

# NeuroPilot: A Real-time Brain-Computer Interface system to enhance concentration of students in online learning

Asif Islam, Farhan Ishtiaque, Md. Muhyminul Haque, Farhana Sarker, Ravi Vaidyanathan, and Khondaker Abdullah-Al-Mamun

**Abstract**—The prevalence of online learning poses a vital challenge in real-time monitoring of students’ concentration. Traditional methods such as questionnaire assessments require manual intervention, and webcam-based monitoring fails to provide accurate insights about learners’ mental focus as it is deceived by mere screen fixation without cognitive engagement. Existing BCI-based approaches lack real-time validation and evaluation procedures. To address these limitations, a Brain-Computer Interface (BCI) system is developed using a non-invasive Electroencephalogram (EEG) headband, FocusCalm, to record brainwave activity under attentive and non-attentive states. 20 minutes of data were collected from each of 20 participants watching a pre-recorded educational video. The data validation employed a novel intra-video questionnaire assessment. Subsequently, collected signals were segmented (sliding window), filtered (Butterworth bandpass), and cleaned (removal of high-amplitude and EOG artifacts such as eye blinks). Time, frequency, wavelet, and statistical features were extracted, followed by recursive feature elimination (RFE) with support vector machines (SVMs) to classify attention and non-attention states. The leave-one-subject-out (LOSO) cross-validation accuracy was found to be 88.77%. The system provides feedback alerts upon detection of a non-attention state and maintains focus profile logs. A pilot study was conducted to evaluate the effectiveness of real-time feedback. Five participants underwent a 10-minute session comprising a 5-minute baseline phase devoid of feedback, succeeded by a 5-minute feedback phase, during which alerts were activated if participants exhibited inattention for approximately 8 consecutive seconds. A paired t-test ( $t = 5.73$ ,  $p = 0.007$ ) indicated a statistically significant improvement in concentration during the feedback phase.

**Index Terms**—brain-computer interface (BCI), electroencephalography (EEG), machine learning (ML), support vector machine (SVM)

## I. INTRODUCTION

CONCENTRATION monitoring has become essential to ensure learning effectiveness in online classes, whose adoption has been steadily increasing across different educational curricula [1]. Prior research shows that learners are more prone to mind-wandering when they are watching lecture videos (being regarded as online learning scenario) than during live lectures (traditional offline classes) [2]. Thus, to track the engagement of students in educational and learning settings, conventional methods of developing questionnaires [3] or camera-based monitoring [4], [5] are used. Although questionnaires are useful for the objective evaluation of students, they only serve as a manual tool that requires human intervention and input. Conversely, webcam-based monitoring systems are limited to analyzing external behavioral cues such as facial expressions, gaze, or posture, and therefore lack the ability to provide direct neurophysiological insights into

learners’ cognitive attention levels [6], alongside inducing privacy issues [7]. Additionally, vision-based models fail to preserve robustness in low-light conditions. For this reason, Brain-Computer Interface (BCI) has emerged as a powerful tool to perform cognitive load evaluation of students while watching lecture videos.

BCI is a form of neurotechnology that is leveraged to establish a direct connection between humans and machines. Specifically, electroencephalography (EEG) signals are increasingly being utilized for various use cases, including bionic intelligence [8], vehicle control [9]–[11], and other assistive technologies [12]–[15]. BCI has gained considerable recognition in the classification of attentive and non-attentive states [16]–[19], including distraction [20] and drowsiness [21]. Despite the fact that this article is particularly oriented to enhance concentration levels, terms like focus, attention, concentration, and engagement, which are different constructs overlapping each other, can be used interchangeably as the goal is to improve online learning for students using BCI. It has been seen from studies that the prefrontal cortex plays a crucial role in evaluating sustained attention [22]–[24]. In the frontal region, the features contributing to concentration level detection are predominantly the changes in delta, theta, alpha, and beta band power. [22], [23], [25].

Current state-of-the-art works use EEG devices for laboratory experimentation that are not fully consumer-grade, which makes them not suitable for real-time implementation. Some of them are Emotiv Insight, Emotiv EPOC+, and Open-BCI Ultracortex. These devices lack dry electrodes, requires accurate placement across multiple scalp regions, not cost-effective and present setup challenges for consumers [26], [27]. Among consumer-grade devices, cost-effective options such as Muse, NeuroSky MindWave, and FocusCalm were studied. However, existing works provide limited details about their data collection strategies in online learning scenarios [28]–[31], leaving a gap in collecting ground truth (referring to correct labeling of neural data containing actual concentration and actual non-concentration states). Other consumer-grade devices from companies such as Cognionics and Wearable Sensing have been promoted for use in sports scenarios, but their high cost and large number of electrodes limit their suitability for consumer-friendly applications in this field. Vital gaps are seen in real-time implementation of BCI research articulated to track the concentration of students in learning environments. Huang et al. [32] focused on real-time cognitive monitoring to improve concentration time with neuro-feedback mechanism. However, the data collection strategy included mental arithmetic tasks, which is not an identical simulation of

a learning environment, and the data validation was done using post regulation assessments such as the Sustained Attention to Response Task (SART) [33], which fails to validate the participants' engagement during the mental arithmetic tasks. Furthermore, the working algorithm of the single-channel EEG headband, along with the feature extraction and model training process, were not discussed, as their work focused more on evaluating the effects of neurofeedback rather than on developing the classification methodology. Conrad and Newman developed a strategy to use EEG signals for evaluating mind-wandering in online classes [34]. However, the use of only pre- and post-assessment questionnaires is insufficient to validate learners' attention levels during the video. The work done by Rehman et al. [35] relied on publicly available datasets, which poses the same limitations as the earlier research gaps. The studies summarized in Table I are selected to represent existing approaches for monitoring learner attention in online learning environments. Inclusion criteria are works that addressed learner engagement or attention in online or remote learning settings, and use of other related modalities (e.g., webcam-based, neurofeedback) whose limitations are being addressed in our proposed approach. Exclusion criteria included works focusing purely on offline classroom learning, cognitive tasks unrelated to learning (e.g., driving simulators), or studies that did not explicitly address attention/engagement measurement.

TABLE I  
SUMMARY OF PRIOR STUDIES

| Work                     | Focus                                                                | Limitations                                                              |
|--------------------------|----------------------------------------------------------------------|--------------------------------------------------------------------------|
| Camelia [3]              | Questionnaire to validate learner engagement                         | Requires manual interventions                                            |
| Patil [4] and Hossen [5] | Computer Vision (webcam) to understand learner engagement            | Lack of insights in cognitive processing                                 |
| Huang [32]               | Validating the effects of neurofeedback in increasing attention span | Lack of proper documentation on classification algorithm                 |
| Conrad [34]              | Evaluating mind-wandering of students in online classes using EEG    | Lack of proper data validation techniques                                |
| Rehman [35]              | Attention classification using their Deep Q-Learning model           | Computationally intensive and less feasible for real-time implementation |

In this study, a eeg based attention classification framework is proposed to address the research gaps, with a focus on real-time implementation in online learning scenarios. A novel data acquisition and validation design through an intra-video questionnaire assessment was done to label the blocks as ground truths. The experimentation was done on classifying two cognitive states of students, namely attention and non-attention. Twenty participants were recruited to participate in data collection experiments containing two sessions, designated for the two classes of data and the device used is a FocusCalm headband. After that, several preprocessing techniques were employed to get clean data, followed by feature extraction, feature selection and classification. Following this, leave-one-subject-out validation was used in order to evaluate

the classifier model when faced with unseen data. Finally, a simple app GUI was developed to have real-time insights of concentration levels along with a feedback mechanism to alert the learner during non-attention states. Thus, this article has contributions in the following areas:

- A data collection paradigm is introduced that uses authentic lecture videos from educational technology platforms as EEG stimuli, offering a closer approximation to real-world online learning environments than traditional controlled tasks.
- A novel intra-video questionnaire assessment is proposed for data validation, which, unlike traditional post-task questionnaires or external measures (e.g., SART), collects learner responses during the lecture itself. This approach minimizes recall bias, provides time-specific ground-truth labels, and enhances the reliability of the EEG dataset.
- A real-time system is developed with feedback alerts to ensure the attention spans of learners improve over time. Preliminary evidence of its effectiveness in enhancing concentration is validated via a pilot study.

## II. DATA COLLECTION

The section gives a detailed explanation of the inclusion criteria of participants, the device used for data collection, the experimental design, and setup.

### A. Participant Selection

A total of 20 students with a mean age of  $22.8 \pm 3.8$  years were recruited for the data collection experiment. A detailed explanation of the experimental setup, collection paradigm, and the study was given to the participants before taking their informed consents. Participants retained the right to leave the data collection process at their discretion at any moment. This data collection protocol was approved by the Institutional Research Ethics Board (IREB), United International University (UIU). Among the participants, 15 are graduate students of Advanced Intelligent Multidisciplinary System Lab (AIMS Lab) under Institute of Research, Innovation, Incubation & Commercialization (IRIIC), UIU and rest are the students of UIU. They were selected to pass the condition of having cognitive abilities necessary to comprehend a standard high school or equivalent curriculum-based educational material. This was ensured due to the video stimuli being exposed to the participants at the time of data collection, which contained online lectures on different topics prepared by national educational technology platforms.

### B. Data Acquisition Device

FocusCalm headband [36] was used as the data acquisition device in this work for its consumer-grade accessibility, dry electrode system, frontal EEG coverage, low setup complexity and a balanced trade-off between signal quality and user comfort. It is a neuro-feedback EEG device designed for meditation purposes. There are three dry electrodes placed close to each other and aligns with the Fp1, Fp2, and Fpz positions of the international 10-10 electrode placement system as shown

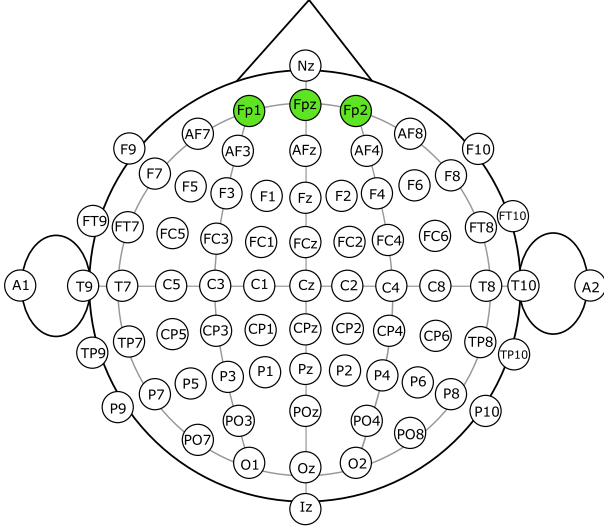


Fig. 1. Positions of Fpz, Fp1 and Fp2 electrodes used by FocusCalm headband according to the international 10-10 electrode placement system.

in Fig. 1. Fpz is used as the primary electrode, while Fp1 and Fp2 serve as the reference and ground electrodes [37]. The device wirelessly sends single channel data at 250 Hz, along with eight distinct values, of which six indicate the delta, theta, alpha, low beta, high beta, and gamma frequency band values, and the other two show attention and meditation scores given by their developed model. The developers of FocusCalm at BrainCoTech have made a software development kit (SDK) named Crimson SDK, which is accessible for taking raw EEG data along with the eight other values in real-time.

### C. Data Collection Experiment

The data collection experiment was designed using the PsychoPy software [38]. The design included a starting instructional video for both classes (attentive and non-attentive) that gives a brief view of the experiment done.

The participants were asked to watch a 10-minute video lecture that was split into 30-second blocks. Subsequently, a question on that exact block was shown on the screen with multiple choices, which the participants had to answer. The duration given for answering the question was set to 15 seconds, as an average human takes around 4 to 5 seconds for reading a sentence [39]. This gives the participants 10 seconds to answer which is considered ample, as humans take around 2 seconds to recognize answers [40]. The questions were set to ensure that the participant has followed the lecturer's quotes and drawings on the board. The questions require no external knowledge to be answered in order to ensure an objective evaluation of a participant's engagement. After the instructional video, a short sample video was played followed by a question-answer block to familiarize the participant with the experiment. A conditional screen was displayed with a fixation cross where the participant has to press a key to start the video lecture. In this way, each participant had to watch 20 video blocks of a lecture followed by questions inserted between them. The experimental setup captures timestamps, which mark the start and end of each video block together

with participants' responses to the questions. These records were subsequently used for data validation.

For the non-attentive data collection session, similar video sequences were presented, but rendered inaudible and blurred to prevent participants from perceiving the content. The participants were instructed to stare at the screen for 5 minutes followed by 5 minutes with eyes closed. In the eyes-open condition, participants viewed the hazy, non-informative video, which was designed as a monotonous, low-load task. Such tasks, which provide minimal external stimulation, are known to induce states of boredom or an "unengaged mind", creating a strong tendency for attention to shift from the external environment to internal thought [41], [42]. This method is based on extensive findings that mind-wandering is most prevalent when cognitive demands are low [43], thereby simulating a common occurrence in real-world scenarios where students face difficulties in maintaining attention during online lectures. The eyes-closed condition was designed to simulate drowsiness or dozing off during online lectures. A brief overview is illustrated in Fig. 2a.

### D. Data Validation

After successful data acquisition, the EEG data were stored with proper timestamps through some modifications of the Crimson SDK, and the responses to the questions were stored with synchronized timestamps. The correct response from a participant validates the EEG data corresponding to the video block that follows the question. All validated blocks were stored as ground truths for attentive data. For non-attentive data, the questionnaire assessment was not performed and thus required no further validation. The procedure is shown in Fig. 2b.

## III. METHODS

The validated data were processed through several steps, namely data pre-processing, feature extraction, and feature selection before being used as training data for classification.

### A. Data Preprocessing and Preparation

Data were pre-processed through steps that include segmentation, filtering, noise removal, and smoothing.

1) *Segmentation*: EEG data of each participant of a total length of  $T$  samples were partitioned to get data segments of optimal length  $L = 1750$  samples (details in section III.D). Although the average sampling frequency of 250 Hz was retained, samples were used instead of time intervals to avoid errors caused by minute temporal delays in packet transmission, which complicate timestamp-based calculations. Through validation, a total of  $K$  blocks were selected from the initially acquired  $n$  blocks for each participant. This number corresponds to the average number of questions (around 20) set for a video lecture per session. A sliding window was used during segmentation to reduce the edge effects and improve statistical reliability [44]. The sliding window is also referred to in this article as the overlap ratio  $r = 0.7$  (details in section

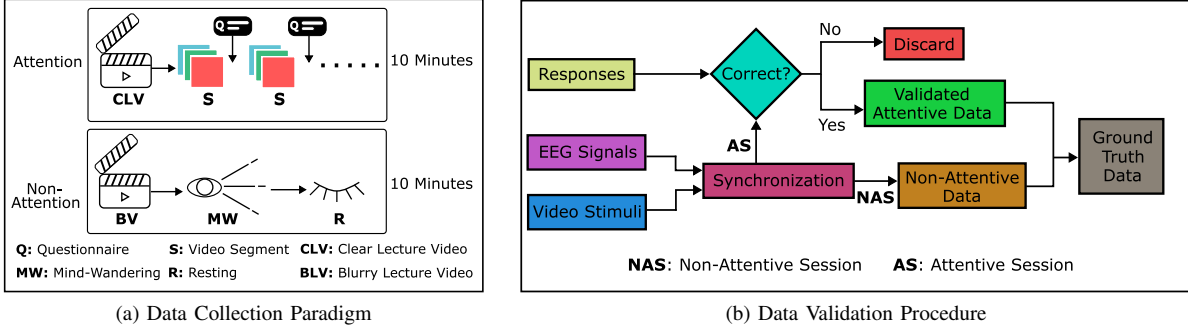


Fig. 2. Overview of the data collection paradigm and validation procedure.

III.D). The total number of segments found for each selected block is:

$$n_k = \left\lfloor \frac{T_k - L}{L(1-r)} \right\rfloor + 1$$

Thus, the total number of segments found is:

$$n = \sum_{k=1}^K n_k = \sum_{k=1}^K \left( \left\lfloor \frac{T_k - L}{L(1-r)} \right\rfloor + 1 \right)$$

The total number of samples is  $x = n \times L$ . During segmentation, two columns were added for labeling and assigning subject-wise user\_ids. These were used later in classification.

To ensure subject-wise label balance, samples for both classes were extracted separately for each subject. Let  $\mathcal{D}_u^0$  and  $\mathcal{D}_u^1$  denote the sets of samples with labels 0 and 1 respectively (0 meaning attentive and 1 meaning non-attentive), for a given user  $u$ . An equal number of samples was retained by trimming the larger class to match the size of the smaller one. Number of samples to be retained,  $n_u = \min(|\mathcal{D}_u^0|, |\mathcal{D}_u^1|)$ , and the last  $n_u$  samples from each label set were selected, which gave  $\mathcal{D}_{u,\text{tail}}^0(n_u)$  and  $\mathcal{D}_{u,\text{tail}}^1(n_u)$ . The balanced dataset for user  $u$  was then

$$\mathcal{B}_u = \mathcal{D}_{u,\text{tail}}^0(n_u) \cup \mathcal{D}_{u,\text{tail}}^1(n_u),$$

and the final dataset was

$$\mathcal{B} = \bigcup_{u \in \mathcal{U}} \mathcal{B}_u$$

Given that the dataset retained a substantial number of segments, class imbalance was addressed by trimming the larger class rather than oversampling the smaller one.

2) *Filtering*: The EEG signals were filtered using a bidirectional Butterworth bandpass with a cutoff frequency of 0.5–64 Hz, which is a standard range in this research field, and an optimal order of 3 found through graphical comparisons. The problem arises due to edge effects caused by filter initialization and group delay as seen in Fig. 3a. To overcome this, 250 samples were trimmed from both ends of each segment to get clean signal with minimal phase distortion after filtering. The procedure is illustrated in Fig. 3b. This method added a constraint to the selection of segment length for experimentation which is  $L > 500$  samples. This means that the new segments found after trimming had a length of  $(L - 500)$  samples, which gave rise to another constraint of setting the ratio  $r$ . The constraint can be understood as  $r \times L > 500$ . The logical explanation is understandable from

the overall segmentation and trimming procedure, depicted in Fig. 3c, where the sequential overlapping and trimming of the segments are shown along with the obtained segments. Finally, the signal was passed through a 50 Hz notch filter to remove power line interference.

3) *Artifact Removal*: Non-invasive EEG signals are often contaminated by artifacts, including eye blinks. Two steps were performed for noise removal. First, the filtered segments were checked for high-amplitude values greater than a threshold of  $150 \mu\text{V}$ , as meaningful EEG data lie below this value [45]. If any segment contained at least one high amplitude value, then that segment was dropped from the dataset. Correlation and interpolation are commonly used for high-amplitude portions [46], [47], but the segment rejection approach was used to exclude artificial continuations, given the abundance of collected data. Following this, eye blinks were removed through Ensemble Empirical Mode Decomposition (EEMD) [48]. Peak indices were extracted from the filtered EEG signal using the `find_peaks` function from the SciPy library [49], which finds all the local maxima. Then three conditions were imposed to take the region around the peak. If a peak index fell within the starting 40 or ending 80 values of the segment, then the starting 120 or ending 120 values were taken for EEMD. However, if the peak index fell within the rest of the region then the 40 values before and 80 values after the peak index were taken. These indices were acquired through rigorous experimentation to get optimal artifact removal. The signal found after performing EEMD was then smoothed using a uniform filter to eliminate residual high-frequency noise.

## B. Feature Extraction

Following pre-processing, feature extraction was applied to the segments. A comprehensive set of 458 features was extracted from the EEG signals to capture time-domain, frequency-domain, wavelet-domain, and statistical characteristics. Time-domain features included commonly used descriptors such as Hjorth parameters (e.g., complexity, activity), median (md), normalized second difference (n2d) etc. Frequency-domain features encompassed spectral moments (e.g., centroid, spread), entropy-based measures, and spectral shape descriptors (e.g., crest, flatness, roll-off). Additionally, wavelet packet transform (WPT) was used to extract sub-band energy and entropy features, which effectively captures local-

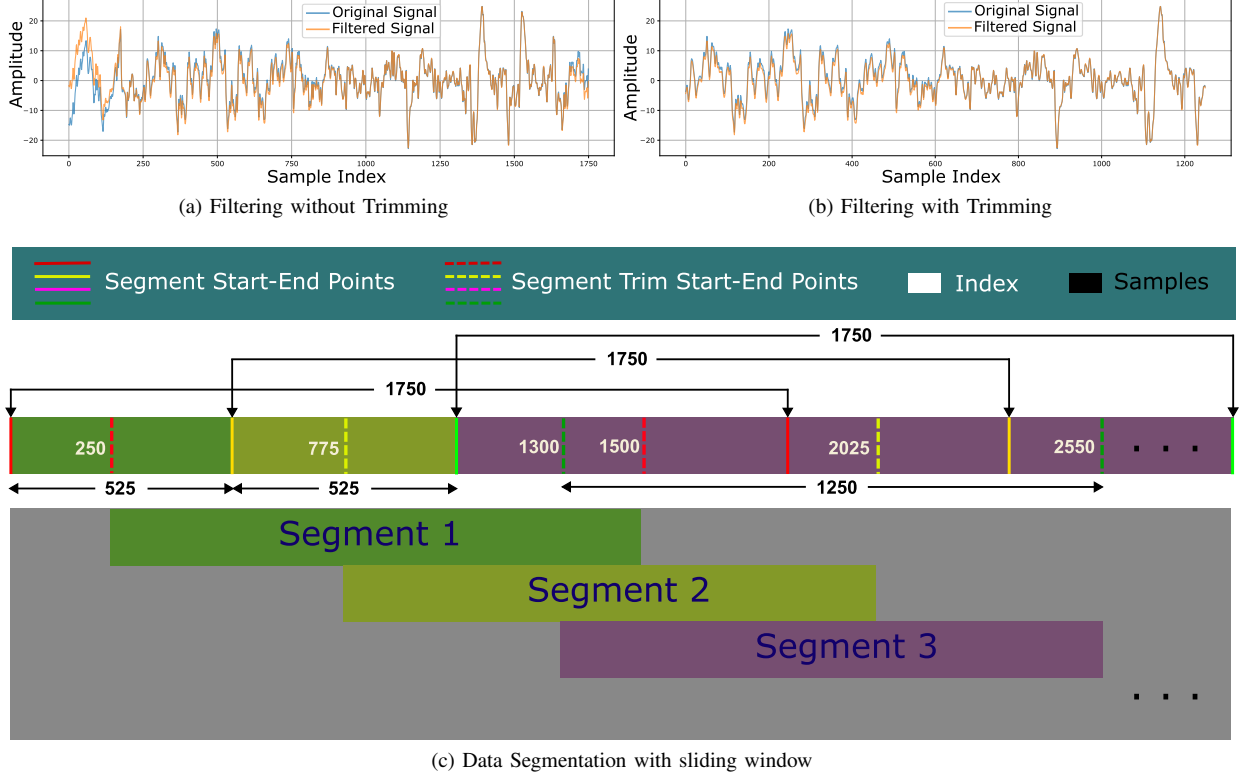


Fig. 3. Subfigures (a) and (b) illustrate the effect of trimming on filtered EEG signals, showing signals without trimming in (a) and with trimming in (b). Subfigure (c) presents the data segmentation pipeline using overlapping windows, highlighting how trimming affects the resulting segments.

ized time–frequency dynamics. These features were selected based on their proven effectiveness in attention evaluation through EEG and biosignal analysis. Eight additional features from the Crimson SDK (six frequency band powers along with attention and meditation scores) were appended to the initial 458 features, for a total of 466. These standard, open-source features were included to explore their potential contribution. Following feature extraction, a data cleaning step was performed. Any features containing NaN or Inf values (arise from divisions by zero or logarithmic transformations of zero-valued data) were replaced with zero. Zero was chosen as a neutral, non-informative value that preserves data dimensionality and ensures numerical stability for model training.

### C. Feature Selection

The next step after feature extraction was selection of the most relevant features for training the data. This was necessary to ensure that the final model was trained on the best set of features that contribute to the classification performance. The selection was done in two steps: Pearson correlation filtering (PCF), followed by Recursive Feature Elimination (RFE) using Support Vector Machine (SVM).

1) *Pearson Correlation Filtering*: PCF was used to reduce high dimensionality and optimize computational efficiency, which works by selecting one from each highly correlated pair when the coefficient of correlation exceeds a threshold  $p$ . The value of  $p = 0.8$  (details in section III.D) was used to remove the highly correlated features, leaving 108 in the

set for further analysis. The Pearson correlation coefficient is defined as:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

where  $x$  and  $y$  are the features,  $\bar{x}, \bar{y}$  are their means and  $n$  represents the number of paired observations.

2) *Recursive Feature Elimination with SVM*: Subsequently, RFE with SVM was applied to find the meaningful set of features required for optimal model training. It is a supervised feature selection method that removes the least important features recursively based on the weights assigned by a linear SVM model. At each iteration, the model was trained on the current feature set, and features with the smallest absolute weights (least contribution to the decision function) were eliminated. This process was continued until the desired number of features,  $d = 50$  (details in section III.D), was retained.

First, RFE was applied for each subject (identified by user\_id) using a linear SVM. The regularization parameter  $C_1$  for this SVM was iterated over a range (0.01 to 100), to account for varied model complexity and feature importance. For each  $C_1$  value, a set of 50 top-ranked features was obtained. These sets were then evaluated through another SVM with a radial basis function (RBF) kernel. Classification performance of this secondary model served as the basis for selecting the optimal feature set. This subject-focused process was designed so that the selected features were not only ranked high based on linear separability but also demonstrated better generalization of the model. Finally, a set of 9 common features was obtained

by finding the most frequently selected features across all subjects. These included the relative powers of alpha and delta band, ratio of low beta to summation of alpha and theta band, average power of low beta band, and median of high beta band. These features showed strong correlations with those reported in the literature as contributing to sustained attention [22]–[25]. These consensus-based features were used for final model training. The selected features for final training along with their meanings are given in Table II.

TABLE II  
SELECTED FEATURES FOR TRAINING

| Feature    | Meaning                                                |
|------------|--------------------------------------------------------|
| RP_A       | Relative power of alpha band                           |
| RP_D       | Relative power of delta band                           |
| attention  | Attention Score from Crimson SDK                       |
| avg_pow_B1 | Average power of low beta band                         |
| en_b_at    | Ratio of low beta to summation of alpha and theta band |
| hc_D       | Hjorth complexity of delta band                        |
| md_B2      | Median of high beta band                               |
| meditation | Meditation Score from Crimson SDK                      |
| n2d_G      | Normalized second difference of gamma band             |

#### D. Hyperparameter Tuning

All parameters defined from pre-processing to feature selection were systematically varied to evaluate their impact on classification performance through rigorous assessments. For the segment length  $L$ , values ranging from 5 to 8, paired with overlap ratios ranging from 0.3 to 0.7, were used for analysis. The pair of  $(L, r) = (1750, 0.7)$  was found to be optimal through comparisons of model performance across all tested values. A similar approach was followed for the Pearson correlation coefficient threshold  $P$  (0.6 to 0.95 in increments of 0.05) and the optimal number of features  $d$  (30 to 80 in increments of 10). The best trade-off between model performance and generalization was found for  $P = 0.8$  and  $d = 50$ . For RFE, the linear SVM was trained on  $C_1 \in \{0.01, 0.1, 1, 10, 100\}$ , and the secondary SVM for identifying the optimal feature set was trained using a nested grid search over two parameters:  $C_2 \in \{0.01, 0.1, 1, 10, 100\}$  and  $\gamma \in \{0.001, 0.01, 0.1, 0.5, 1, 10\}$ . The combination yielding the highest performance determined the best feature set corresponding to the optimal  $C_1$ . Final classification was further tuned using a separate grid search on the same range of  $C_2$  and  $\gamma$ , applied to the common features obtained across all users. The optimal pair  $(C_2, \gamma) = (0.01, 0.5)$  was found to yield the best classification performance.

#### E. Classification

SVMs are widely adopted for EEG classification tasks due to their robustness in high-dimensional spaces, effectiveness with small sample sizes, and ability to model non-linear decision boundaries through kernel functions. Therefore, an SVM with an RBF kernel was selected to capture non-linear relationships in the feature space. The SVM finds a hyperplane that maximizes the margin between classes while allowing some misclassifications, which are controlled by the regularization parameter  $C_2$ . The kernel coefficient  $\gamma$  determines

the influence of individual training samples on the decision boundary, thereby shaping how tightly the model fits local patterns in the feature space.

The training pipeline included fitting the model on a nested grid search of  $C_2$  and  $\gamma$ . The evaluation was performed through leave-one-subject-out (LOSO) cross-validation. LOSO is an appropriate choice to effectively assess the model's generalizability to unseen subjects, thereby closely reflecting real-world application scenarios.

After a rigorous grid search, the pairs of  $C_2$  and  $\gamma$  yielding the best classification performance for the subjects were recorded. From these pairs, the most frequent combination was selected to fit the final model, which was then used for real-time evaluation.

#### F. Evaluation Metrics

The performance of the classifier was evaluated using standard metrics, including accuracy, precision, recall, and F1-score. Since the dataset was balanced per class during segmentation, these metrics were sufficient to estimate the model performance.

### IV. EXPERIMENTAL RESULTS

This section presents the performance evaluation of the trained model and the ablation studies that led to its optimized version.

#### A. Cross-Validation Results

The model performance was analyzed using the LOSO cross-validation procedure (discussed in section III.E). The optimized model's performance on each fold (fold implying the data of one subject as test data while taking all the others as train data) is shown in Fig. 4. The optimized model achieved an average test accuracy of 88.77%. Additionally, it attained a precision of 90.49%, recall of 88.77%, and F1-score of 89.03%, which indicates strong predictive capability with a good balance between correctly identifying positive cases and minimizing false positives.

#### B. Model Ablation Studies

Initially, 458 features excluding the 8 features provided by the Crimson SDK, were used to train the SVM, which gave an accuracy of 76.44%. Subsequently, the eight SDK features were incorporated, and the model was retrained, achieving the current best performance. Among these additional features, the attention and meditation scores were included in the final feature set. To assess whether the model's performance is overly reliant on these two scores, it was compared against the baseline performance obtained using only the Crimson SDK's scores as the feature set. The Crimson SDK was selected as the baseline, as it provides the scores associated with cognitive load. To evaluate the predictive accuracy of these scores for the EEG signals, two methods were employed. The first method included training the model on the two features and evaluating the metrics, which gave an accuracy of 84.13%. The second method involved the ratio of attention to

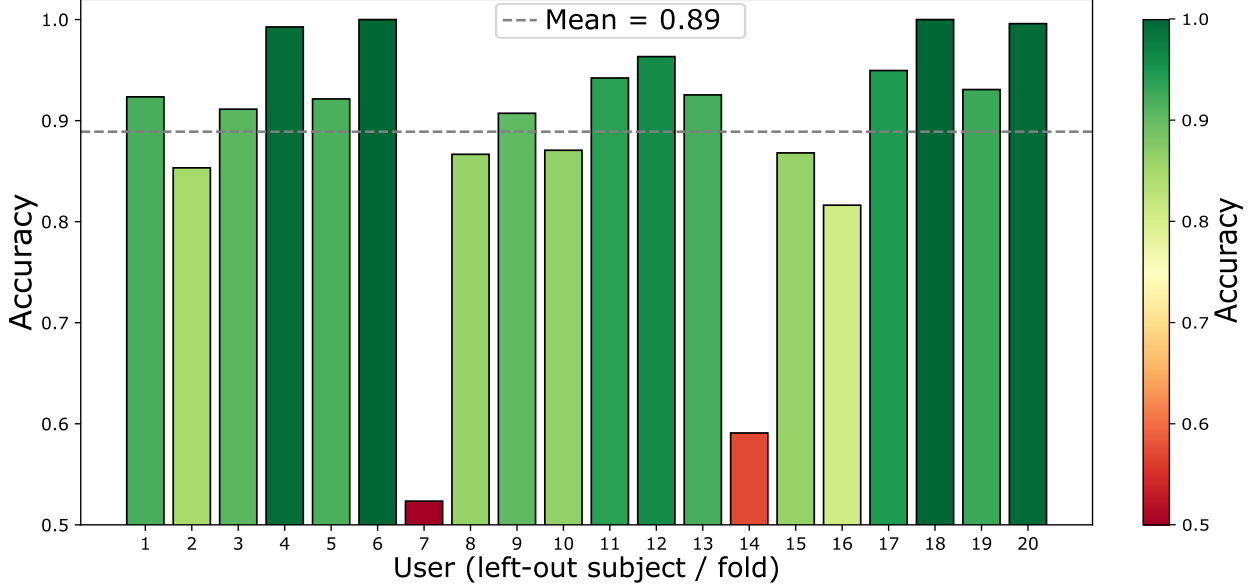


Fig. 4. Accuracy of the model evaluated using Leave-One-Subject-Out (LOSO) cross-validation, where each iteration treats one subject as the test fold and the remaining subjects as the training set.

meditation score  $r_{am}$  as the condition to classify the data. For all the  $n$  segments found, the predictions were taken to be attentive if  $r_{am} > 1$  and non-attentive otherwise. The accuracy was found to be 82.27%. This shows that the proposed methodology outperforms the current interface of the FocusCalm headband modified for attention classification in online learning environments (details discussed in section V). A brief summary is provided in Table III for a clear overview.

TABLE III  
MODEL ABLATION PERFORMANCE

| Method                                                 | Performance |
|--------------------------------------------------------|-------------|
| 458 features excluding attention and meditation scores | 76.44%      |
| Ratio of attention and Meditation scores               | 82.27%      |
| Attention and meditation scores as final features      | 84.13%      |
| 466 features including attention and meditation scores | 88.77%      |

## V. DISCUSSION

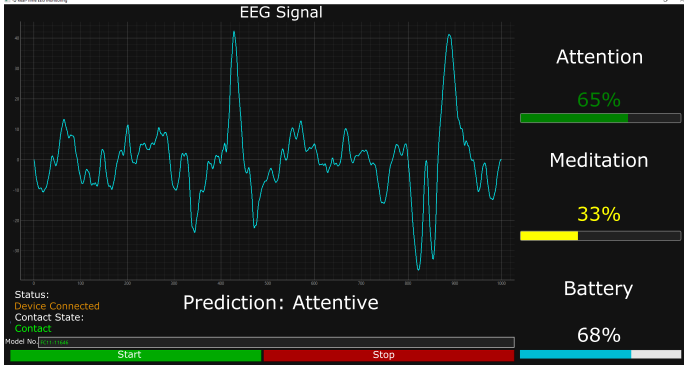
The overall research methodology differs from baseline methods (given the limited prior research addressing this objective within this specific environment). Therefore, the discussion focuses on the unique aspects of the approach, the rationale for their selection, their implementation, and the specific methods they outperform. Finally, this section addresses the limitations of the current research and provides a logical justification for the methods selected.

The proposed EEG-based attention classification achieves 6.5% better accuracy than the Crimson SDK, which suggests that the combination of wavelet packet decomposition along with entropy features are effective in capturing attention state dynamics. Although baseline methods achieve satisfactory results, the data collection strategies of other baseline studies

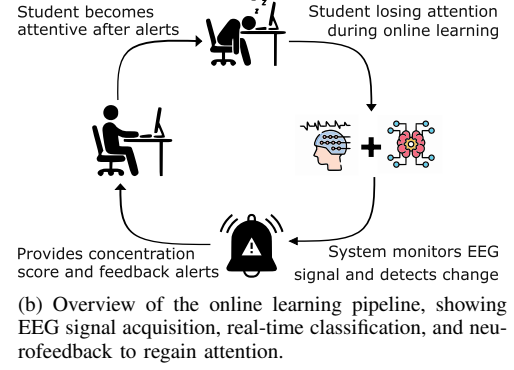
such as Rehman et al. and Huang et al. are based on instructions and there was room for data validation strategies which are implemented in this research. The steps followed in this research ensure correct predictions which pertain to ground truths and this is crucial for real-time implementation.

Initially, device selection was emphasized, as most previous research relied on non-consumer-grade devices. For this study, the FocusCalm headband was chosen for its portability, wearability, and ease of hands-on use, despite the potential for experiments with the Muse headband, which offers greater spatial coverage in real-time applications. Given our objective of developing a state-of-the-art real-time binary classification system, the FocusCalm device has proven sufficient for our purposes. The data collection protocol was designed to reflect the environmental setting of classrooms, with video stimuli consisting of lecture content prepared for students. The data validation process was established under the principle that only ground-truth labels are required. Instances can occur where a participant experiences cognitive load yet fails to answer a question. However, this did not affect labeling, as the protocol ensures that all selected data correspond to verified ground truths.

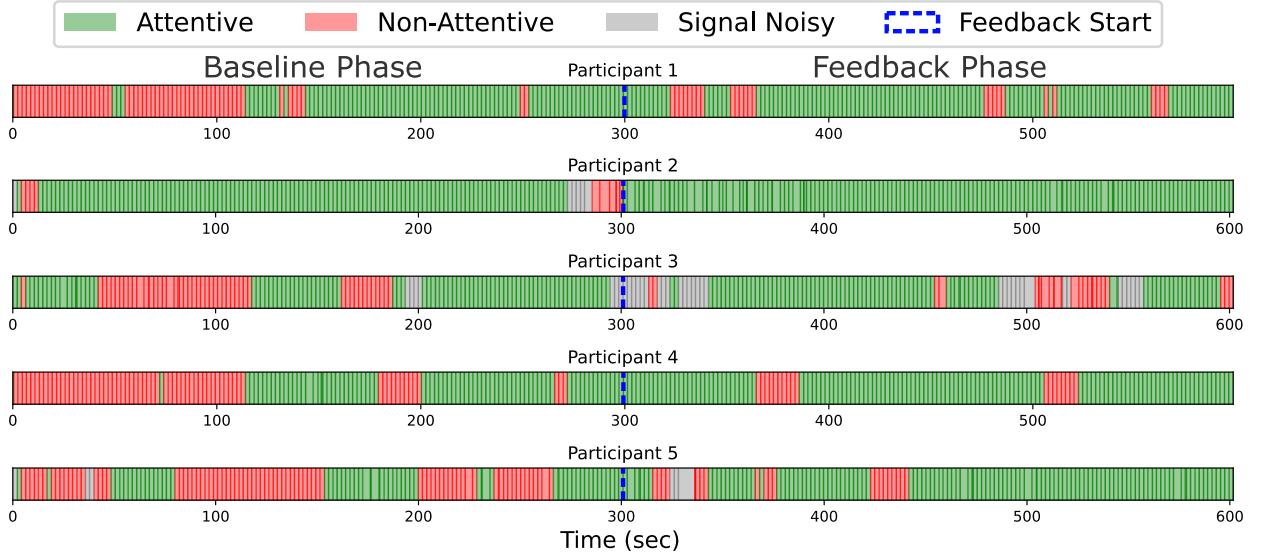
The methods were selected through extensive trials. For instance, both causal forward filtering and zero-phase second-order section (SOS) filtering were tested. While zero-phase SOS filtering was more effective at removing phase distortion, its non-causal nature performed poorly at segment edges. To address this, those edge portions were removed, ensuring both the absence of distortion and the independence from future data. Furthermore, to prevent data loss, the segmentation process used an overlapping ratio. For classification, SVM was initially selected due to its proven effectiveness in EEG signal analysis. Conceptually, SVM is viewed as a mechanism that iteratively identifies the most discriminative features in



(a) Graphical user interface (GUI) for real-time EEG data visualization and classification.



(b) Overview of the online learning pipeline, showing EEG signal acquisition, real-time classification, and neurofeedback to regain attention.



(c) Graphical visualization of cognitive states in two phases of conducted pilot study.

Fig. 5. Subfigures (a) and (b) represent the graphical user interface and the real-time working pipeline. Subfigure (c) denotes the changes between attentive and non-attentive states during the real-time system test on five participants.

a high-dimensional space, analogous to how traditional EEG analyses iteratively isolate regions or frequency bands of interest to distinguish cognitive states. An initial assessment using various machine learning models was conducted, and the SVM achieved the highest performance, which is consistent with observations reported in the literature.

After validating the data using LOSO, the optimal  $(C_2, \gamma)$  values were determined in two ways. The first method identified the most common  $(C_2, \gamma)$  pair that achieved the best metrics across all subjects. The second method selected the  $(C_2, \gamma)$  pair that yielded the highest metrics, regardless of its frequency among subjects. The mode pair,  $(C_2, \gamma) = (0.01, 0.5)$ , produced an average accuracy of 88.77%, whereas  $(C_2, \gamma) = (0.01, 0.1)$  achieved a slightly higher average accuracy of 89.63%. Despite the higher accuracy, the former pair was adopted for real-time testing due to its more frequent favorable performance, which makes the model generalized and objective. The LOSO validation results given in Fig. 4 show that while the model performed well on most subjects, it performed poorly on two. This indicates the model's robustness could be improved with additional data.

## VI. REAL-TIME SETUP

### A. Data-handling Procedure

The real-time system was implemented in Python using a circular buffer for continuous EEG data processing. Fixed-length segments of 1750 samples were extracted with a step size of 525 samples (70% overlap), resulting in a new classification output approximately every 2.1 seconds. Each segment was processed using noise removal, feature extraction, and classification by the trained model.

### B. Graphical User Interface Development

The graphical user interface (GUI), developed with PyQt6, displays real-time EEG signals, SDK metrics, the algorithm's prediction, and headband status. The GUI is shown in Fig. 5a. The system provides real-time feedback: an audio-visual alert is triggered after five consecutive non-attention segments. Similar warnings prompt the user to adjust the headband if contact is lost or if significant artifacts are detected. At the end of a session, a summary screen displays detailed attention statistics. This feedback loop is visualized in Fig. 5b, which aims to promote cognitive engagement.

### C. Pilot Evaluation

To preliminarily assess the effectiveness of the implemented feedback mechanism, a pilot study was conducted involving five participants. Each participant watched a 10-minute educational video. The video was split in two consecutive phases: a 5-minute baseline phase without feedback and a 5-minute feedback phase. During the baseline phase, the system continuously monitored cognitive states without alerts. For the feedback phase, the same monitoring was done and the system issued visual and audio alerts whenever a participant remained non-attentive as per the model for approximately 8.4 consecutive seconds (4 overlapped segments).

A paired t-test was conducted on the mean durations of non-attentive states in the two phases. At first, non-attentive segments with less than 8 seconds duration were discarded as the effect of feedback mechanism was being evaluated. The paired t-test yielded a t-statistic of 5.73, and a p-value of 0.007. Moreover, the mean duration of non-attention periods for the participants was found to be 37.50 seconds in the baseline phases and 14.97 seconds in the feedback phases. Both analyses indicated a significant improvement in concentration during the feedback phase, suggesting that the real-time alerts effectively supported participants in regaining and sustaining attention. Fig. 5c gives a graphical overview for better visualization, portraying the timeline of sustained non-attention in the baseline phase and the feedback phase.

## VII. CONCLUSION

As traditional methods like questionnaire and observations require manual interventions, they are not suitable for online learning. For this reason, an automated real-time attention classification with neurofeedback mechanism has been developed in this article. The methodology, based on data from the FocusCalm headband, employed advanced preprocessing, WPT-based feature extraction, and an SVM classifier. The proposed algorithm achieved an average accuracy of 88.77% on leave-one-subject-out validation. This result demonstrates the robustness of the model on unseen subjects and proves the feasibility of using consumer-grade BCI for real-time attention monitoring.

Despite these promising results, some limitations are acknowledged. The model's performance was validated on a limited participant pool. Furthermore, the optimal feature set includes proprietary scores from the Crimson SDK, which may limit the model's transparency and adaptability.

Future work will directly address these limitations. We will first focus on expanding the dataset to include a more diverse participant pool. Better and more robust methods will be developed to ensure better performance of the model without requiring proprietary Crimson SDK scores. The application will be analyzed through implementation in multiple use cases, followed by systematic result collection. The use of multi-channel, consumer-grade and ready-to-use EEG devices will be adopted for better spatial coverage to capture and classify visual, auditory, and task distractions alongside the current attention and non-attention states. Another possible implication is integrating this research with multi-modal inputs

such as facial expressions and behavioral logs, to evaluate the key dimensions of a class: learner engagement and content quality. This research thus paves the way for practical and scalable real-time BCI applications.

## ACKNOWLEDGMENTS

We gratefully acknowledge the Advanced Intelligent Multidisciplinary Systems Lab, under the Institute of Research, Innovation, Incubation & Commercialization at United International University, for their support and resources, Institute of Advanced Research for funding this project (Code No. UIU-IAR-02-2024-SE-39) and Onnorokom EdTech Limited for testing our project in their learning environment.

## REFERENCES

- [1] P. Qiao, X. Zhu, Y. Guo, Y. Sun, and C. Qin, "The development and adoption of online learning in pre- and post-covid-19: Combination of technological system evolution theory and unified theory of acceptance and use of technology," *Journal of Risk and Financial Management*, vol. 14, no. 4, p. 162, Apr. 2021. [Online]. Available: <http://dx.doi.org/10.3390/jrfm14040162>
- [2] J. D. Wammes and D. Smilek, "Examining the influence of lecture format on degree of mind wandering," *Journal of Applied Research in Memory and Cognition*, vol. 6, no. 2, p. 174–184, Jun. 2017. [Online]. Available: <http://dx.doi.org/10.1037/h0101808>
- [3] F. Camelia and T. L. J. Ferris, "Validation studies of a questionnaire developed to measure students' engagement with systems thinking," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 4, p. 574–585, Apr. 2018. [Online]. Available: <http://dx.doi.org/10.1109/TSMC.2016.2607224>
- [4] S. Patil, A. Jani, D. Jondhale, D. Pathak, and A. S. Savyanavar, "Student's attention and alertness monitoring system using AI," in *2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon)*. IEEE, Apr. 2024, p. 1–6. [Online]. Available: <http://dx.doi.org/10.1109/MITADTSoCiCon60330.2024.10575441>
- [5] M. K. Hossen and M. S. Uddin, "Attention monitoring of students during online classes using XGBoost classifier," *Computers and Education: Artificial Intelligence*, vol. 5, p. 100191, 2023. [Online]. Available: <http://dx.doi.org/10.1016/j.caeai.2023.100191>
- [6] Q. Liu, X. Jiang, and R. Jiang, "Classroom behavior recognition using computer vision: A systematic review," *Sensors*, vol. 25, no. 2, p. 373, Jan. 2025. [Online]. Available: <http://dx.doi.org/10.3390/s25020373>
- [7] S. A. L. Nair and R. K. Megalingam, "Human attention detection system using deep learning and brain-computer interface," *Neural Computing and Applications*, vol. 36, no. 18, p. 10927–10940, Mar. 2024. [Online]. Available: <http://dx.doi.org/10.1007/s00521-024-09628-8>
- [8] R. Wei, C. Hua, J. Chen, D. Mu, and J. Zhao, "Attention-based multiscale tCNN for SSVEP classification and its application to bionic intelligent soft gripper control," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, p. 1–10, 2025. [Online]. Available: <http://dx.doi.org/10.1109/TSMC.2025.3578533>
- [9] J. Wang, L. Bi, and W. Fei, "Multitask-oriented brain-controlled intelligent vehicle based on human-machine intelligence integration," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 4, p. 2510–2521, Apr. 2023. [Online]. Available: <http://dx.doi.org/10.1109/TSMC.2022.3212744>
- [10] J. Zhuang, K. Geng, and G. Yin, "Ensemble learning based brain-computer interface system for ground vehicle control," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 9, p. 5392–5404, Sep. 2021. [Online]. Available: <http://dx.doi.org/10.1109/TSMC.2019.2955478>
- [11] S. Lin, C. Fan, D. Han, Z. Jia, Y. Peng, and S. Kwong, "HATNet: EEG-based hybrid attention transfer learning network for train driver state detection," *IEEE Transactions on Cybernetics*, vol. 55, no. 5, pp. 2437–2450, 2025.
- [12] C.-T. Lin and T.-T. N. Do, "Direct-sense brain-computer interfaces and wearable computers," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, p. 298–312, Jan. 2021. [Online]. Available: <http://dx.doi.org/10.1109/TSMC.2020.3041382>

- [13] W. Dang *et al.*, “Flashlight-Net: A modular convolutional neural network for motor imagery EEG classification,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 54, no. 7, p. 4507–4516, Jul. 2024. [Online]. Available: <http://dx.doi.org/10.1109/TSMC.2024.3382828>
- [14] N. Kaongoen, M. Yu, and S. Jo, “Two-factor authentication system using P300 response to a sequence of human photographs,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 3, p. 1178–1185, Mar. 2020. [Online]. Available: <http://dx.doi.org/10.1109/TSMC.2017.2756673>
- [15] M. Du, Y. Yao, J. Xiang, X. Chen, Y. Jin, K. Zhang, G. Bao, and S. Cai, “A method for improving online active engagement during lower limb rehabilitation training based on eeg signals,” *IEEE Transactions on Human-Machine Systems*, vol. 55, no. 4, pp. 650–659, 2025.
- [16] D.-H. Lee, J.-H. Jeong, B.-W. Yu, T.-E. Kam, and S.-W. Lee, “Autonomous system for EEG-based multiple abnormal mental states classification using hybrid deep neural networks under flight environment,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 10, p. 6426–6437, Oct. 2023. [Online]. Available: <http://dx.doi.org/10.1109/TSMC.2023.3282635>
- [17] S. Cai, P. Li, E. Su, Q. Liu, and L. Xie, “A neural-inspired architecture for eeg-based auditory attention detection,” *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 4, pp. 668–676, 2022.
- [18] Y. Liu, Y. Yu, Z. Ye, M. Li, Y. Zhang, Z. Zhou, D. Hu, and L.-L. Zeng, “Fusion of spatial, temporal, and spectral eeg signatures improves multilevel cognitive load prediction,” *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 2, pp. 357–366, 2023.
- [19] N. Z. Zenia, S. Targ, L. G. Dizaji, and Y. Hu, “Eeg features to quantify the nasa-tlx factors of cognitive workload,” *IEEE Transactions on Human-Machine Systems*, vol. 55, no. 3, pp. 372–382, 2025.
- [20] Y.-W. Choi, H.-B. Shin, and S.-W. Lee, “Brain-guided self-paced curriculum learning for adaptive human-machine interfaces,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 55, no. 7, p. 4693–4704, Jul. 2025. [Online]. Available: <http://dx.doi.org/10.1109/TSMC.2025.3560832>
- [21] T. K. Reddy, V. Arora, V. Gupta, R. Biswas, and L. Behera, “EEG-based drowsiness detection with fuzzy independent phase-locking value representations using Lagrangian-based deep neural networks,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 1, p. 101–111, Jan. 2022. [Online]. Available: <http://dx.doi.org/10.1109/TSMC.2021.3113823>
- [22] P. K. Sahu and K. Jain, “Sustained attention detection in humans using a prefrontal theta-eeg rhythm,” *Cognitive Neurodynamics*, vol. 18, no. 5, p. 2675–2687, May 2024. [Online]. Available: <http://dx.doi.org/10.1007/s11571-024-10113-0>
- [23] P. Kaushik, A. Moye, M. v. Vugt, and P. P. Roy, “Decoding the cognitive states of attention and distraction in a real-life setting using eeg,” *Scientific Reports*, vol. 12, no. 1, Nov. 2022. [Online]. Available: <http://dx.doi.org/10.1038/s41598-022-24417-w>
- [24] S. Zhang, Z. Yan, S. Sapkota, S. Zhao, and W. T. Ooi, “Moment-to-moment continuous attention fluctuation monitoring through consumer-grade eeg device,” *Sensors*, vol. 21, no. 10, p. 3419, May 2021. [Online]. Available: <http://dx.doi.org/10.3390/s21103419>
- [25] L.-W. Ko, O. Komarov, W. D. Hairston, T.-P. Jung, and C.-T. Lin, “Sustained attention in real classroom settings: An eeg study,” *Frontiers in Human Neuroscience*, vol. 11, Jul. 2017. [Online]. Available: <http://dx.doi.org/10.3389/fnhum.2017.00388>
- [26] P. Sawangjai, S. Hompoonsup, P. Leelaarporn, S. Kongwudhikunakorn, and T. Wilaprasitporn, “Consumer grade EEG measuring sensors as research tools: A review,” *IEEE Sensors Journal*, vol. 20, no. 8, p. 3996–4024, Apr. 2020. [Online]. Available: <http://dx.doi.org/10.1109/JSEN.2019.2962874>
- [27] J. Sabio, N. S. Williams, G. M. McArthur, and N. A. Badcock, “A scoping review on the use of consumer-grade eeg devices for research,” *PLOS ONE*, vol. 19, no. 3, p. e0291186, Mar. 2024. [Online]. Available: <http://dx.doi.org/10.1371/journal.pone.0291186>
- [28] S. Sudharsan, S. Siddharth, M. Uma, and R. Kaviyaraj, “Learning behavior analysis for personalized E-Learning using EEG signals,” in *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, 2024, pp. 1–9.
- [29] G. Ruqeyya and S. M. U. Saeed, “Player engagement classification in mobile games using MUSE headband,” in *2022 17th International Conference on Emerging Technologies (ICET)*, 2022, pp. 83–88.
- [30] M. Beiramvand, M. Shahbakhti, N. Karttunen, R. Koivula, J. Turunen, and T. Lipping, “Assessment of mental workload using a Transformer network and two prefrontal EEG channels: An unparameterized approach,” *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–10, 2024.
- [31] A. Arsalan, M. Majid, A. R. Butt, and S. M. Anwar, “Classification of perceived mental stress using a commercially available EEG headband,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2257–2264, 2019.
- [32] H. Huang *et al.*, “Real-time attention regulation and cognitive monitoring using a wearable EEG-based BCI,” *IEEE Transactions on Biomedical Engineering*, vol. 72, no. 2, p. 716–724, Feb. 2025. [Online]. Available: <http://dx.doi.org/10.1109/TBME.2024.3468351>
- [33] I. H. Robertson, T. Manly, J. Andrade, B. T. Baddeley, and J. Yiend, “‘Oops!’: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects,” *Neuropsychologia*, vol. 35, no. 6, p. 747–758, May 1997. [Online]. Available: [http://dx.doi.org/10.1016/s0028-3932\(97\)00015-8](http://dx.doi.org/10.1016/s0028-3932(97)00015-8)
- [34] C. Conrad and A. Newman, “Measuring mind wandering during online lectures assessed with eeg,” *Frontiers in Human Neuroscience*, vol. 15, Aug. 2021. [Online]. Available: <http://dx.doi.org/10.3389/fnhum.2021.697532>
- [35] A. Ur Rehman, N. Ul Eman, F. Ullah, D. Cacciagrano, L. Mostarda, and N. Raza, “Advanced EEG signal processing and deep Q-Learning for accurate student attention monitoring,” *IEEE Access*, vol. 13, p. 47260–47270, 2025. [Online]. Available: <http://dx.doi.org/10.1109/ACCESS.2024.3523225>
- [36] K. Flanagan and M. J. Saikia, “Consumer-grade electroencephalogram and functional near-infrared spectroscopy neurofeedback technologies for mental health and wellbeing,” *Sensors*, vol. 23, no. 20, p. 8482, Oct. 2023. [Online]. Available: <http://dx.doi.org/10.3390/s23208482>
- [37] FocusCalm, “Clinics & consultants,” [Online]. Available: <https://focuscalm.com/pages/clinics-consultants>, [Accessed: 18-08-2025].
- [38] J. Peirce *et al.*, “PsychoPy2: Experiments in behavior made easy,” *Behavior Research Methods*, vol. 51, no. 1, p. 195–203, Feb. 2019. [Online]. Available: <http://dx.doi.org/10.3758/s13428-018-01193-y>
- [39] B. Kapteijns and F. Hintz, “Comparing predictors of sentence self-paced reading times: Syntactic complexity versus transitional probability metrics,” *PLOS ONE*, vol. 16, no. 7, p. e0254546, Jul. 2021. [Online]. Available: <http://dx.doi.org/10.1371/journal.pone.0254546>
- [40] R. Ley and D. Locascio, “Study time, recall time, meaningfulness, and recognition latency in paired-associate learning,” *Psychonomic Science*, vol. 28, no. 2, p. 88–90, Feb. 1972. [Online]. Available: <http://dx.doi.org/10.3758/BF03328671>
- [41] J. D. Eastwood, A. Frischen, M. J. Fenske, and D. Smilek, “The unengaged mind: Defining boredom in terms of attention,” *Perspectives on Psychological Science*, vol. 7, no. 5, p. 482–495, Sep. 2012. [Online]. Available: <http://dx.doi.org/10.1177/1745691612456044>
- [42] S. Bench and H. Lench, “On the function of boredom,” *Behavioral Sciences*, vol. 3, no. 3, p. 459–472, Aug. 2013. [Online]. Available: <http://dx.doi.org/10.3390/bs3030459>
- [43] J. Smallwood and J. W. Schooler, “The science of mind wandering: Empirically navigating the stream of consciousness,” *Annual Review of Psychology*, vol. 66, no. 1, p. 487–518, Jan. 2015. [Online]. Available: <http://dx.doi.org/10.1146/annurev-psych-010814-015331>
- [44] P. Saideepthi, A. Chowdhury, P. Gaur, and R. B. Pachori, “Sliding window along with EEGNet-based prediction of EEG motor imagery,” *IEEE Sensors Journal*, vol. 23, no. 15, p. 17703–17713, Aug. 2023. [Online]. Available: <http://dx.doi.org/10.1109/JSEN.2023.3270281>
- [45] Z. He, K. Yang, N. Zhuang, and Y. Zeng, “Processing of affective pictures: A study based on functional connectivity network in the cerebral cortex,” *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, Jan. 2021. [Online]. Available: <http://dx.doi.org/10.1155/2021/5582666>
- [46] P. S. Ledwidge, C. N. McPherson, L. Faulkenberg, A. Morgan, and G. C. Baylis, “A comparison of approaches for motion artifact removal from wireless mobile EEG during overground running,” *Sensors*, vol. 25, no. 15, p. 4810, Aug. 2025. [Online]. Available: <http://dx.doi.org/10.3390/s25154810>
- [47] X. Jiang, G.-B. Bian, and Z. Tian, “Removal of artifacts from EEG signals: A review,” *Sensors*, vol. 19, no. 5, p. 987, Feb. 2019. [Online]. Available: <http://dx.doi.org/10.3390/s19050987>
- [48] Z. WU and N. E. HUANG, “Ensemble empirical mode decomposition: A noise-assisted data analysis method,” *Advances in Adaptive Data Analysis*, vol. 01, no. 01, p. 1–41, Jan. 2009. [Online]. Available: <http://dx.doi.org/10.1142/S1793536909000047>
- [49] P. Virtanen *et al.*, “SciPy 1.0: Fundamental algorithms for scientific computing in Python,” *Nature Methods*, vol. 17, no. 3, p. 261–272, Feb. 2020. [Online]. Available: <http://dx.doi.org/10.1038/s41592-019-0686-2>