Optimal EEG Time Window Length for Boredom Classification using Combined Non-linear Features

Rajamanickam Yuvaraj[*,1], S Samyuktha[1], Jack Fogarty[1], Jun Song Huang[1], Samuel Tan[2] and, Teck Kiong Wong[2]
[1]Science of Learning in Education Centre, National Institute of Education, Nanyang Technological University, Singapore.
[2]Education Technology Division (ETD), Ministry of Education (MOE), Singapore 138675.

***Abstract*—In this paper, we investigate the effect of different time window lengths on boredom detection to find the optimal length for extracting EEG features related to boredom. Continuous EEG was recorded from 9 healthy participants (mean age = 31.8, 5 female) as they watched an education video stimulus to induce boredom. A range of non-linear features, such as fractal dimension, entropy, and chaotic features, were computed over time windows of 0.5, 1, 2, 3, 4, 5, 6, 10, 15, 20, 25, and 30 seconds. Feature combination was achieved with the help of feature selection using the Gini impurity score. Classifiers, namely, XG Boost, random forest (RF), and multilayer perceptron were employed to assess boredom classification accuracy systematically and compared the signal features performance calculated in different time windows. The proposed framework yields the highest mean accuracy of 88.61%±0.10 with 5-fold cross-validation (CV) and leave-one-subject-out (LOSO) CV accuracy of 88.92%±0.25 when using a 0.5-sec window length based on the combined feature set and RF classifier. These results provide a clear reference for selecting epoch duration when analyzing EEG signals related to boredom, an important methodological consideration for any EEG-driven emotion recognition systems.**

***Index Terms*—boredom, education, learning, EEG, emotion, machine learning**

I. Introduction

Emotions play a major role in decision-making, learning, and other aspects of human behaviour [1]. Accordingly, measuring emotions in educational settings can offer significant information explaining students' learning outcomes. Student emotions in class, for instance, can indicate how they feel about course material and their engagement, which is associated with academic performance [2]. Feedback regarding students' emotional responses during class may also be used to develop and optimize the learner's experience [2], [3]. Thus, exploring students' emotions as they learn can be of substantial value to teachers and students.

Among the various emotions that have been targeted in the education field, there is one emotion, "boredom", which has gained far less attention from educational researchers. Several studies showed that boredom can have a negative impact on the efficiency of learning [4]. Therefore, intervention is needed to sustain students' attention on the learning content and promote their learning efficiency. Evidence shows that feedback regarding boredom states has been effective in better regulating their learning and contributes to student's well-being. For instance, if a computer-driven system recognizes that the learner is experiencing boredom, it can alter the learning material to engage and motivate the learner. This, in turn, reduces the level of boredom. Thus, using this countermeasure, we could expect the learner to focus on the learning content.

In the literature, numerous automated emotion recognition methods have been published using various modalities, including physiological signals (e.g., electroencephalogram (EEG), electrocardiogram (ECG), and galvanic skin resistance) and behavioral data (e.g., facial expressions). While all these modalities have their specific advantages and disadvantages, as [5] suggest, physiological measurements are considered the most objective. Among other physiological measurements, EEG provides real-time measures of the working brain and does not require an overt response from the participant. From a physiological perspective, EEG signals record the brain's electrical activity, which belongs to the central nervous system and is deeply related to cognition, including emotion processing. Despite the large number of studies conducted on EEG-based emotion recognition, very few researchers have tried to build boredom detection models using EEG data and machine learning (ML) approaches [6]–[9]. Shen et al. [6] used a non-overlapped 1-sec (256 samples/sec) for extracting EEG frequency sub-bands to detect several target emotions, with one of them being boredom, which resulted in an accuracy of 67.10% for boredom detection. Kim et al. [7] took EEG data of 1 sec (220 samples per sec) and applied a thresholding technique to classify boredom and non-boredom stages. Seo et al. [8] set the window length to 7 sec (220 samples per sec) without overlapping on channels of the EEG data and obtained the best mean boredom detection accuracy of 79.98%. In another study, Seo et al. [9] applied a non-overlap 1-sec EEG time window (220 samples per sec) for extracting features to perform boredom recognition and achieved the highest accuracy of 86.73%. From these studies, it can be observed that a window length of 1 second was used in most research to date; however, the impact of this methodological choice on boredom emotion detection has not been explored systematically. There is currently no criterion or prior knowledge on the temporal scale (i.e., time window length) to measure and classify EEG data for boredom emotion recognition. Moreover, the impact of window length may also be critical given the temporal dynamics of emotion processing and response stages can vary in the order of milliseconds, seconds, and even minutes [10];

This research was financed by the Singapore Ministry of Education (MOE) through the Education Research Funding Programme (ERFP) (Grant No: OER 14/22 YR), which was overseen by the National Institute of Education (NIE), Nanyang Technological University (NTU), Singapore. *Corresponding author email: yuvaraj.rajamanickam@nie.edu.sg

it is unclear then if short or long window lengths are more appropriate for capturing temporal dynamics in emotions and what will be most effective for boredom in learning contexts.

The present study aims to determine an optimal EEG time window for boredom detection in learning contexts. To attain this, a range of non-linear features are examined to identify the most significant and generalizable EEG window length distinguishing boredom and non-boredom states. Emotion classification accuracy will be assessed systematically using a range of signal features calculated in time windows of 0.5 sec, 1 sec, 2 sec, 3 sec, 4 sec, 5 sec, 6 sec, 10 sec, 15 sec, 20 sec, 25 sec, and 30 sec. The non-linear feature sets that are studied include fractal dimension (FD), entropy, and chaotic features. Feature selection is performed using Gini importance score. The performance of three common classifiers namely, random forest (RF), XGBoost (XGB), and multilayer perceptron (MLP) were assessed; Accuracy (AC), sensitivity (SE), specificity (SP), F1-score (F1-S), and area under curve (AUC) serves as the main performance metrics in this investigation. In this way, we aim to recommend the most useful and generalizable EEG time window length for detecting boredom state in a learning context and to guide the future development of boredom-aware systems.

II. Materials

A. Participants and physiological signals acquisition

Nine healthy pre-service teachers (right-handed, mean age = 31.8, SD = 6.5; 5 females, 4 males) from Singapore's National Institute of Education (NIE) volunteered to participate in this study. All participants provided their informed consent prior to participating. They reported normal or corrected to normal hearing and vision and had no history of neurological difficulties or head trauma. All procedures were approved by the Institutional Review Board (IRB) at the Nanyang Technological University (NTU), Singapore.

Continuous EEG, electrocardiogram (ECG), galvanic skin resistance (GSR), and eye-gaze data were recorded at 1000Hz using an AntNeuro EEGO sports amplifier referenced to CPz. The raw EEG signals were collected from 64 scalp channels, including the left and right mastoids. Only EEG data were extracted and analyzed in this study.

B. Emotion stimulus and experiment protocol

After the EEG set up, participants were seated in a sound-suppressed room to watch emotion induction videos presented via E-Prime. To set a baseline state of non-boredom, participants first watched a short 4.2-minute video taken from the BBC documentary Planet Earth [11], showing colorful scenes of marine life accompanied by narration and music; this control video was identified as a useful baseline for boredom in [11]. After the baseline video, a 10-minute-long educational video was shown to the participants, involving a science teacher explaining input/output energy concepts in standard lecture format. This video was selected after a pilot was conducted to ensure the video reliably induced strong feelings of boredom in adults. Participants were required to sit still and watch the boredom video for a minimum of 6 minutes, after which they were free to end the video and move on; this was to maximize data collection while ensuring a minimum of 6 minutes of EEG data related to a boring video stimulus. In our analysis, among the nine preservice teachers, the shortest play time for boredom video was 7.08 minutes, and the average play time was 8.82 minutes. At the end of the baseline and boredom videos, the participants completed a state affect questionnaire (SAQ) to indicate what emotions they felt (including excitement, disgust, neutral, boredom, and distress) and the intensity of those emotions on the scale of 0-8. Ground truth was based on the subject's answer to the boredom SAQ.

III. Method

A. EEG Signal preprocessing and time window lengths

Raw EEGs were first down-sampled to 256 Hz and re-referenced to a common average before applying a 50 Hz notch filter to eliminate electrical noise. Electrooculogram (EOG) correction was then applied utilizing a wavelet denoising technique based on thresholding [12]; Daubechies wavelet (db9) with level 6 was chosen as mother wavelet, and Stein's Unbiased Risk Estimate (SURE) thresholding algorithm was used [12]. EEG data during the first minute of the boredom video condition was excluded from the analysis to minimize variability in the latency of emotion induction.

Preprocessed continuous EEG data were segmented into non-overlapping time window lengths for analysis. A total of 12 different time window lengths were examined to identify the optimal length to achieve the highest boredom classification accuracy. The window lengths 0.5, 1, 2, 3, 4, 5, and 6 seconds were chosen based on the existing emotion recognition studies using EEG [13]. To refine the result of the investigation, additional window lengths (10, 15, 20, 25, and 30) were also included with the existing lengths.

B. Discrete Wavelet Transform

Discrete wavelets transform (DWT), a popular technique for analyzing signals in the time-frequency domain that decomposes EEG signals in several approximations (A) and details (D) levels of wavelet coefficients corresponding to subbands of EEG (i.e., delta, theta, alpha, beta and gamma). Using Daubechies wavelet, six levels of decomposition were performed on each pre-processed-segmented epoch from each channel to extract delta (0-4Hz – A6), theta (4-8 Hz – D6), alpha (8-16 Hz-D5), beta (16-32 Hz-D4), and gamma (32-64 Hz-D3) sub-bands of EEG. All the decomposed band coefficients were then applied to feature generation.

C. Non-linear feature generation

In this study, non-linear measures such as fractal dimension, entropy, and chaotic features were extracted from each sub-band wavelet coefficient for boredom and non-boredom classification. FD quantifies the dynamic complexity of EEG signals by exploiting their stochastic nature. This study employed two commonly used FDs for EEG signal characterization, namely

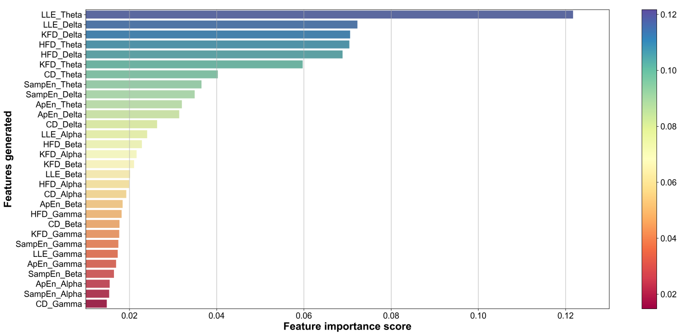

Fig. 1. Feature ranking based on Gini impurity score for 0.5 second EEG time window length

Higuchi's fractal dimension (HFD) and Katz's fractal dimension (KFD) [14]. Entropy is another non-linear feature that measures randomness in a signal. We calculated approximate entropy (ApEn) and sample entropy (SampEn)in this study, which has been successfully applied to EEG applications [13]. The chaotic features calculated are the correlation dimension, which measures complexity, and the largest Lyapunov exponent (LLE), which measures the chaos of the signal. A description of the procedure for calculating these non-linear features can be found in [13], [14]. These non-linear features were computed for each EEG segment of five bands, resulting in an N-dimensional feature vector i.e., 6 (sub-bands) $\times$ no. EEG segments per channel $\times$ 61 (channels).

D. Feature selection via Gini impurity score

Feature selection is a technique to select the desired subset of features, reduce the feature space dimension, and improve pattern recognition model classification performance [15]. Implicit feature selection using tree methods like a random forest classifier can be visualized as a plot indicating the relative importance of features. The resultant Gini impurity score is a by-product of binary splits, based on the highest difference in scores at each tree node, and gives a general indication of feature relevance.

E. Boredom classification and performance evaluation

To compare the success of the explored non-linear EEG features for various time window lengths, three classical classifiers were employed for boredom and non-boredom classification, including XGB, MLP, and RF. The performance of each classification model was measured in terms of AC, SE, SP, and F1-S. Along with these performance metrics, the receiver operating characteristics (ROC) curve and area under the ROC curve (AUROC) were also computed to evaluate the model. Two different cross-validation methods were adopted to evaluate the robustness of the proposed classifier models. The five-fold cross-validation (CV) method was utilized to obtain consistent recognition performance, where the trained model contains a mixed amount of data from all subjects, 4-fold for training in each iteration, and the remaining fold was used for testing. The other method used for testing the results is the leave-one-subject-out cross-validation (LOSO-CV). In LOSO-CV, the subject folds used for testing are not considered for training, which better evaluates performance in testing new subjects. In both methods, the distribution of the features for the non-boredom and boredom classes was kept equal over each fold (i.e., balanced classification). Classifier model training involved 90% of each training fold data for initial learning and the remaining 10% for validation. For each time window length, the final AC, SE, SP, F1-S, and AUC were averaged with standard deviation (SD) across the folds to compute the overall boredom performance of the given classifier model.

IV. Experimental results and discussion

In this section, the performance of the different EEG time window lengths to classify boredom and non-boredom based on EEG signals is evaluated in several aspects. In the first part, the importance of individual features using Gini score measurement was analyzed. In the second part, the obtained 5-fold-CV numerical results graph for each feature set, including the combination of all non-linear feature sets (i.e., combined), different window lengths, and the various classifiers in the proposed framework, are provided. Finally, LOSO-CV execution on the most successful EEG time window length with the corresponding feature set is reported.

Fig. 1 shows features ranked in the order of discriminating capability. For brevity, only the feature importance score for the top-performing time window length is reported here. Feature importance scores are used to understand the relative importance of features and only features that score higher than the cut-off is considered meaningful. The shape of the graph can determine the cut-off, and the ones that score lowest are those considered to have fewer trees and no significant discriminatory power. Based on a visual inspection of the feature importance score plot, a cutoff of five was chosen to remove the lower five features. Accordingly, the sub-band entropy features are shown to be of the least importance, whereas the remaining features are of reasonable importance. Overall, FDs were found to carry substantial discriminative power relative to all other methods.

Fig. 2(a)-2(d) shows the average boredom detection accuracy using different time window lengths, EEG feature sets, and classifiers. Visual inspection shows that shorter window lengths generally showed higher performance across all feature sets, irrespective of the classifier used. Indeed, the highest average boredom detection accuracy was achieved when using a 0.5-sec window length (AC = $88.61\% \pm 0.10\%$), relative to other window lengths when using either FD, entropy, chaotic, or combined feature sets. This outcome may be driven by sample size, as in general, a longer length of window will lead to fewer samples for analysis. Shorter window lengths will expand the input scale of features from the temporal dimension, which is expected to be advantageous in capturing transient changes in EEG and emotion processing while also increasing sample size. Employing a longer window length will undermine reading temporal EEG data, while using a

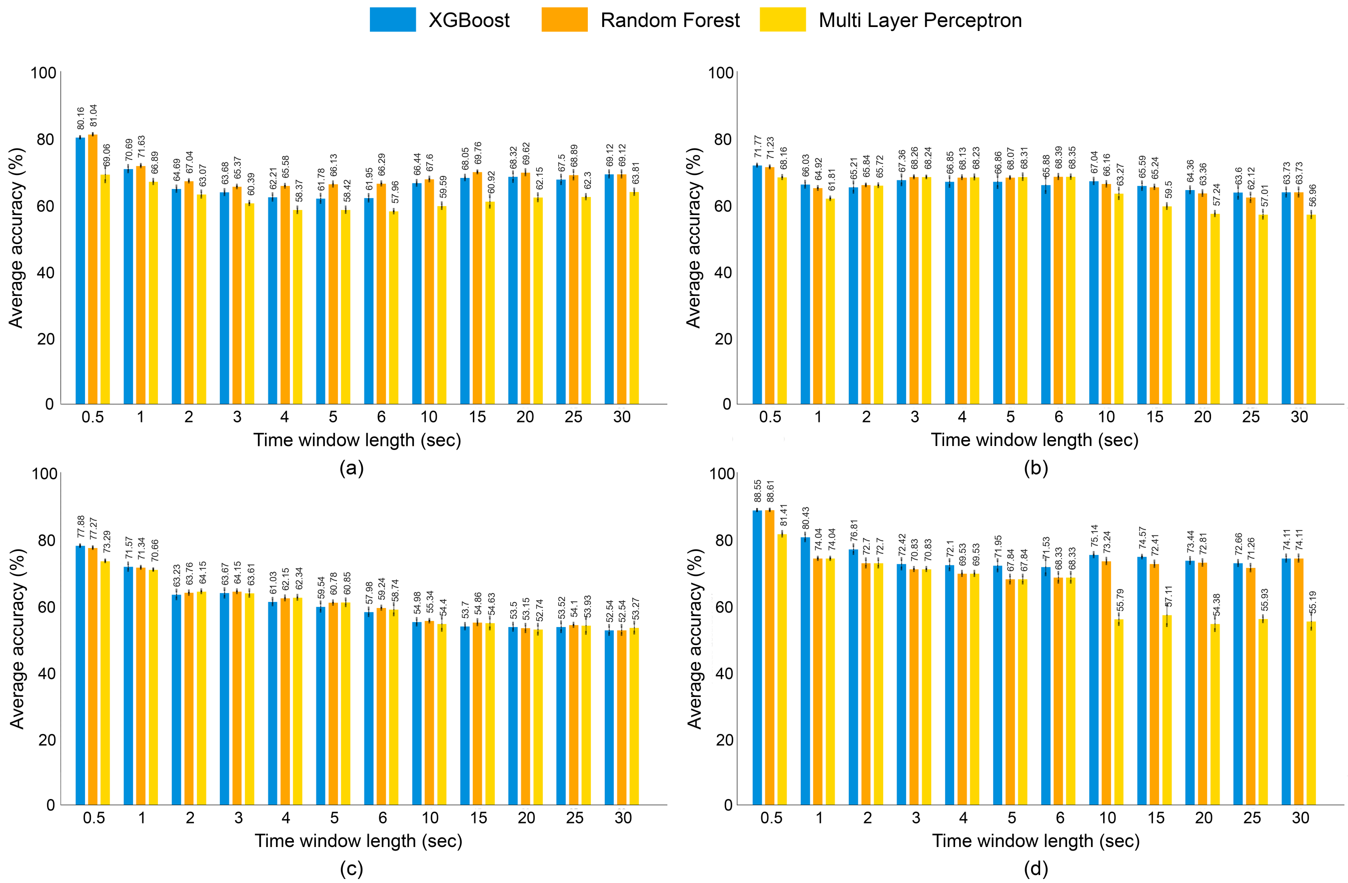

Fig. 2. Comparison of average boredom classification results using different time window lengths, EEG feature sets, and classifiers in terms of the % of AC. (a) Fractal dimension, (b) Entropy, (c) Chaotic, and (d) Combined feature set.

shorter length will extend the computing time that is inconvenient to online affective computing. Given that different window lengths have different advantages and disadvantages, a suitable time window length that can balance the contradiction between them is required, and our results show that the 0.5-sec EEG time window length is an optimal choice for boredom detection in education settings or learning contexts.

As demonstrated by the average classification accuracy across feature sets in Fig. 2(a) - 2(d), the combined feature set performed higher for classifying boredom and non-boredom states relative to other features when using RF, XGB, or MLP classifiers. This outcome is consistent with past research showing that a combination of features integrates strengths and reduces the weaknesses of individual features, resulting in performance improvement [14]. Furthermore, the combined feature set achieved classification results with the lowest SD of accuracy, showing they perform more consistently than other feature sets in this study. The results also suggest that all three sets of features (FD, entropy, and chaotic) used here carry discriminative and diverse information that is capable of reliably detecting boredom in the education context. Another important finding of the present study is that the RF classifier performed better for boredom classification compared to XGB and MLP. Using the combined feature set, we obtained the highest mean accuracy of $88.61\% \pm 0.10\%$ with the RF classifier. This was found to be the case across all time windows and feature sets utilized in this study and in line with previous research supporting the utility of RF for emotion recognition [14]. Given the 0.5 window length was most successful, the subsequent results primarily focus on outcomes for the 0.5-time windows with the combined feature set and RF classifier.

Fig. 3 shows the mean ROC curve and the mean AUROC value across the folds. The SD of AUROC values shows that the AUROC deviation for each fold is much less (< 0.02), and the average value of AUROC is 0.950. These results indicate that the proposed framework performs well in detecting boredom from EEG signals. Table I presents the boredom detection results obtained with the LOSO-CV. While combining feature set and RF classifier, the 0.5-sec long EEG segments achieved an average boredom detection performance of AC = $88.49\%\pm0.24\%$, SE = $83.26\%\pm0.42\%$, SP = $93.73\%\pm0.25\%$, F1-S = $87.86\% \pm 0.27\%$, and AUC = $94.97\% \pm 0.15\%$. This performance demonstrates the robustness of the developed boredom detection model evaluated the accuracy in testing

TABLE I
BOREDOM CLASSIFICATION RESULTS OBTAINED WITH THE LOSO-CV FOR 0.5 SEGMENTS USING COMBINED FEATURES AND RF CLASSIFIER

Subject ID	Performance metrics (%)				
	AC	**SE**	**SP**	**F1-S**	**AUROC**
1	89.05	83.87	94.23	88.45	95.36
2	89.32	84.51	94.13	88.78	95.59
3	88.66	83.18	94.14	88	94.81
4	88.88	83.94	93.81	88.3	95.4
5	89.15	83.96	94.33	88.55	95.32
6	88.65	83.42	93.87	88.02	95.17
7	88.6	83.03	94.18	87.93	95.09
8	89.21	84.25	94.18	88.65	95.42
9	88.77	83.45	94.09	88.14	95.25
Average	**88.92± 0.25**	**83.73± 0.47**	**94.11± 0.16**	**88.31± 0.29**	**95.27± 0.21**

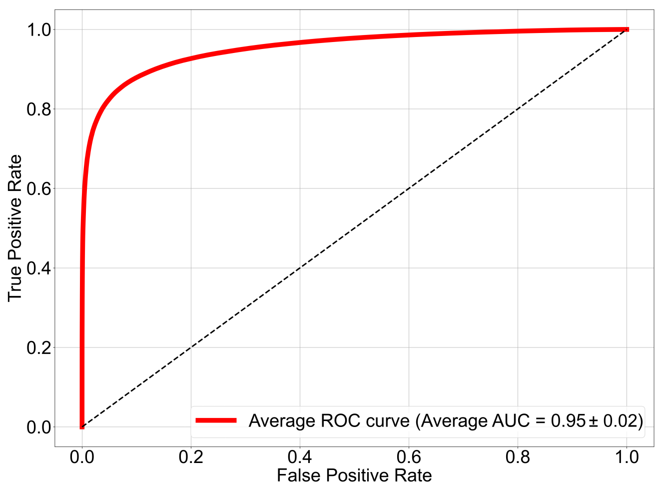

Fig. 3. Average ROC curve for 5-fold-CV

new subjects compared to 5-fold-CV methods.

V. CONCLUSION

In this preliminary study, the effectiveness of different time window lengths of EEG on boredom detection in a learning context was investigated. For this purpose, nine preservice teachers volunteered for this ongoing study and watched a non-boring and boring lecturer video clip while EEG was recorded. Non-linear features were extracted from the pre-processed EEG signals using FD, entropy, and chaotic methods and quantitative comparison of feature sets with three different classifiers, XGB, RF, and MLP was conducted. To select the best subset of the extracted features and improve the discrimination performance, Gini importance score was utilized. Our experimental results revealed that 0.5-sec time window length is the most optimal size in detecting boredom emotion categorized in terms of boredom and non-boredom state. With 0.5-sec window length, the combined feature set and RF classifier method achieves the highest mean accuracy of $88.61\% \pm 0.10\%$. The study findings are a further step towards developing applications that may be utilized to monitor and provide feedback related to learning emotions, which could be used to guide pedagogical design and self-regulated learning. However, this is an ongoing study, and we aim to recruit a larger sample size to strengthen the findings' reliability. This could help us create future boredom-aware systems in education and other areas (e.g., entertainment). We will also explore the neurophysiological mechanism of boredom with a larger sample. This could lead to a deeper science of learning insights to boredom in education contexts. Furthermore, we seek to analyze the other physiological signals (namely GSR, ECG, and eye-gaze) and SAQ data that are recorded (but not used in this study) and find a way to develop a multimodal boredom detection system that could help to improve our understanding of boredom.

REFERENCES

[1] A. Bechara, H. Damasio, and A. R. Damasio, "Emotion, decision making and the orbitofrontal cortex," *Cerebral cortex*, vol. 10, no. 3, pp. 295–307, 2000.

[2] M. R. Reyes, M. A. Brackett, S. E. Rivers, M. White, and P. Salovey, "Classroom emotional climate, student engagement, and academic achievement." *Journal of educational psychology*, vol. 104, no. 3, p. 700, 2012.

[3] Y. Rajamanickam, "Electroencephalogram based emotion recognition in parkinson's disease using non-linear methods," Ph.D. dissertation, School of Mechatronic Engineering, Universiti Malaysia Perlis, 2015.

[4] L. S. Kanevsky, "A comparative study of children's learning in the zone of proximal development," *European Journal of High Ability*, vol. 5, no. 2, pp. 163–175, 1994.

[5] J. Kim and D. R. Fesenmaier, "Measuring emotions in real time: Implications for tourism experience design," *Journal of Travel Research*, vol. 54, no. 4, pp. 419–429, 2015.

[6] L. Shen, M. Wang, and R. Shen, "Affective e-learning: Using "emotional" data to improve learning in pervasive learning environment," *Journal of Educational Technology & Society*, vol. 12, no. 2, pp. 176–189, 2009.

[7] J. Kim, J. Seo, and T. H. Laine, "Detecting boredom from eye gaze and eeg," *Biomedical Signal Processing and Control*, vol. 46, pp. 302–313, 2018.

[8] J. Seo, T. H. Laine, and K.-A. Sohn, "An exploration of machine learning methods for robust boredom classification using eeg and gsr data," *Sensors*, vol. 19, no. 20, p. 4561, 2019.

[9] S. Jungryul, T. H. Laine, and K.-A. Sohn, "Machine learning approaches for boredom classification using eeg," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, pp. 3831–3846, 2019.

[10] P. A. Kragel, A. R. Hariri, and K. S. LaBar, "The temporal dynamics of spontaneous emotional brain states and their implications for mental health," *Journal of cognitive neuroscience*, vol. 34, no. 5, pp. 715–728, 2022.

[11] C. Merrifield and J. Danckert, "Characterizing the psychophysiological signature of boredom," *Experimental brain research*, vol. 232, pp. 481–491, 2014.

[12] N. G. Seshadri, B. K. Singh, and R. B. Pachori, "Eeg based functional brain network analysis and classification of dyslexic children during sustained attention task," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 4672–4682, 2023.

[13] R. Yuvaraj, M. Murugappan, N. M. Ibrahim, K. Sundaraj, M. I. Omar, K. Mohamad, and R. Palaniappan, "Optimal set of eeg features for emotional state classification and trajectory visualization in parkinson's disease," *International Journal of Psychophysiology*, vol. 94, no. 3, pp. 482–495, 2014.

[14] R. Yuvaraj, P. Thagavel, J. Thomas, J. Fogarty, and F. Ali, "Comprehensive analysis of feature extraction methods for emotion recognition from multichannel eeg recordings," *Sensors*, vol. 23, no. 2, p. 915, 2023.

[15] B. H. Menze, B. M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, and F. A. Hamprecht, "A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data," *BMC bioinformatics*, vol. 10, pp. 1–16, 2009.