when using the 66% training set, the accuracy was 84.88%.

The above results show that we can determine with the best accuracy of *90.12%* (*86.05%*) to which image order of display every given instance belongs to (i.e., which was the order of the picture the participant viewed when answering the question). Because of the fact that the participants were *isolated* from the audience when viewing the first image and *with* the audience when viewing the second image, we can determine with the accuracy of 90.12% if the participant was in the presence of the audience when answering the question. In the case of image display order, the best results were achieved when using the *time duration* feature together with the emotional state features.

5.1.2. Type of the Image. In this classification attempt, we had two classes: Image A and Image B. Classification was performed with and without (using only EEG data (emotional

state) features) the *time duration* feature as a classification attribute.

When using the *time duration* feature, the best classification results were achieved using the *K-nearest neighbors classifier*, with $k = 1$, and Euclidean distance function as a search algorithm (Table 4). From a total of 253 instances, the correct classification was done for 227 of them, which is an accuracy of *89.72%* for cross-validation (Table 4(a)) and *89.53%* with 66% of the available data (Table 4(b)). We tried the classification with $k = 2$ (84.98% cross-validation; 84.88% with 66% training set) and $k = 3$ (87.75% cross-validation; 88.37% with 66% training set).

The second best result when considering *type of the displayed image* class, with the *time duration* feature, was achieved using the *random forest classification algorithm* with 100 iterations and unlimited tree depth. The percentage of correctly classified instances was 83.40% for cross-validation and 86.05% with 66% training set (Table 5).

If the *time duration* feature as an attribute was not used, the best results were achieved using the *K-nearest neighbors classifier*, with $k = 1$, and Euclidean distance search algorithm. Using only the emotional state features, we had 229 (out of 253; *90.51%*) correctly classified instances with cross-validation (Table 6(a)) and 77 (out of 86; *89.53%*) correctly classified instances when using 66% of the data for training and the rest for testing the classifier (Table 6(b)). This was the best prediction result for classification based on *type of the displayed image* class. With $k = 2$, the percentage of correctly made classifications was 86.56% with cross-validation (87.21% for 66% training set), and for $k = 3$, the percentage was 89.72% with cross-validation (88.37% for 66% training set).

The second best result using EEG data features only was achieved using the *random forest classification algorithm*, with number of iterations set to 100. With cross-validation, the obtained classification accuracy was 85.38%, and when using 66% of data in the training set, the accuracy was 83.72%.

Result analysis shows that we can determine with the best accuracy of *90.51%* (*89.53%*) to which image (A or B, regardless if it was displayed first or second in a row) every given instance belongs to (i.e., which image the participant was viewing before answering the question). In the case of image type, the best results were achieved when using only EEG data-based features without the *time duration* feature. In a similar study [25], by using EEG data gathered from eight participants, the authors were able to successfully classify image clips from a broad area image with an accuracy of 78–95%. In another study by Kawakami et al. [26], participants watched random images from 101 different categories. The recommended algorithm shows that it is possible to correctly detect the image class with an accuracy of 52–74% by using only EEG data. In the paper [27], the authors tested the invariance of brainwave representations of simple patches of colors and simple visual shapes and their names. By using the developed method, they were able to correctly recognize from 60% to 75% of the test-sample brainwaves. The general conclusion was that simple shapes, such as circles, and single color displays generate brainwaves

TABLE 4: Type of the image, *K*-nearest neighbors classifier (*k* = 1) results with time duration feature.

Image A	Image B	<– Classified as
	(a) Cross-validation	
117	15	Image A
11	110	Image B
	(b) Training set with 66% of data	
37	3	Image A
6	40	Image B

TABLE 5: Type of the image, random forest classifier results with time duration feature.

Image A	Image B	<– Classified as
	(a) Cross-validation	
120	12	Image A
30	91	Image B
	(b) Training set with 66% of data	
36	4	Image A
8	38	Image B

TABLE 6: Type of the image, *K*-nearest neighbors classifier (*k* = 1) results without time duration feature (EEG data features only).

Image A	Image B	<– Classified as
	(a) Cross-validation	
119	13	Image A
11	110	Image B
	(b) Training set with 66% of data	
37	3	Image A
6	40	Image B

surprisingly similar to those generated by their verbal names. Our results confirmed that it was possible to distinguish different images based on EEG data, with higher accuracy.

5.1.3. Correctness of the Answer. In this case, an instance can belong to one of three classes: viewing the image, correct answer (true) or wrong answer (false). We did not use the *time duration* feature in any of these classification algorithms, because the time interval for viewing the image (which was 60 seconds) was much longer than the intervals for answering the questions. It should also be noted that we were dealing with an imbalanced dataset (i.e., number of instances belonging to different classes was significantly different). Thus, the number of instances that belong to *viewing the image* was 23, the number of instances belonging to the *correct answer* was 156 and the number of *wrong answer* instances was 74.

Having that in mind, we did not achieve any significant classification results using the *correctness of the answer* attribute as a class. It can be said that the best result was achieved by using the *K*-nearest neighbors classifier, with *k* = 3, and Euclidean distance function as a search algorithm, where we had more correct than incorrect classifications in two

classes: image and true (Table 7), while with all the other used classification algorithms, there were more correct than incorrect classifications in only one class (true). The *K*-nearest neighbor classifier gave *61.66%* correctly classified instances using cross-validation and *61.63%* correctly classified instances with the 66% training set.

5.2. Qualitative Analysis. After performing the classification attempts using three different classes, we analyzed the changes in emotional state features with regard to each of these classes.

5.2.1. Order of Displaying the Image. When analyzing EEG data with regard to the order of displaying the image, we tried to determine the relation between individual feature and the class by comparing the average value of the features from all participants for each of two possible classes (first and second). After that, we selected the features with the biggest difference in average values and grouped them based on belonging to the same emotional state (Table 8).

As can be seen from the results, the top six features that changed the most with regard to different classes were related to the emotional state stress and relaxation. Furthermore, values for both stress and relaxation features (except for Stress avg/MAX) are higher (in average) in case of the class *first*, which means that both stress and relaxation were greater when the participant was doing the test for the first time without the audience as compared to the second time with the audience. The reasons for this require further analysis. This confirms the conclusions from other papers [28–30] that there is a strong correlation between the presence of the audience and stress level (blood pressure and heart rate).

5.2.2. Type of the Image. In the case of *type of the image* class, we used the same method for determining the relation of the individual feature with the class in Section 5.2.1. We grouped the features with the biggest difference in average values with regard to the two classes based on the emotional state they belonged to (Table 9).

The results show that *stress* is the emotional state with the highest change with regard to the type of image the participant was viewing. It was higher, on average, in the case when the participant was viewing Image A as compared to Image B, regardless if the image was shown first or second in a row. The reasons for this will be the subject of our future work.

5.2.3. Correctness of the Answer. As described in Section 5.1.3, in the case of *correctness of the answer*, the instance could belong to one of the three classes. Although we have used the same method of calculating the average value of every individual feature for each of these classes, in this case, we made two separate comparisons. The first comparison was between the class *viewing the image*, on the one hand, and the *wrong* and *correct answer* classes, on the other hand (Table 10), and the second comparison was between the *wrong* and *correct* answer classes (Table 11).

The results from Table 10 show that there is a significant decrease in *stress*, *engagement*, *relaxation*, and *focus* when the participant switched from viewing the image to

TABLE 7: Correctness of the answer, K-nearest neighbors classifier ($k = 3$) results.

Image	True	False	<– Classified as
	(a) Cross-validation		
16	5	2	Image
5	119	32	True
3	50	21	False
	(b) Training set with 66% of data		
4	1	1	Image
3	43	12	True
2	14	6	False

TABLE 8: Comparison of average values for features for class first and second.

Difference in %	Feature	Emotional state
-6.32	Stress max	Stress
-4.27	Stress avg	Stress
+3.76	Stress avg/MAX	Stress
-3.30	Rel max	Relaxation
-2.93	Rel avg	Relaxation
-2.62	Rel min	Relaxation

TABLE 9: Comparison of average values for features for class Image A and Image B.

Difference in %	Feature	Emotional state
-7.93	Stress max	Stress
-6.30	Stress avg	Stress
+4.98	Eng avg/MAX	Engagement
-4.40	Stress min	Stress
-3.66	Exc avg/MAX	Excitement

TABLE 10: Comparison of average values of features for class viewing the image and classes wrong and correct answer.

Average difference in %	Feature	Emotional state
-32.28	Stress max	Stress
-25.29	Eng max	Engagement
-15.85	Rel max	Relaxation
-13.58	Foc max	Focus

TABLE 11: Comparison of average values for features for class wrong answer and correct answer.

Difference in %	Feature	Emotional state
+3.59	Rel max	Relaxation
+3.15	Rel avg/MAX	Relaxation
+2.97	Rel avg	Relaxation

answering the questions. For example, the maximal stress value was, on average, 32.28% higher while the user was viewing the picture.

When comparing the average emotional feature values for the wrong and correct answers, the only significant difference was in *relaxation*, which was higher (on average) for the questions that were answered correctly.

6. Conclusion

In this research, we used EEG device to gather data on human performance while doing the electronic visual short-term working memory test. There were 12 subjects participating in the experiment. At the beginning of the testing, the subject was presented an image for 1 minute, with adequate instructions to remember as many details as possible. The first group of participants was presented Image A as the first image in the isolated environment without the audience and Image B as the second image with the audience. The second group of subjects was presented with the reversed order of images.

The values of six emotional states (interest, engagement, excitement, stress, relaxation, and focus) were used in different classification attempts with regard to three classes: *order of displaying the image*, *type of the image*, and *correctness of the answer*. There are several conclusions that can be drawn from the classification results. First, the classification results show that, by using the emotional states of the participants and question duration interval, we can determine with the best accuracy of 90.12% which was the order of the image the participant viewed when answering the question. Because of the fact that the second image was displayed in front of the audience, this means that we can determine with an accuracy of 90.12% if the participant was in the presence of the audience when answering the question. Second, results obtained from the quantitative analysis of the emotional states of the participants enable us to determine with an accuracy of 90.51% which image (Image A or Image B, regardless if it was displayed first or second in a row) the participant was viewing before answering the question.

Using the qualitative analysis, by comparing overall changes in the emotional states of the participants, we were able to conclude that both stress and relaxation were higher (on average) when the participant was doing the test for the first time (without the audience) as compared with the second time (with the audience). Furthermore, the stress was higher (on average) when the participant was viewing Image A as compared with Image B. Also, there was a significant decrease in stress, engagement, relaxation, and focus when the participant switched from viewing the image to answering the questions, and the relaxation value was higher (on average) for the questions that were answered correctly.

In future work, we plan to deal with some questions that occurred during the analysis of these results. First, there is a question about the cause for the elevated average stress and relaxation values when doing the test the first time (without the audience) as compared to doing it the second time (with the audience). Second, the higher average stress level that was measured when participants were viewing Image A also

requires further analysis. Using the presented classification methods, we were not able to get significant results when trying to find the correlation between EEG data and the correctness of the answer. One of the goals of our future work will be to find a method to predict the outcome of the test (i.e., number of correct answers) by using EEG data and possibly data from other sensors.

Data Availability

The gathered data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

Research was performed as part of the employment of the authors at Singidunum University, Belgrade.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] A. Baddeley, "The episodic buffer: a new component of working memory?," *Trends in Cognitive Science*, vol. 4, no. 11, pp. 417–423, 2000.

[2] R. Luria, P. Sessa, A. Gotler, P. Jolicoeur, and R. Dell'Acqua, "Visual short-term memory capacity for simple and complex objects," *Journal of Cognitive Neuroscience*, vol. 22, no. 3, pp. 496–512, 2010.

[3] N. Cowan, E. M. Elliott, J. Scott Saults et al., "On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes," *Cognitive Psychology*, vol. 51, no. 1, pp. 42–100, 2005.

[4] A. D. Baddeley and G. J. Hitch, "Working memory," in *The Psychology of Learning and Motivation*, G. H. Bower, Ed., pp. 47–90, Academic Press, New York, 1974.

[5] A. D. Baddeley, *Working Memory*, Oxford University Press, New York, 1986.

[6] T. P. Alloway and R. G. Alloway, "Investigating the predictive roles of working memory and IQ in academic attainment," *Journal of Experimental Child Psychology*, vol. 106, no. 1, pp. 20–29, 2010.

[7] T. N. Tombaugh, *Test of Memory Malingering: TOMM*, Multi-Health Systems, North Tonawanda, NY, 1996.

[8] M. R. Asato, J. A. Sweeney, and B. Luna, "Cognitive processes in the development of TOL performance," *Neuropsychologia*, vol. 44, no. 12, pp. 2259–2269, 2006.

[9] A. Jevremovic, S. Arsic, M. Antonijevic, A. Ioannou, and N. Garcia, "Human-computer interaction monitoring and analytics platform - Wisconsin card sorting test application," in *HealthyIoT 2018 - 5th EAI International Conference on IoT Technologies for HealthCare*, Guimaraes, Portugal, November 2018.

[10] H. Ekanayake, "P300 and Emotiv EPOC: does Emotiv EPOC capture real EEG?1–82011, http://neurofeedback.visaduma.info/emotivresearch.htm.

[11] S. Debener, F. Minow, R. Emkes, K. Gandras, and M. de Vos, "How about taking a low-cost, small, and wireless EEG for a walk?," *Psychophysiology*, vol. 49, no. 11, pp. 1617–1621, 2012.

[12] P. S. Inventado, R. Legaspi, M. Suarez, and M. Numao, "Predicting student emotions resulting from appraisal of its feedback," *Research and Practice in Technology Enhanced Learning*, vol. 6, no. 2, pp. 107–133, 2011.

[13] P. Bobrov, A. Frolov, C. Cantor, I. Fedulova, M. Bakhnyan, and A. Zhavoronkov, "Brain-computer interface based on generation of visual images," *PLoS One*, vol. 6, no. 6, article e20674, 2011.

[14] M. K. Petersen, C. Stahlhut, A. Stopczynski, J. E. Larsen, and L. K. Hansen, "Smartphones get emotional: mind reading images and reconstructing the neural sources," *Lecture Notes in Computer Science*, vol. 6975, no. 2, pp. 578–587, 2011.

[15] E. T. Esfahani and V. Sundararajan, "Using brain–computer interfaces to detect human satisfaction in human–robot interaction," *International Journal of Humanoid Robotics*, vol. 8, no. 1, pp. 87–101, 2011.

[16] M. Antonijevic, G. Shimic, A. Jevremovic, M. Veinovic, and S. Arsic, "The potential for the use of EEG data in electronic assessments," *Serbian Journal of Electrical Engineering*, vol. 15, no. 3, pp. 339–351, 2018.

[17] Z. Khayyer, V. Nejati, and J. Fathabadi, "Stress induction and visual working memory performance: the effects of emotional and non-emotional stimuli," *Biotechnology and Health Sciences*, vol. 4, no. 2, 2017.

[18] S. Vogel and L. Schwabe, "Learning and memory under stress: implications for the classroom," *NPJ Science Of Learning*, vol. 1, no. 1, article 16011, 2016.

[19] S. Bona and J. Silvanto, "Accuracy and confidence of visual short-term memory do not go hand-in-hand: behavioral and neural dissociations," *PLoS One*, vol. 9, no. 3, article e90808, 2014.

[20] Z. Mohamed, M. El Halaby, T. Said, D. Shawky, and A. Badawi, "Characterizing focused attention and working memory using EEG," *Sensors*, vol. 18, no. 11, article 3743, 2018.

[21] Emotiv, "Neurotech for the global community," https://www.emotiv.com.

[22] "Advanced EEG technology - backed by science," https://www.emotiv.com/the-science/.

[23] F. Sivrikaya and B. Yener, "Time synchronization in sensor networks: a survey," *IEEE Network*, vol. 18, no. 4, pp. 45–50, 2004.

[24] D. L. Mills, "Internet time synchronization: the network time protocol," *IEEE Transactions on Communications*, vol. 39, no. 10, pp. 1482–1493, 1991.

[25] N. Bigdely-Shamlo, A. Vankov, R. R. Ramirez, and S. Makeig, "Brain activity-based image classification from rapid serial visual presentation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 5, pp. 432–441, 2008.

[26] T. Kawakami, T. Ogawa, and M. Haseyama, "Novel image classification based on decision-level fusion of EEG and visual features," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5874–5878, Florence, Italy, May 2014.

[27] P. Suppes, B. Han, J. Epelboim, and Z.-L. Lu, "Invariance of brain-wave representations of simple visual images and their names," *Proceedings of the National Academy of Sciences*, vol. 96, no. 25, pp. 14658–14663, 1999.

[28] S. Taylor, T. Seeman, N. Eisenberger, T. Kozanian, A. Moore, and W. Moons, "Effects of a supportive or an unsupportive audience on biological and psychological responses to stress,"

Journal of Personality and Social Psychology, vol. 98, no. 1, pp. 47–56, 2010.

[29] N. Bolger and D. Amarel, "Effects of social support visibility on adjustment to stress: experimental evidence," *Journal of Personality and Social Psychology*, vol. 92, no. 3, pp. 458–475, 2007.

[30] N. Bolger, A. Zuckerman, and R. C. Kessler, "Invisible support and adjustment to stress," *Journal of Personality and Social Psychology*, vol. 79, no. 6, pp. 953–961, 2000.