

# Pilot Study: The Use of Electroencephalogram to Measure Attentiveness towards Short Training Videos

Paul Alton Nussbaum and Rosalyn Hobson Hargraves  
Department of Electrical and Computer Engineering  
Virginia Commonwealth University  
Richmond, VA USA

**Abstract**—Universities, schools, and training centers are seeking to improve their computer-based [3] and distance learning classes through the addition of short training videos, often referred to as podcasts [4]. As distance learning and computer based training become more popular, it is of great interest to measure if students are attentive to recorded lessons and short training videos.

The proposed research presents a novel approach to this issue. Signal processing of electroencephalogram (EEG) has proven useful in measuring attentiveness in a variety of applications such as vehicle operation and listening to sonar [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15]. Additionally, studies have shown that EEG data can be correlated to the ability of participants to remember television commercials days after they have seen them [16]. Electrical engineering presents a possible solution with recent advances in the use of biometric signal analysis for the detection of affective (emotional) response [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27].

Despite the wealth of literature on the use of EEG to determine attentiveness in a variety of applications, the use of EEG for the detection of attentiveness towards short training videos has not been studied, nor is there a great deal of consistency with regard to specific methods that would imply a single method for this new application. Indeed, there is great variety in EEG signal processing and machine learning methods described in the literature cited above and in other literature [28] [29] [30] [31] [32] [33] [34].

This paper presents a novel method which uses EEG as an input to an automated system that measures a participant's attentiveness while watching a short training video. This paper provides the results of a pilot study, including a structured comparison of signal processing and machine learning methods to find optimal solutions which can be extended to other applications.

**Keywords**—*Electroencephalogram; EEG; Signal Analysis; Machine Learning; Attentiveness; Training; Videos*

## I. INTRODUCTION

For the purposes of this experiment, the affective state of the participant watching the short training video is deemed to be attentive when the participant has a high positive affect, including feelings of satisfaction, engagement, interest, and involvement. This definition of short training video attentiveness is not the only possible definition, but it is justified by related research in similar applications.

In [20] the authors propose a brief and easy to administer mood scale called the Positive and Negative Affect Schedule

(PANAS), where the orthogonal axis of emotion are “negative affect” and “positive affect” and high positive affect is associated with terms such as high energy, full concentration, and pleasurable engagement.

In [21] affect is explored not from the point of view of the learner having an emotional response, but rather a computerized tutoring system having an avatar that uses facial expressions and body movements to communicate an affective state to the learner. With regard to the range of emotions expressed by the avatar, the researchers explain “...because their role is primarily to facilitate positive learning experiences, only a critical subset of the full range of emotive expression is useful for pedagogical agents.

For example, they should be able to exhibit body language that expresses joy and excitement when learners do well, inquisitiveness for uncertain situations (such as when rhetorical questions are posed), and disappointment when problem-solving progress is less than optimal.

In [22] the authors defines a model of emotions and learning that explains the emotional state of the learner as being in one of four quadrants, depending on the positive or negative value on two axes. The first axis is “learning,” and the second axis is “affect.” The positive side of the affect axis is associated with terms such as awe, satisfaction, curiosity, and hopefulness.

In [32] the researchers use machine learning to predict a participant's self-assessed emotion when presented with the results of a quiz and personality test. The participant finds out the results, and then expresses their emotions by selecting one of the following words: disappointment, distress, joy, relief, satisfaction or fear-confirmed.

In [26] researchers analyze the affective state of participants while they are performing a learning task with a computerized natural language tutor using both online and offline self-reports by participants as well as peers and trained judges observing facial features. Words that the study used in both self-reports and reports by observers are defined in TABLE I.

It is useful to note that this study found the inter-rater reliability coefficient, Kappa, of trained judges who watch the facial expressions of the participants and code facial actions to provide affective judgments was only 0.36. Although Kappa is a conservative measure of agreement, it still shows that emotions are tough to judge even by those who are trained to do so.

TABLE I. DEFINITION OF AFFECTIVE STATES

Affective state	Definition
Boredom	State of being weary and restless through lack of interest
Confusion	Failure to differentiate similar or related ideas/ noticeable lack of understanding
Flow	State of interest that results from involvement in an activity
Frustration	Making vain or ineffectual efforts however vigorous; a deep chronic sense or state of insecurity and dissatisfaction arising from unresolved problems or unfulfilled needs; dissatisfaction or annoyance
Neutral	No apparent emotion or feeling
Surprise	Wonder or amazement, especially from the unexpected

The Mathematical definition of Kappa is shown in Equation (1) where  $\text{Pr}(a)$  is the relative observed agreement among raters, and  $\text{Pr}(e)$  is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly saying each category.

$$\kappa = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)} \quad (1)$$

If the raters are in complete agreement then  $\kappa = 1$ . If there is no agreement among the raters other than what would be expected by chance, as defined by  $\text{Pr}(e)$ ,  $\kappa = 0$ .

In [23] the researchers attempt to dive deeper into the affective state of the learner by gathering, in their own words, the definition of what they mean by “engagement in courses” and “What makes a course engaging to you?” The learners in this study were male and female freshman and sophomores in engineering, and the researchers were able to group the responses into categories, as well as give specific wording terms that the students used. Some notable popular categories and terms used include: active participation, hands on, faculty enthusiasm and interest, discussion, interaction between faculty and students, as well as other similar terms, and other less popular terms.

So for the purposes of this pilot study, the definition of attentiveness is when the participant has a high positive affect towards a short training video.

## II. EXPERIMENTAL SETUP

Recent research has demonstrated the use of a single dry frontal electrode for the capture of EEG in similar applications. For this research experiment, the MindWave Mobile device from NeuroSky was used.

In [28], the same make and model of EEG data collection device, showed a positive correlation between EEG data collected and participant self-reported attention during an experiment where the participant interacts with a computer generated three-dimensional avatar.

Also using the same make and model of EEG data collection equipment is [17], where 78% of the time the brain wave data accurately identified the participant’s self-categorization of affective state during an experiment which presented increasingly stressful mental challenges to the participant. [27] also used the same device to predict with accuracy levels significantly better than chance whether or not the participant was reading an easy or a hard sentence. Finally, the presentation related to [36] indicated that the correlation analysis could just as effectively use only the frontal contact.

In [25], the participant’s affective state while using an Intelligent Tutoring System (ITS) is measured using a single frontal lobe EEG electrode with earlobe grounds showing an accurate detection (82%) of affective state from the brainwaves. [25] was based on earlier work [31] which also used the same EEG sensor setup and revealed an ability to use brainwaves to accurately predict participant self-assessed emotion according to the Self-Assessment Manikin (SAM) scale. [33] also uses only a frontal electrode (two electrodes are used, however the difference in voltage between the two is used as the only single EEG data stream over time) to achieve an 85.7% accuracy in identifying if the participant is watching an interesting video versus an uninteresting video. In [35], the single electrode is placed at F3 or F4 alternatively in an experiment that also collected facial expressions when watching emotion inducing videos, revealing that noisy correlated EEG-facial expression data could also be correctly recovered to as high in 95.2% cases.

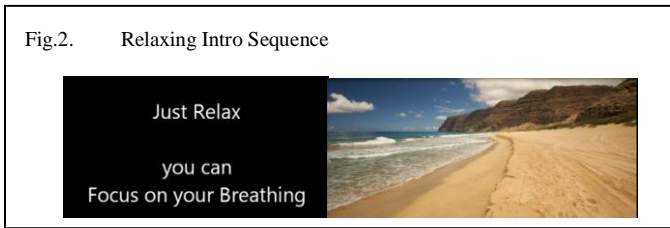
### A. Watching Short Training Videos

Participants are fitted for the dry contact EEG, sitting in front of the data collection and video playback laptop as seen in Figure 1. The headset is adjusted until a strong signal is generated without interruption, and blinks can be seen on the EEG graph displayed in real-time.

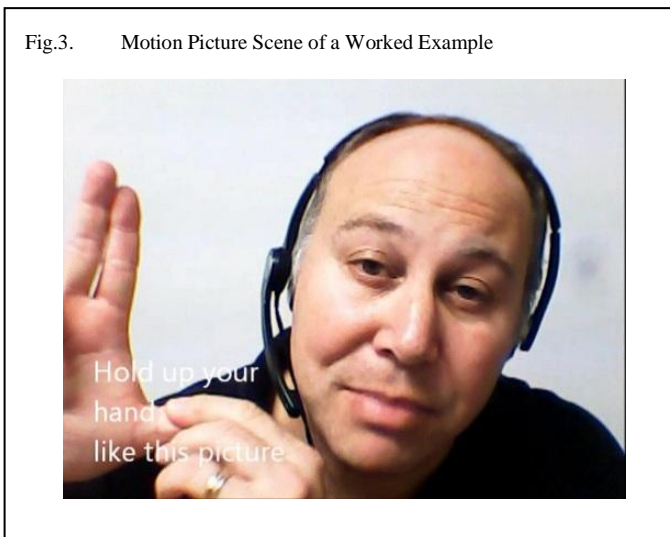


Fig.1. Participant with single dry contact EEG sensor

Once the setup is operational, the recording is started, and the video playback begins. The video starts with a relaxation segment that begins with a text message as in Figure 2.



After 2.5 minutes of relaxing empty-beach photos with the sound of waves and seagulls in the background, the short training video plays for an additional 2.5 minutes. The training videos explain a hypothetical technical lesson regarding an imaginary device with components names such as “Alpha Module” and “Beta Module”. The participant must learn how to troubleshoot the device and decide whether or not to replace a component. The videos include enhanced podcast (essentially a PowerPoint slide show that is narrated) as well as worked example style training that asks the participant to act along with the still pictures and live video of the trainer. The participant is encouraged to act along with the instructor in the video. That acting may be to hold up their hand, or to grasp their thumb or grasp their index and middle finger. This is done as a memory aid because the fictional device purposely looks like the fingers of a hand. One scene is shown in Figure 3.



### B. Data Collection

EEG signal is collected at 512 Hz in 1 second frames which can be concatenated to any length. To test some of the procedures and mathematical algorithms prior to the proposed research study, a set of seven participants were asked to participate, and detailed notes were taken during this testing by an observer who also sat in the room. After the videos were done, the participant removed the EEG headset and answers a second questionnaire, including a set of closed ended questions (yes or no) and a multiple choice quiz relating to the lesson provided.

Once the EEG data is collected, the data is segmented into areas where the participant is expected to be inattentive (during the beach pictures) and where the participant is expected to be attentive (during the short training video). The goal is to collect exemplary data of attentive and inattentive epochs gathered

from participants observing videos that are purposely designed to induce an inattentive or attentive state. Once the system is trained, it can then be used as an automated detection mechanism for attentiveness, because it has learned by example, what the EEG data contains in these situations. Even if the participants are not perfectly attentive for the entire duration of the training video, they are attentive for the majority of the time, and outlier data points will be in the minority.

For the Pilot study, each participant has 150 seconds of inattentive EEG data (watching the beach photos), and 150 seconds of attentive EEG data (watching the training video), for a total of 300 seconds (5 minutes) of EEG data. Since data collection is in 1-second epochs, a total of 2100 sample epochs are collected across 7 participants.

### III. ANALYSIS OF EEG DATA

Analysis consists of windowing, transformation, data reduction, and machine learning and pattern recognition.

#### A. Windowing

Non-overlapping 512 sample one-second windows are used. While many possible windowing functions are available this Pilot Study uses the Hanning window shown in Equation (2) which reduces ripple in the frequency power spectrum.

$$x'_n = x_n \frac{(1 - \cos \frac{2\pi n}{N})}{2} \quad n = 0, \dots, N - 1 \quad (2)$$

#### B. Transformation

After windowing is complete, the transformation of the data is required. This is done to extract important features that will be used by machine learning, and also in large part to remove instance by instance timing variations that are not important to the problem at hand.

The Pilot study compares Discrete Fourier Transform (DFT) with Discrete Wavelet Transform (DWT). Both DFT and DWT provide banded spectrum information. Equation (3) shows the transformation of the sampled data  $x_n$  into the frequency domain  $X_k$  using DFT and Equation (4) shows the magnitude of the Fourier transform  $|X_k|$  used to remove phase information leaving only the frequency information.

$$X_k = \sum_{n=0}^{N-1} x'_n e^{-i2\pi k \frac{n}{N}} \quad k = 0, \dots, N - 1 \quad (3)$$

$$|X_k| = \sqrt{X_k X_k^*} \quad (4)$$

Each element  $|X_k|$  now represents the power of that frequency within the original signal,  $f_k$  ranging from zero Hz (sometimes called DC for electrical signals) all the way up to the Nyquist frequency of the sampled data window of width T seconds, as seen in Equation (5).

$$f_k = \frac{k}{T} \quad k = 0, \dots, \frac{N}{2} - 1 \quad (5)$$

Note that the sequence of values  $|X_k|$  are symmetrical about  $N/2$  as shown in Equation (6).

$$|X_{N-k}| = |X_k| \quad k = 1, \dots, \frac{N}{2} \quad (6)$$

Also used for comparison in the Pilot study is the DFT as shown in equations (7) and (8) where the original windowed signal  $x_n$  is transformed into two signals, each of half the duration (down sampled by a factor of two) called  $X_{appn}$  and  $X_{detn}$ , the approximation coefficients and the detail coefficients, respectively. The signals  $h_n$  and  $g_n$  are a specially matched set of filters, low-pass and high-pass respectively (called Mallat wavelet decomposition), which are generated from the desired wavelet function.

$$X_{appn} = \sum_{k=0}^{N-1} x'_k h_{2n-k} \quad n = 0, \dots, N-1 \quad (7)$$

$$X_{detn} = \sum_{k=0}^{N-1} x'_k g_{2n-k} \quad n = 0, \dots, N-1 \quad (8)$$

DWT can sometimes be a preferred transformation over DFT, because it preserves the sequencing of time varying events, and has greater resolution with regard to the timing of signal events (rather than only phase information). DFT can be made to approximate DWT benefits by using smaller window size with a tradeoff in low frequency resolution, and variations on such Short Time Fourier Transform (STFT) including window overlay and window averaging appear in literature. Similarly, a limitation of DWT is that it provides scale information, but not specifically frequency information. Here too, techniques described in literature show a means to extract frequency info from DWT based on the down sampling nature of the algorithm. For example, the selection of a window size that is a power of 2 (such as 512 samples used in the Pilot study) allows repeated iterations of the wavelet transform to focus on different frequency bands.

DWT decompositions can be repeated again and again, saving the detail coefficients, and then taking the approximation coefficients and passing them through Equations (7) and (8) again. Each such repetition is called a Level of DWT, so the first level is on the original signal using a window size of  $N$ . The second iteration is Level 2, using a window size of  $N/2$ . The  $l$ th iteration is Level  $l$ , using a window size of  $N/2^l$ . Equations (9) and (10) describe how the detail coefficients  $X_{detn}$  of the Level  $l$  DWT decomposition represent a subset of the frequencies contained within the original signal.  $F_{Nyquist}$  is the Nyquist frequency of the sampled signal, and the range of frequencies in the detail coefficients range from a low of  $F_{det\_L}$  to a high of  $F_{det\_H}$ .

$$F_{det\_Hl} = \frac{F_{Nyquist}}{2^{l-1}} \quad (9)$$

$$F_{det\_Ll} = \frac{F_{Nyquist}}{2^l} \quad (10)$$

In the Pilot study, windows of 512 samples are used, representing one-second of collected data (512 Hz sampling rate). The Nyquist frequency,  $F_{Nyquist}$ , is therefore half the sampling rate (256 Hz) and represents the highest frequency represented in the sample. Any frequency higher than 256 Hz that appears in the signal prior to sampling (analog to digital conversion) will result in aliasing, and so must be removed prior to sampling (using hardware filters in the EEG data

collection device). In the Pilot study, a 6th level DWT decomposition was used employing the Daubechies db1 wavelet (equivalently the Haar wavelet, which is a step function with values 1 and -1 of equal duration). This was repeated six times, and the equivalent frequency ranges for each level are shown in **Error! Reference source not found.**

TABLE II. MAPPING OF DWT DECOMPOSITION LEVELS TO EEG FREQUENCY BANDS

Decomposition transform	Frequencies contained in the transformed data	Common EEG Band Name (literature naming and ranges vary)
Xdet1	128 – 256 Hz	Noise [12]
Xdet2	64 – 128 Hz	Gamma+ [25]
Xdet3	32 – 64 Hz	Gamma
Xdet4	16 – 32 Hz	Beta
Xdet5	8 – 16 Hz	Alpha
Xdet6	4 – 8 Hz	Theta
Xapp6	1 – 4 Hz	Delta

The Pilot study compares both DFT and DWT algorithms.

### C. Data Reduction

Data reduction is needed to speed the machine learning algorithms, and to reduce the issues of high dimension data causes machine learning algorithms to over-fit on the training data. This causes them to have reduced ability for generalization (to correctly categorize data that is not part of the training set). This is sometimes called the curse of dimensionality.

Data reduction can take place using transformation-independent algorithms such as Principal Component Analysis or Fischer Discriminant analysis. These algorithms seek to find an optimal linear transformation that reduces the number of dimensions while keeping the data as spread out as possible, thereby keeping the dimensional information most useful to distance metric machine learning algorithms. Data reduction techniques can also be specific to the transformation algorithm, using the knowledge of the algorithm to find an optimal data reduction scheme.

Data reduction of DFT is described in literature by throwing away frequency information, or by banding adjacent frequency power values by averaging them into a single band-power value. The Pilot study compares different data reduction algorithms. For the DFT, frequency data is discarded at the high end in increments of the powers of 2, to compare machine learning accuracy with reduced data sets. So the Pilot study examines presenting the machine learning algorithm with data sets having dimensionality 256, 128, 64, 32, 16, 8, 4, and 2.

Data reduction of DWT is described in literature as grouping the various detail coefficients and the approximation coefficients into one or more descriptive attributes regarding that set of data. For example, literature describes taking a set of detail coefficients and combining them into a set of parameters including Energy, Power, Median, Entropy, Mean, Min, Max,

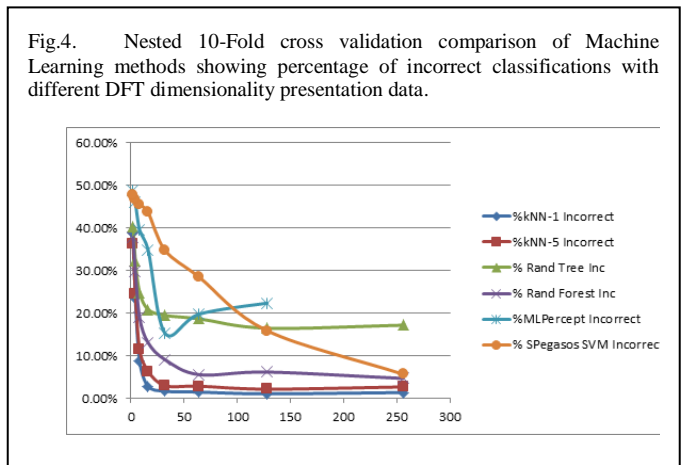


and Slope. For the Pilot study, the DWT is examined at each of 6 levels of decomposition, where at each level the detail coefficients are used to calculate the aforementioned eight parameters, and at the 6th level of decomposition, the approximation coefficients are also used to calculate these parameters. So the Pilot study uses 7 sets of coefficients each yielding 8 parameters giving the data sets the dimensionality of 8x7 or 56.

#### D. Machine Learning and Pattern Recognition

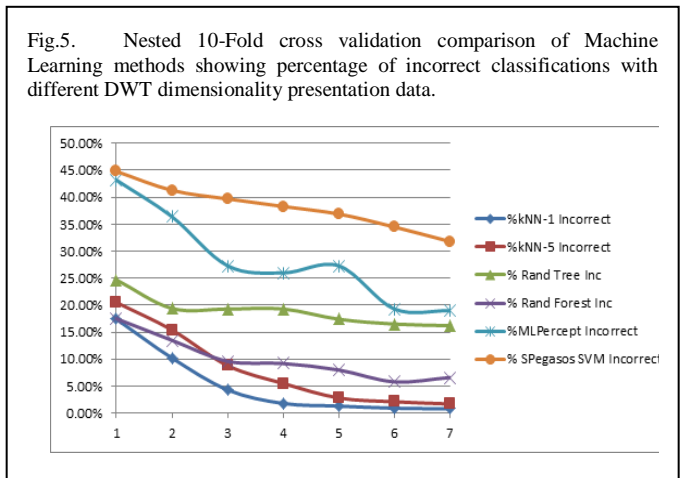
The purpose of N-Fold cross validation is to put the problem solution through rigorous quality assurance using the data that is available. Nested N-fold cross validation adds the additional step of changing the EEG signal analysis parameters. So for example, if 10 different possible ways of performing EEG signal analysis are to be compared using 10-fold cross validation, then 100 different tests will be performed.

Figure 4 shows such a nested 10-Fold cross validation comparison showing percentage of incorrect classifications for DFT across a number of dimension reduction options, and a number of machine learning algorithms. Shown in the figure is the DFT with the full 256 spectral power values, as well as a number of reduced dimensions achieved by keeping only the lowest frequency values (keeping only the lowest 2, 4, 8, 16, 32, 64, and 128 power frequency values). Furthermore, a number of machine learning algorithms, some from the literature, and others used for comparison, are compared. They are, in order of listing in the graph key, kNN algorithm with k-1 and k-5, the Random Tree algorithm, the Random Forest algorithm, a Multi-Layer Perceptron with the number of hidden layer neurons =  $(\# \text{ attributes} + 2)/2$ , and a Support Vector Machine using a Hinge loss function.



Similarly, Figure 5 - also shows a nested 10-Fold cross validation comparison showing percentage of incorrect classifications, this time for DWT, across a number of dimension reduction options, and a number of machine learning algorithms. Shown in the figure is the DWT with the full 7 Coefficient sets, as well as a number of reduced dimensions achieved by keeping only the lowest frequency coefficient sets (keeping only the lowest 1, 2, 3, 4, 5, and 6 coefficient sets). As noted above, the coefficient sets for a sixth level DWT decomposition include Detail Coefficient Sets 1 through 6, plus the Approximation Coefficient Set for level 6.

Note that in the Pilot study, for each coefficient set, we calculate 8 parameters for each coefficient set. Just as in Figure 4 - the same set of machine learning algorithms are compared.



#### IV. CONCLUSIONS AND FUTURE WORK

The results and review encourage the author to do more investigation in this area. The conclusions reached from the pilot study is that some future work on improvements is needed, which should not be very difficult to accomplish in order to achieve the goal of solving the problem of signal processing of EEG for the detection of attentiveness under the constraints of watching short training videos.

##### A. Experimental Setup

There were a number of improvements that will be needed on the experimental setup. These improvements were discovered as part of conducting the trial run of the experiments, and by having the educational podcasts (short training videos) reviewed by an expert in the field. The improvements planned include:

- Beach scene may be made to induce further inattentiveness (more boring). May use single candle burning or other simpler video without audio.
- Improve the training video based on feedback from the professional review. Have the videos instruct the participant to complete simple tasks that allow external measurement if the participant is on task or not.
- Ask participants to go to the bathroom first.
- Insist that cellphones have to be turned off.
- Find a more quiet room to do experiment - occasional voices through the walls, and slamming doors outside distracting.
- Rescale the "age" question. Negative reaction to having to check "45+"

The post-experiment questions will be improved to include more data collected. Open-ended questions will be added. This will allow the proposed research, which is basic research and intended to produce an automated system, also collect data for applied research, such as may research that employs the final method and apparatus under direction of an educational

psychologist. Open ended questions may include formations such as "tell me what you were thinking when...?" and "what were you feeling when...?"

Data collection improvements are also needed, with some ideas including:

- Use a fresh AAA battery before every experiment or two, as the wireless EEG headset loses connection after 2 or 3 uses.
- Have a more precise time stamping method for synchronization of video playback and start of EEG collection (because sometimes the PC is slow to start up the video playback). Perhaps a stopwatch with lap timer separate from the experiment laptop, so it is unaffected by CPU usage dependency that might delay the recording of when the operator clicks a button.
- Even though outside distractions will be minimized as per the above, there should still be a time stamping method of recording precisely when observed unplanned events take place (a cough, the time when a participant sighs or lifts their hand, etc.)

### B. Signal Processing

There were some improvements needed for the proposed research that were discovered during the Pilot study. These include the application of Principal Component Analysis (PCA) and other methods of data reduction to ensure the maximum amount of useful information is obtained. Also, it may be beneficial to eliminate noise such as muscle and eye movement prior to feature extraction. Based on test results, the most suitable methods will be selected in terms of both performance and speed.

Finally, the Pilot study confirms the need for a structured and orderly comparison using a nested n-fold cross validation in which all such options are tested, compared and optimized.

### C. Other Future Work

A few of the references in this paper are older. These are wonderful foundational documents; however additional reference search will look for advances which can be helpful to the research.

Also noted is the need for additional detail on the results obtained in comparison. This reaffirms the need for structured nested n-fold cross validation which not only provides an interpretation of the results, but also definitively reveals the optimal solution for the experimental problem through direct side-by-side comparison.

#### REFERENCES

- [1] R. D. Bakera, S. D'Mello, M. Rodrigo and A. Graesser, "Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments," *Int. J. Human-Computer Studies*, vol. 68, p. 223-241, 2010.
- [2] R. H. Kay, "Exploring the use of video podcasts in education: A comprehensive review of the literature," *Computers in Human Behavior*, vol. 28, p. 820-831, 2012.
- [3] D. Sandberg, T. Åkerstedt, A. Anund, G. Kecklund and M. Wahde, "Detecting Driver Sleepiness Using Optimized Nonlinear Combinations of," *IEEE trans. Intelligent transportations systems*, vol. 12, no. 1, pp. 97-108, 2010.
- [4] K. HAYASHI, K. ISHIHARA, H. HASHIMOTO and K. OGURI, "Individualized Drowsiness Detection during Driving by Pulse Wave Analysis with Neural," *Intelligent Transportation Systems, 2005. Conference Proceedings.*, pp. 901-906, 2005.
- [5] B. J. Wilson and T. D. Bracewell, "Alertness Monitoring Using Neural Networks for EEG Analysis," *Neural Networks for Signal Processing X, 2000. Proceedings of the 2000 IEEE Signal Processing Society Workshop*, vol. 2, pp. 814-820, 2002.
- [6] T. Jung, S. Makeig, M. Stensmo and T. Sejnowski, "Estimating Alertness from the EEG power Spectrum," *IEEE Transactions on bio-medical engineering*, vol. 44, no. 11, pp. 60-70, 1997.
- [7] A. Subasi, M. K. Kiymik, M. Akin and O. Erogul, "Automatic recognition of vigilance state by using a wavelet-based artificial neural networks," *Neural Computations and Applications*, vol. 14, no. 1, pp. 45-55, 2005.
- [8] V. Khare, J. Santhosh, S. Anand and M. Bhatia, "Classification of Five mental Tasks from EEG Data Using Neural Networks Based on Principal Component Analysis," *The IUP Journal of Science & Technology*, vol. 5, no. 4, pp. 31-38, 2009.
- [9] R. Palaniappan, "Utilizing Gamma Band to Improve Mental Task Based Brain- Computer Interface Design," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 3, pp. 299-303, 2006.
- [10] A. Chakraborty, P. Bhowmik, S. Das, A. Halder, A. Konar and K. Nagar, "Correlation between Stimulated Emotion Extracted from EEG and its Manifestation on Facial Expression," *Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 3132-3137, 2009.
- [11] R. Khosrowabadi, M. Heijnen, A. Wahab and H. Chai Quek, "The Dynamic Emotion Recognition System Based on Functional Connectivity of Brain Regions," *2010 IEEE Intelligent Vehicles Symposium*, pp. 377-381, 2010.
- [12] M. Murugappan, "Human Emotion Classification using Wavelet Transform and KNN," *2011 International Conference on Pattern Analysis and Intelligent Robotic*, pp. 148-153, 2011.
- [13] S. H. Lee, B. Abibullaev, W.-S. Kang, Y. Shin and J. An, "Analysis of Attention Deficit Hyperactivity Disorder in EEG Using Wavelet Transform and Self Organizing Maps," *International Conference on Control, Automation and Systems 2010 in KINTEX*, pp. 2439-2442, 2010.
- [14] F. De Vico Fallani and e. al, "Structure of the cortical networks during successful memory encoding in TV commercials," *Clinical Neurophysiology*, vol. 119, pp. 2231-2237, 2008.
- [15] K. Crowley, A. Sliney, I. Pitt and D. Murphy, "Evaluating a Brain-Computer Interface to Categorise Human Emotional Response," *IEEE International Conference on Advanced Learning Technologies*, vol. 10th, pp. 276-278, 2010.
- [16] C. Berka, D. Levendowski, M. Cvetinovic, M. Petrovic, G. David, M. Limicao, V. Zivkovic, M. Popovic and R. Olmstead, "Real-Time Analysis of EEG Indexes of Alertness, Cognition, and memory Acquired with a Wireless EEG Headset," *International Journal of Human-Computer Interaction*, vol. 17 (2), pp. 151-170, 2004.
- [17] C. Lisetti and F. Nasoz, "Using Noninvasive Wearable Computers to Recognize Human," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 11, pp. 1672-1687, 2004.
- [18] D. Watson, L. A. Clark and A. Tellegen, "Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales," *Journal of Personality and Social Psychology*, vol. Vol. 54, no. No. 6, pp. 1063-1070, 1988.
- [19] J. C. Lester, S. G. Towns and P. J. Fitzgerald, "Achieving Affective Impact: Visual Emotive Communication in Lifelike Pedagogical Agents," *International Journal of Artificial Intelligence in Education*, vol. (IJAIED) 10, pp. 278-291, 1999.
- [20] B. Kort, R. Reilly and R. Picard, "An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy—Building a Learning Companion," *Proc. Int. Conf. Advanced Learning Technologies*, vol. 2002, p. 43-48, 2002.

- [21] R. S. HELLER, C. BEIL, K. DAM and B. HAERUM, "Student and Faculty Perceptions of Engagement in Engineering," *Journal of Engineering Education*, pp. 253-261, July 2010.
- [22] N. Garnefski, V. Kraaij and P. Spinhoven, "Negative life events, cognitive emotion regulation and emotional problems," *Personality and Individual Differences*, vol. 30, no. 2001, pp. 1311-1327, 2001.
- [23] C. Frasson and P. Chalfoun, "Managing Learner's Affective States in Intelligent Tutoring Systems," *Advances in Intelligent Tutoring Systems*, vol. SCI, no. 308, p. 339-358, 2010.
- [24] S. K. D'Mello, S. D. Craig and A. C. Graesser, "Multimethod assessment of affective experience and expression during deep learning," *Int. J. Learning Technology*, vol. Vol. 4, no. Nos. 3/4, pp. 165-187, 2009.
- [25] J. Mostow, K.-m. Chang and J. Nelson, "Toward Exploiting EEG Input in a Reading Tutor," *Proceedings of the 15th International Conference on Artificial Intelligence in Education*, vol. 15th, pp. 230-237, 2011.
- [26] G. Rebolledo-Mendez, I. Dunwell, E. A. Martínez-Mirón, M. D. Vargas-Cerdán, S. de Freitas, F. Liarokapis and A. R. García-Gaona, "Assessing NeuroSky's Usability to Detect Attention Levels in an Assessment Exercise," *Human-Computer Interaction*, vol. Part I, no. HCI 2009, p. 149-158, 2009.
- [27] H. Nagashino, T. Kawano, M. Akutagawa, Q. Zhang, Y. Kinouch, F. Shichijo and S. Nagahiro, "Application of Neural Networks to Dynamics Identification by EEG," *Seventh International Conference on control, Automation, Robotics And Vision*, vol. 1, pp. 554-559, 2002.
- [28] S.-i. Ito, Y. Mitsukura, K. Sato, S. Fujisawa and M. Fukumi, "Study on Association between User's Personality and Individual Characteristic of Left Prefrontal Pole EEG Activity," *2010 Sixth International Conference on Natural Computation (ICNC 2010)*, pp. 2163-2166, 2010.
- [29] A. Heraz and C. Frasson, "Predicting the Three Major Dimensions of the Learner's Emotions from Brainwaves," *World Academy of Science, Engineering and Technology*, vol. 31, pp. 323-329, 2007.
- [30] P. Chalfoun, S. Chaffar and C. Frasson, "Predicting the Emotional Reaction of the Learner with a Machine Learning Technique," *Workshop on Motivational and Affective Issues in ITS*, vol. ITS'06, 2006.
- [31] A. Belle, R. Hobson Hargraves and K. Najarian, "An Automated Optimal Engagement and Attention Detection System Using Electrocardiogram," *Computational and Mathematical Methods in Medicine*, vol. 2012, no. Article ID 528781, pp. 1-12, 2012.
- [32] S. Selvan and R. Srinivasan, "Removal of Ocular Artifacts from EEG Using an Efficient Neural Network Based Adaptive Filtering Technique," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 330-332, 1999.
- [33] T. Green and K. Najarian, "Correlations between emotion regulation, learning performance, and cortical activity," in *Proceedings of the 29th Annual Conference of the Cognitive Science Society*, Nashville, TN, 2007.
- [34] A. Chakraborty, P. Bhowmik, S. Das, A. Halder, A. Konar and A. K. Nagar, "Correlation between Stimulated Emotion Extracted from EEG and its Manifestation on Facial Expression," *Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics*, vol. 2009, pp. 3132-3137, 2009.
- [35] P. A. Nussbaum, "LITERATURE REVIEW OF SIGNAL PROCESSING OF ELECTROENCEPHALOGRAPH FOR THE DETECTION OF ATTENTIVENESS TOWARDS SHORT TRAINING VIDEOS," *Advances in Artificial Intelligence*, 2013 (Under Review, not yet published).