

Evaluating an Elevated Signal-to-Noise Ratio in EEG Emotion Recognition

Zachary Estreito, University of Nevada, Reno, USA*

Vinh Le, University of Nevada, Reno, USA

Frederick C. Harris Jr., University of Nevada, Reno, USA

 <https://orcid.org/0000-0002-0857-6931>

Sergiu M. Dascalu, University of Nevada, Reno, USA

ABSTRACT

Predicting valence and arousal values from EEG signals has been a steadfast research topic within the field of affective computing or emotional AI. Although numerous valid techniques to predict valence and arousal values from EEG signals have been established and verified, the EEG data collection process itself is relatively undocumented. This creates an artificial learning curve for new researchers seeking to incorporate EEGs within their research workflow. In this article, a study is presented that illustrates the importance of a strict EEG data collection process for EEG affective computing studies. The work was evaluated by first validating the effectiveness of a machine learning prediction model on the DREAMER dataset, then showcasing the lack of effectiveness of the same machine learning prediction model on cursorily obtained EEG data.

KEYWORDS

Affective Computing, Arousal, BCI, Data Analysis, Data Integrity, Emotion Classification, Machine Learning, User Study, Valence

INTRODUCTION

The recognition of human emotions from electroencephalography (EEG) signals has been a steadfast research topic within the field of affective computing. In the past two decades, many valid techniques to interpret human emotional states from EEG signals have been documented, and a standardized workflow has been established. This standardized process involves EEG signal collection, band separation, feature extraction, and emotion classification. Information about proper EEG data collection processes, however, is scarce. There are no commonly accepted guidelines for proper EEG data collection techniques. As such, this scarcity of information leads to an artificial bottleneck for new researchers seeking to incorporate EEGs within their research workflow.

EEG signals are extremely vulnerable to external artifacts, and a controlled environment is essential for obtaining usable EEG signal data. While there are techniques that can be used to remove artifacts and noise from EEG data, these techniques are limited in their efficacy. A participant's

DOI: 10.4018/IJSI.333161

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

involuntary movement, a random sound emitted near the participant, or a distracting conversation can easily be enough to compromise the data being collected. Furthermore, the collection of data could also be compromised by factors not easily controlled by an experiment's facilitator, such as the participant's hair type, length, or density or a psychoactive medication being taken by the participant.

In this paper, an EEG user study is presented in which external artifacts, such as noise or involuntary movement, accumulated within the EEG data collected due to an uncontrolled testing environment. From there, the results of multiple different machine learning prediction models on the established DREAMER dataset are compared to the results of the same prediction models on the EEG data obtained in the aforementioned poorly controlled environment. This study demonstrates the importance of a meticulous and methodical approach to EEG data collection by showing the ineffectuality of validated data analysis techniques when applied to EEG data with a high signal-to-noise ratio.

The remainder of this paper is structured as follows: the second section presents a brief background of the concepts of valence and arousal within affective EEG studies and the related techniques utilized during the study, the third section provides details of the experiment methodology, the fourth section showcases a comparison between the analysis of the EEG data collected during the experiment and the analysis of the EEG data from the DREAMER dataset, the fifth section provides a brief discussion of the experiment and EEG data collection, and sixth section wraps up the paper with the conclusion and planned future work.

BACKGROUND

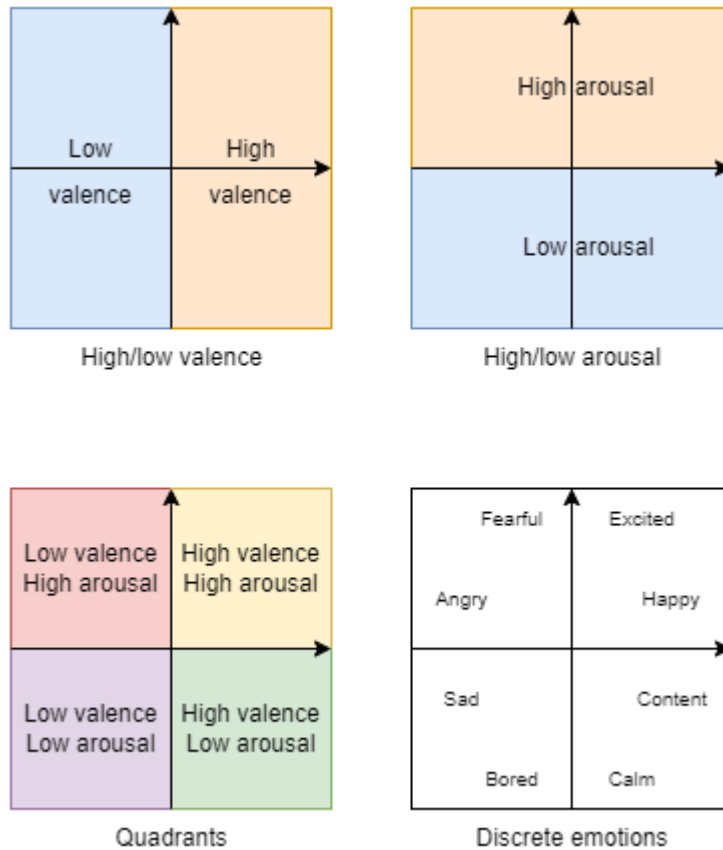
Valence–Arousal Model

Russell's valence–arousal model (Russell et al., 1979) is a human emotion classification model consisting of two dimensions: valence and arousal. *Valence* represents the positivity of the emotion being felt, with positive emotions existing on one side of the axis and negative emotions on the other side. *Arousal* represents the degree of stimulation of the emotion being felt, with high-stimulation emotions being placed on one side of the axis perpendicular to the valence axis and low-stimulation emotions on the other side. The combination of these two axes allows emotions to be categorized into four unique quadrants: high valence with high arousal, high valence with low arousal, low valence with high arousal, and low valence with low arousal. These quadrants provide a convenient way to group similar emotions. With the quadrants drawn out, discrete emotions can then be placed in their associated quadrants, as seen in Figure 1.

Pre-Processing

The goal of pre-processing EEG data is to remove noise from the signals, thereby yielding a higher signal-to-noise ratio. EEG electrodes are highly prone to artifacts and interferences. Notch filtering is often used to remove alternating current power line interference from EEG signals, most commonly at 50 Hz or 60 Hz. Bandpass filtering is used to remove frequencies outside of the useful range. Signals can also be detrended to compensate for the dehydration of wet electrodes over the course of a recording. This dehydration would lead to signals becoming weaker over time, so detrending seeks to compensate for that. Artifacts originating from the wearer's movement, including facial muscle movement and speech, can only be removed with limited success depending on the severity of the movement. Regression analysis is an effective approach to remove artifacts, but this technique requires a reference channel. Both EEG devices used in this study provide a reference channel, so this approach was utilized. Artifactual segment rejection involves removing artifacted sections of EEG data (Islam et al., 2016). This segment rejection can be performed manually by searching for large spikes in EEG signal activity or automatically by removing sections that contain signal outliers that are many standard deviations from the mean. A visualization of EEG signal pre-processing can

Figure 1. The representation of emotions in the valence–arousal model



be seen in Figure 2. Multiple approaches to pre-processing were taken, including the use of notch filtering, bandpass filtering, signal detrending, and regression analysis. Unfortunately, despite every possible combination of pre-processing techniques being utilized, no combination was found to have any significant effects on the usability of the data obtained in this study for emotion classification.

EEG Frequency Bands

EEG signals represent neural oscillations, which are grouped into five frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30+ Hz). These different frequency bands can be seen in Figure 3. Delta waves are associated with unconscious mental activities, and theta waves are associated with subconscious mental activities, so the delta and theta band brain waves were filtered out of the data used in this study. Alpha waves are associated with relaxed conscious thoughts, beta waves are associated with active conscious perception, and gamma waves are associated more with hyperactive perception and, thus, are all useful for the classification of emotions. For this reason, EEG data used in this study was run through a bandpass filter to separate it into the alpha, beta, and gamma components. Additionally, asymmetry in brain wave activity is useful for the classification of valence, especially on the alpha band (Alarcao et al., 2017), so alpha band features were converted to alpha asymmetry differential features by subtracting the right-side alpha features from the left side alpha features (Thammasan et al., 2017).

Figure 2. EEG signal pre-processing: (a) Raw EEG signal, (b) EEG signal with power line notch filter applied, (c) EEG signal with power line notch filter and regression analysis artifact removal applied

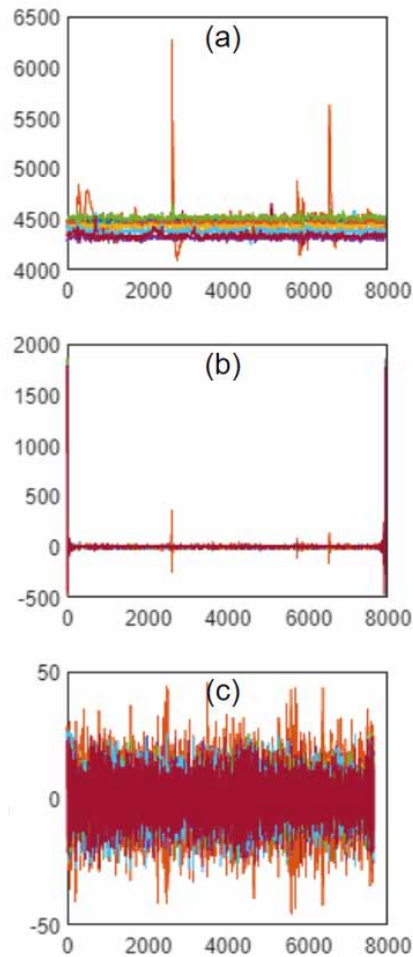
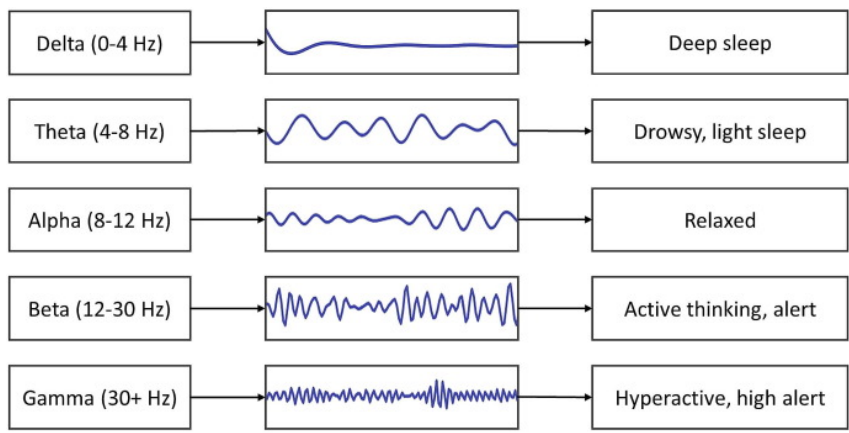


Figure 3. EEG frequency bands



EEG Feature Extraction

EEG feature extraction can be divided into four domains: frequency domain methods, time domain methods, time–frequency domain (wavelet transform) methods, and nonlinear methods (Acharya et al., 2013). For the purposes of emotion classification, all of these have been found to be useful. Galvão et al. (2021) found that the best features to use with the random forest classifier are the time domain Hjorth activity parameter, the time-frequency domain wavelet energy, and wavelet entropy parameters of the alpha differential asymmetry, beta, and gamma bands. This study utilizes the Hjorth activity parameter on the alpha differential asymmetry, beta, and gamma bands, which was found to be nearly as successful as the combination of the three aforementioned features.

Emotion Classification

Classification of emotions from EEG features with machine learning techniques has been proven to be a valid approach to EEG signal analysis. The random forest (RF) and k -nearest neighbor (k -NN) classifiers are two of the most successful machine learning methods for emotion classification in recent literature, with RF showing slightly better performance when provided with fewer features (Galvão et al., 2021; Xu et al., 2012). Because only the Hjorth activity parameter is being used in this study, the random forest classifier was chosen. A tree depth of 100 was used due to the conclusions from Giannakaki et al. (2017), and further testing revealed that increasing the tree depth past 100 did not yield better results.

METHODOLOGY

Participants

Eight participants between the ages of 22 and 35 were recruited for this study. Participants were recruited by word of mouth and an email message sent through university channels. There were no specific characteristics being sought in participants. The only known characteristics of participants that could be considered relevant to this study would be the size of the participant's head, as well as the type, thickness, and length of the participant's hair. The size of participants' heads was relevant because the EEG headsets being used in this study have limitations on the head size that they can accommodate. The different hair characteristics listed can also affect EEG signal quality. Participants were not selected based on their hair, and the participants had a wide variety of hair types.

Hardware

The Emotiv EPOC X (Emotiv, n.d.-a) and OpenBCI EEG Electrode Cap (OpenBCI Shop, n.d.-a) were selected for this study. Both devices are among the most commonly used consumer-grade EEG devices in affective computing research, and they have both been validated in multiple studies. Dadebayev et al. (2022) reviewed eight different studies involving the use of commercial EEG devices for emotion recognition. They found that the OpenBCI EEG Electrode Cap and Emotiv EPOC were both effective when used to predict valence and arousal in participants exposed to visual stimuli. The Emotiv EPOC is a lower model of the Emotiv EPOC X with a smaller electrode count. Because the Emotiv EPOC X contains all of the electrodes of the Emotiv EPOC and more, it can be extrapolated that the Emotiv EPOC X would perform equally or better than the Emotiv EPOC.

The Emotiv EPOC X, shown in Figure 4, reads EEG signals through the conductive felt pads, which are hydrated with a saline solution before use. These felt pads must contact the scalp of the wearer. Adequate contact quality can be difficult to achieve in wearers with curly, thick, and/or long hair, and the wearer's hair may need to be parted where the hydrated felt pads make contact. The Emotiv EPOC X requires the use of Emotiv's proprietary software, and there is innate pre-processing done by the device. Both of these factors make the Emotiv EPOC X less useful as a research device due to their enigmatic nature.

Figure 4. The experimental setup with the Emotiv EPOC X



The OpenBCI EEG Electrode Cap, shown in Figure 5, uses a saline gel to create a contact surface that can vary in size as desired, as well as passing through hair more easily. The cap is put on the wearer before the saline gel is applied. While the cap is being worn, the saline gel is drawn into a blunt-tip syringe to be injected into the gaps in the contact points. The amount of gel used is variable and depends on the hair type, density, and length of the wearer. Hair that is curly, dense, or long requires a greater volume of saline gel.

Software

Throughout this user study, a couple of software packages were used in order to collect the EEG data. Because Emotiv and OpenBCI have their own proprietary designs, the decision was made to use the built-in interfaces to affect each headset. To consolidate the experience and make it a bit more autonomous, a software suite designed to automatically label events within EEG streams was utilized.

EmotivPRO (Emotiv, n.d.-b) is the software package required to stream and record EEG data from the Emotiv EPOC X. EmotivPRO is proprietary and requires a subscription-based license to access the most important features, such as lab streaming layer (LSL) streaming. EmotivPRO also includes features to assist with EEG device setup by showing a live display of the contact quality of each electrode. EmotivPRO was used in this study to verify contact quality and generate LSL streams to work in conjunction with the Generalized EEG Data Acquisition and Processing System (GEDAPS) software suite.

The OpenBCI GUI (OpenBCI, n.d.-b) is an open-source GUI tool that allows streaming and recording of EEG data from OpenBCI devices. There are also visualization tools, including a contact quality visualizer to assist with cap setup. It offers the ability to stream EEG data with the BrainFlow library but not LSL. This was used in conjunction with the OpenBCI LSL (OpenBCI, n.d.-a) command-line plugin to utilize the LSL protocol.

GEDAPS (Le et al., 2023) is the software suite used to autonomously display the emotional stimuli images presented to participants in this study. GEDAPS removes the need for a facilitator to manually record timings and metadata within an EEG user study. It works together with a backend

Figure 5. The experimental setup with the OpenBCI all-in-one EEG electrode cap



that utilizes PyLSL to generate event marker timestamps and inserts it directly into an EEG headset's data stream. This is so that EEG data can later be divided up into time-based sections specific to the content being perceived by the EEG headset wearer. This suite was designed to work with the Emotiv EPOC X but was modified as part of this research to accommodate the OpenBCI EEG Electrode Cap for the purpose of this study.

Datasets

The International Affective Picture System (Lang et al., 2005) is a picture dataset with standardized emotion labels. Each picture is accompanied by self-reported ratings of valence, arousal, and dominance. This dataset contains many pictures with highly graphic content. In lieu of this, the media used for this study were specifically curated to invoke positive and negative reactions in both valence and arousal. To this end, there were four combinations used for the experiment in this paper. These combinations were: high valence with low arousal, low valence with low arousal, low valence with high arousal, and high valence with high arousal.

DREAMER (Katsigiannis & Ramzan, 2017) is a dataset of EEG signals and ECG signals accompanied by self-reported valence and arousal labels, conducted over 18 video samples. The EEG signal analysis methods used in this study were first verified on the DREAMER data alone, by creating a machine learning model out of the EEG data for 22 of the 23 entries within the DREAMER database, then using that model to predict the valence and arousal labels of the EEG data of the 23rd entry and comparing the predicted values to the self-reported values. The feature extraction and classification model used in this study was approximately 80% accurate at predicting both valence and arousal values when tested in this manner.

Procedure

All participants were tested with both the Emotiv EPOC X and the OpenBCI EEG Electrode Cap. Four of the eight participants started with the Emotiv EPOC X, and four of the participants started with the OpenBCI EEG Electrode Cap. Participants were asked not to speak, move their body, or move any facial muscles during the task but were also informed that they would be free to leave at any point should they wish to not continue. After being fitted with the first EEG device and confirming that the participant was in a comfortable position and was ready to proceed, the EEG task was initiated.

After the task was completed, the EEG device was removed from the participant's head, and they were given towels to dry their scalps on the contact points. The fitting process was then repeated with the other EEG device, and the task was performed again once the participant was ready. Upon completion of the task with the second device, the participant was once again given towels to dry their scalps, and then their involvement in the user study ended.

Tasks

As part of the user study, the tasks that were performed involved the participant watching a series of images from the IAPS dataset chosen specifically to invoke emotions that fit firmly in one of the four valence-arousal quadrants indicated above in the dataset section. Before viewing the images, participants were prompted to hold their eyes open for 60 seconds to the best of their ability and then close their eyes for an additional 60 seconds to establish a baseline. The image series sequence began being displayed after establishing this baseline. Each series of images contained 12 images displayed for 5 seconds each, for a combined total of 60 seconds. A gray screen was displayed for 5 seconds between each series to allow the participant's affect to return to baseline. While this was happening, the GEDAPS software was autonomously logging the EEG data and labeling these events for future processing. After the last series, participants were once again prompted to hold their eyes open for 60 seconds, then closed for an additional 60 seconds. This was done to further confirm the participant's baseline and to compensate for any EEG data trending that could have occurred due to electrodes becoming dehydrated over the course of the task.

RESULTS AND COMPARATIVE ANALYSIS

The first part of the comparative analysis for this study was to confirm that the feature extraction and classification methods chosen were valid. This was accomplished using the EEG data from the DREAMER dataset. Two machine learning models were created. The first model used the random forest regressor with a tree depth of 100 trained on the Hjorth activity of the alpha differential asymmetry, beta, and gamma bands. This model was created 23 times (once for each entry in DREAMER) to generate a continuous number prediction. The second model used the random forest classifier with a tree depth of 100 trained on the Hjorth activity of the alpha differential asymmetry, beta, and gamma bands. This model was created 23 times (once for each DREAMER entry) to generate a discrete number classification. In each case, the model was trained on the EEG data and self-reported labels from the other 22 entries to predict the self-reported valence and arousal values of the selected entry. The self-reported valence and arousal labels in the DREAMER dataset are integers from 1 to 5. An example of the continuous prediction results generated for all of DREAMER's 18 video samples for subjects 1–3 using Machine Learning Model 1 can be seen in Figure 6.

To determine significance, the discrete accuracy of the predicted values generated by Machine Learning Model 2 was compared to the expected 20% accuracy of a random number generator constrained to the integer values between 1 and 5, as shown in Figure 7. The combined results of all 23 subjects in the DREAMER dataset easily reached statistical significance ($p < 0.001$) when compared to the expected results of a random number generator.

With the validity of the model verified, the next step was to analyze the data collected from the participants in this study. Two approaches were taken for this analysis. Approach 1 involved using Machine Learning Model 1 to predict the values of the participants in this study. The expected results would be as follows: High valence ≥ 3.0 , low valence < 3.0 , high arousal ≥ 3.0 , and low arousal < 3.0 . Unfortunately, the data obtained from the participants in this study was not able to reach statistical significance for either of the two EEG devices using this approach,

Figure 6. Valence and arousal prediction results for DREAMER subjects 1–3 using the 22-subject continuous DREAMER model

Video Sample	Participant 1				Participant 2				Participant 3			
	Valence		Arousal		Valence		Arousal		Valence		Arousal	
	Self	Calc	Self	Calc	Self	Calc	Self	Calc	Self	Calc	Self	Calc
1	4.00	3.14	3.00	3.91	3.00	3.00	2.00	2.91	3.00	2.42	4.00	3.06
2	3.00	2.02	3.00	2.64	2.00	3.68	4.00	2.98	4.00	2.84	5.00	2.06
3	5.00	4.30	4.00	3.99	5.00	4.00	5.00	4.00	5.00	4.01	5.00	3.74
4	4.00	1.85	3.00	4.02	1.00	1.10	5.00	4.00	2.00	2.00	5.00	4.24
5	4.00	3.29	4.00	3.98	2.00	3.60	2.00	3.90	3.00	4.00	4.00	4.00
6	1.00	1.00	2.00	2.06	2.00	2.56	3.00	3.73	3.00	1.01	3.00	3.42
7	5.00	4.95	4.00	2.09	4.00	5.00	2.00	2.94	5.00	5.00	3.00	3.00
8	1.00	1.01	2.00	4.00	2.00	1.00	5.00	4.00	1.00	1.00	4.00	4.00
9	1.00	1.00	2.00	2.00	2.00	1.00	3.00	2.30	1.00	1.00	2.00	3.16
10	5.00	2.04	3.00	2.98	1.00	3.37	4.00	2.24	3.00	1.86	2.00	3.92
11	4.00	4.00	2.00	2.00	4.00	3.79	2.00	1.00	4.00	4.00	1.00	2.00
12	4.00	4.00	3.00	2.95	4.00	4.00	3.00	2.87	5.00	4.00	3.00	2.95
13	4.00	4.90	3.00	4.99	5.00	4.02	4.00	3.25	4.00	4.85	4.00	4.18
14	3.00	2.28	1.00	3.00	2.00	2.91	3.00	2.76	3.00	2.00	2.00	2.99
15	2.00	2.99	5.00	3.05	1.00	2.82	5.00	3.00	4.00	2.02	2.00	3.00
16	3.00	4.00	2.00	3.00	2.00	3.97	5.00	2.32	4.00	3.99	2.00	3.00
17	1.00	1.39	2.00	3.00	1.00	1.02	3.00	4.00	2.00	1.05	2.00	3.89
18	3.00	3.00	5.00	4.14	3.00	2.96	5.00	3.00	3.00	2.90	3.00	3.67
Mean error	0.6900		0.8422		0.9033		1.1533		0.7083		0.8989	
All participant mean error	0.7977		0.8156									
Std dev error	0.8034		0.8184		0.7539		0.6579		0.6618		0.7819	
All participant std dev error	0.7208		0.6903									
Relative accuracy	82.75%		78.94%		77.42%		71.17%		82.29%		77.53%	
All participant relative accuracy	80.06%		79.61%									

and the data did not seem to follow even a non-significant trend toward expected results, as presented in Figure 8.

The second comparative analysis approach involved training a model using data from 7 of the 8 participants to predict the valence–arousal quadrant of the 8th participant. The classifier labels used for this were binary, with 1s representing high values for valence and arousal and 0s representing low values for valence and arousal. Due to the smaller training sample size, Machine Learning Model 1 was re-validated on the DREAMER dataset for a smaller sample size by training it on every possible combination involving 8 subjects (7 to train and 1 to predict). Instead of using the provided discrete ratings from 1 to 5, DREAMER data was re-labeled to match the binary classifier labels of high/low valence and high/low arousal to check quadrant accuracy. DREAMER data was re-labeled to match the participant data, with high valence/arousal being represented by values ≥ 3.0 and low valence being represented by values < 3.0 . Mean quadrant accuracy for all participants was approximately 52%, which is much higher than the 25% that could be expected from a random number generator. This approach also reached statistical significance with a z -score of 3.3 and $p < 0.001$. A table of these predictions is presented in Figure 9.

Figure 7. Valence and arousal prediction results for DREAMER subjects 1–3 using the 22-subject discrete DREAMER model

Video Sample	Participant 1				Participant 2				Participant 3			
	Valence		Arousal		Valence		Arousal		Valence		Arousal	
	Self	Calc	Self	Calc	Self	Calc	Self	Calc	Self	Calc	Self	Calc
Discrete 1	4	3	3	4	3	3	2	3	3	2	4	3
Discrete 2	3	2	3	3	2	4	4	3	4	3	5	2
Discrete 3	5	4	4	4	5	4	5	4	5	4	5	4
Discrete 4	4	2	3	4	1	1	5	4	2	2	5	4
Discrete 5	4	3	4	4	2	4	2	4	3	4	4	4
Discrete 6	1	1	2	2	2	3	3	4	3	1	3	3
Discrete 7	5	5	4	2	4	5	2	3	5	5	3	3
Discrete 8	1	1	2	4	2	1	5	4	1	1	4	4
Discrete 9	1	1	2	2	2	1	3	2	1	1	2	3
Discrete 10	5	2	3	3	1	3	4	2	3	2	2	4
Discrete 11	4	4	2	2	4	4	2	1	4	4	1	2
Discrete 12	4	4	3	3	4	4	3	3	5	4	3	3
Discrete 13	4	5	3	5	5	4	4	3	4	5	4	4
Discrete 14	3	2	1	3	2	3	3	3	3	2	2	3
Discrete 15	2	3	5	3	1	3	5	3	4	2	2	3
Discrete 16	3	4	2	3	2	4	5	2	4	4	2	3
Discrete 17	1	1	2	3	1	1	3	4	2	1	2	4
Discrete 18	3	3	5	4	3	3	5	3	3	3	3	4
Discrete accuracy	55.56%		55.56%		66.67%		88.89%		61.11%		66.67%	
All participant discrete accuracy	62.80%		62.80%									
Discrete accuracy z-score	5.55		5.55									
Discrete accuracy p-value	0.0000000141		0.0000000141									

Unfortunately, this approach was also unsuccessful in generating statistically significant results for both the Emotiv EPOC X and the OpenBCI EEG Electrode Cap. As was the case with Approach 1, there was also no visible nonsignificant trend towards expected results. The mean valence and arousal values for all four quadrants were all within a single standard deviation of each other, indicating that the model could not reliably identify valence or arousal in the participant EEG data. These results are shown in Figure 10.

DISCUSSION

Despite all of the techniques used being validated, the analytical processes and statistical operations could not, in the end, overcome the sheer amount of artifact pollution within the collected data. Although statistical operations can indeed suppress or clean up some of the noise within an EEG dataset, there is an upper limit on what is achievable with EEG data collected from a poorly controlled environment. Techniques such as notch filtering, bandpass filtering, signal detrending, and regression analysis (Islam et al., 2016) exist to remove artifacts from EEG data. Unfortunately, despite every possible combination of these techniques being used, none of the combinations of techniques were

Figure 8. Valence and arousal prediction results for study participants using the 22-subject DREAMER model

User	Axis	Emotiv				OpenBCI			
		HVHA	HVLA	LVHA	LVLA	HVHA	HVLA	LVHA	LVLA
1	Valence	2.182	2.000	2.182	2.000	2.182	2.000	2.182	2.000
	Arousal	1.909	1.727	2.364	2.818	4.000	4.000	4.000	4.000
2	Valence	3.545	3.818	3.818	3.818	2.727	2.182	2.182	2.182
	Arousal	1.455	2.091	1.818	1.636	2.364	2.182	2.455	2.455
3	Valence	3.818	3.818	3.273	3.455	5.000	5.000	5.000	5.000
	Arousal	1.727	1.273	1.636	1.727	3.000	3.000	3.000	3.000
4	Valence	4.182	4.818	4.727	4.727	4.091	4.091	4.091	4.000
	Arousal	2.545	2.818	3.000	3.000	2.455	2.455	2.091	2.455
5	Valence	3.545	3.364	3.727	3.818	2.182	2.364	2.000	2.364
	Arousal	2.545	3.091	2.182	2.182	2.000	2.000	2.000	2.182
6	Valence	3.818	2.909	3.273	3.636	4.000	4.091	4.000	4.000
	Arousal	1.909	2.000	2.091	2.000	2.727	3.000	2.909	2.818
7	Valence	2.545	3.636	3.455	3.455	2.182	2.182	2.182	2.182
	Arousal	2.091	1.182	1.636	1.818	2.091	2.000	2.000	2.000
8	Valence	3.636	4.182	3.091	3.455	4.273	4.182	4.091	4.182
	Arousal	3.000	2.818	2.273	2.636	3.000	3.000	3.000	3.000
Mean	Valence	3.409	3.568	3.443	3.545	3.330	3.261	3.216	3.239
	Arousal	2.148	2.125	2.125	2.227	2.705	2.705	2.682	2.739
StDev	Valence	0.475	0.565	0.515	0.428	1.038	1.081	1.122	1.066
	Arousal	0.500	0.708	0.440	0.469	0.381	0.444	0.442	0.371

found to be sufficient to make the EEG data collected from participants to be usable. In order to avoid or mitigate these issues in future studies, it is important to make a conscious effort to create a controlled and secluded environment that is free of distractions. Such an environment would be conducive to EEG data collection.

In order to establish a better solution of a more controlled environment, it is important to revisit some of the issues that occurred as part of the data collection portion in this study. Despite being asked to remain still and silent, several participants involuntarily spoke or fidgeted during the task. The combination of these factors added a lot of noise to the EEG signals, resulting in a poor signal-to-noise ratio. It is also likely that some less commonly discussed external factors played a role in making the data unusable, namely participant distraction due to nearby audible conversations. If participants were distracted by nearby conversations, they may not have been processing the visual stimuli of the images being shown to them. Participant distraction does not appear as noise or artifacts on EEG data, but rather, it produces EEG data that simply does not match the expected results. This is because the participants are still feeling emotions normally, but the source of their emotions is an external factor, rather than the intended visual stimuli. For this reason, participant distraction cannot be mitigated with any data analysis approach. Lastly, the participants also had varying hair types, and it was observed during the data collection process that thicker and curlier hair significantly worsened EEG contact quality.

Figure 9. Valence and arousal prediction results for DREAMER subjects 1–3 using 7-subject quadrant DREAMER model

Video Sample	Participant 1				Participant 2				Participant 3			
	Valence		Arousal		Valence		Arousal		Valence		Arousal	
	Self	Calc	Self	Calc	Self	Calc	Self	Calc	Self	Calc	Self	Calc
1	High	High	High	High	High	High	Low	Low	High	Low	High	High
2	High	Low	High	Low	Low	High	High	Low	High	Low	High	Low
3	High	High	High	High	High	High	High	High	High	High	High	High
4	High	Low	High	High	Low	Low	High	High	Low	Low	High	High
5	High	High	High	High	Low	High	Low	High	High	High	High	High
6	Low	Low	Low	Low	Low	Low	High	High	High	Low	High	High
7	High	High	High	Low	High	High	Low	Low	High	High	High	High
8	Low	Low	Low	High	Low	Low	High	High	Low	Low	High	High
9	Low	Low	Low	Low	Low	Low	High	Low	Low	Low	Low	High
10	High	Low	High	Low	Low	High	High	Low	High	Low	Low	High
11	High	High	Low	Low	High	High	Low	Low	High	High	Low	Low
12	High	High	High	Low	High	High	High	Low	High	High	High	Low
13	High	High	High	High	High	High	High	High	High	High	High	High
14	High	Low	Low	High	Low	Low	High	Low	High	Low	Low	Low
15	Low	Low	High	High	Low	Low	High	High	High	Low	Low	High
16	High	High	Low	High	Low	High	High	Low	High	High	Low	High
17	Low	Low	Low	High	Low	Low	High	High	Low	Low	Low	High
18	High	High	High	High	High	Low	High	High	High	Low	High	High
Quadrant accuracy	50.00%				55.56%				38.89%			
Mean quadrant accuracy	51.93%											
Quadrant accuracy z-score	5.4											
Quadrant accuracy p-value	0.0000000334											

CONCLUSION

The primary goal of this study was to compare the efficacy of a machine learning model for predicting valence and arousal values from both clean and noisy EEG signal data. Data collection was performed in an uncontrolled environment, with numerous sources of external influence. The results of the captured data and the established dataset, DREAMER, were placed in the same analytical model to perform a comparative analysis. It was discovered that the combination of involuntary artifacts and external factors caused a too-severe reaction regarding the noise within the stream. Greater care and effort must be made in order to record EEG data, as technology has not reached the point in which there is a reliable way to filter out excessive noise within the data stream.

In the future, it would be worthwhile to pursue further research to determine the effects of different hair types on EEG signal quality. The goal would be to find an applicable solution that could be included in the analysis to compensate for hair type. Additionally, this research can be expanded by testing the effects of uncontrolled environments on events that would maybe benefit

Figure 10. Valence and arousal prediction results for participant data of each quadrant, separated by device

User	Axis	Emotiv				OpenBCI			
		HVHA	HVLA	LVHA	LVLA	HVHA	HVLA	LVHA	LVLA
1	Valence	0.042	0.099	0.097	0.126	1.000	1.000	1.000	1.000
	Arousal	1.000	1.000	1.000	1.000	0.808	0.786	0.755	0.762
2	Valence	0.686	0.337	0.553	0.224	0.999	0.991	0.995	0.995
	Arousal	0.657	0.119	0.510	0.327	0.962	0.989	0.998	1.000
3	Valence	0.511	0.747	0.856	0.707	1.000	1.000	1.000	1.000
	Arousal	0.752	0.646	0.730	0.860	0.146	0.166	0.194	0.205
4	Valence	0.188	0.187	0.436	0.423	1.000	0.999	1.000	1.000
	Arousal	0.856	0.952	0.999	0.944	0.655	0.832	0.672	0.465
5	Valence	0.754	0.995	0.838	0.831	1.000	1.000	1.000	1.000
	Arousal	0.108	0.055	0.260	0.045	0.933	0.967	0.971	0.971
6	Valence	0.103	0.064	0.047	0.308	0.000	0.090	0.000	0.000
	Arousal	0.673	0.562	0.685	0.665	0.051	0.159	0.023	0.022
7	Valence	0.004	0.002	0.001	0.010	0.166	0.085	0.083	0.104
	Arousal	0.272	0.004	0.047	0.545	0.019	0.021	0.040	0.055
8	Valence	0.000	0.000	0.427	0.086	1.000	1.000	1.000	1.000
	Arousal	1.000	0.979	0.758	0.889	0.029	0.023	0.031	0.015
Mean	Valence	0.286	0.304	0.407	0.339	0.771	0.771	0.760	0.762
	Arousal	0.665	0.540	0.624	0.659	0.450	0.493	0.460	0.437
StDev	Valence	0.299	0.363	0.315	0.284	0.417	0.412	0.433	0.429
	Arousal	0.294	0.387	0.301	0.307	0.403	0.420	0.415	0.403

or remain indifferent regarding artifacts, such as confusion detection. Lastly, with the recent release of OpenBCI's Gelfree EEG Cap (OpenBCI Shop, n.d.-b), a direct comparison between the efficacy and usability of consumer-grade saline hydrated felt electrode EEG devices and consumer-grade gel electrode EEG devices can be studied without the confounding variables of different hardware manufacturers and proprietary software.

ACKNOWLEDGMENT

This research was supported by the National Science Foundation (OAC-2209806), the National Science Foundation (OIA-2019609), and the National Science Foundation (OIA-2148788).

REFERENCES

- Acharya, U. R., Sree, S. V., Swapna, G., Martis, R. J., & Suri, J. S. (2013). Automated EEG analysis of epilepsy: A review. *Knowledge-Based Systems*, 45, 147–165. doi:10.1016/j.knosys.2013.02.014
- Alarcao, S. M., & Fonseca, M. J. (2017). Emotions recognition using EEG signals: A survey. *IEEE Transactions on Affective Computing*, 10(3), 374–393. doi:10.1109/TAFFC.2017.2714671
- Dadebayev, D., Goh, W. W., & Tan, E. X. (2022). EEG-based emotion recognition: Review of commercial EEG devices and machine learning techniques. *Journal of King Saud University: Computer and Information Sciences*, 34(7), 4385–4401. doi:10.1016/j.jksuci.2021.03.009
- Emotiv. (n.d.-a). *Emotiv EPOC X: 14 channel mobile EEG headset*. Emotiv. <https://www.emotiv.com/epoc-x/>
- Emotiv. (n.d.-b). *EmotivPRO*. Emotiv. <https://www.emotiv.com/emotivpro/>
- Galvão, F., Alarcão, S. M., & Fonseca, M. J. (2021). Predicting exact valence and arousal values from EEG. *Sensors (Basel)*, 21(10), 3414. doi:10.3390/s21103414 PMID:34068895
- Giannakaki, K., Giannakakis, G., Farmaki, C., & Sakkalis, V. (2017). Emotional state recognition using advanced machine learning techniques on EEG data. *Proceedings of the 2017 IEEE 30th International Symposium on Computer-Based Medical Systems*, (pp. 337–342). IEEE. doi:10.1109/CBMS.2017.156
- Islam, M. K., Rastegarnia, A., & Yang, Z. (2016). Methods for artifact detection and removal from scalp EEG: A review. *Neurophysiologie Clinique. Clinical Neurophysiology*, 46(4–5), 287–305. doi:10.1016/j.neucli.2016.07.002 PMID:27751622
- Katsigiannis, S., & Ramzan, N. (2017). Dreamer: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. *IEEE Journal of Biomedical and Health Informatics*, 22(1), 98–107. doi:10.1109/JBHI.2017.2688239 PMID:28368836
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International Affective Picture System (IAPS): Instruction manual and affective ratings, Technical Report A-8*. Gainesville: The Center for Research in Psychophysiology, University of Florida.
- Le, V. D., Carthen, C. D., Kamaruddin, N., Tavakkoli, A., Dascalu, S. M., & Harris, F. C., Jr. (2023). Generalized EEG data acquisition and processing system. *Proceedings of the 20th International Conference on Information Technology: New Generations*. Springer. doi:10.1007/978-3-031-28332-1_20
- OpenBCI. (n.d.-a). *OpenBCI_LSL*. OpenBCI. https://github.com/openbci-archive/OpenBCI_LSL
- OpenBCI. (n.d.-b). *The OpenBCI GUI*. OpenBCI. https://github.com/OpenBCI/OpenBCI_GUI
- OpenBCI Shop. (n.d.-a). *All-in-one EEG electrode cap starter kit*. OpenBCI. <https://shop.openbci.com/products/openbci-eeeg-electrocap-kit>
- OpenBCI Shop. (n.d.-b). *All-in-one gel-free BCI electrode cap kit*. OpenBCI. <https://shop.openbci.com/products/all-in-one-gelfree-bci-electrode-cap-kit>
- Russell, J. A. (1979). Affective space is bipolar. *Journal of Personality and Social Psychology*, 37(3), 345–356. doi:10.1037/0022-3514.37.3.345
- Thammasan, N., Moriyama, K., Fukui, K.-i., & Numao, M. (2017). Familiarity effects in EEG-based emotion recognition. *Brain Informatics*, 4(1), 39–50. doi:10.1007/s40708-016-0051-5 PMID:27747819
- Xu, H., & Plataniotis, K. N. (2012). Affect recognition using EEG signal. *Proceedings of the IEEE 14th International Workshop on Multimedia Signal Processing*, (pp. 299–304). IEEE.

Zachary Estreito is currently a master's student in computer science and engineering at the University of Nevada, Reno. Zach works part-time as both a graduate teaching assistant and a research assistant. He has been a teaching assistant for the Computer Science and Engineering Capstone Projects course as well as the Human-Computer Interaction course. As a research assistant, he works on an NSF-funded interdisciplinary research project involving the collection, processing, storage, visualization, and exportation of climate data collected from the Lake Tahoe Basin. His own personal research is focused broadly around the field of human-computer interaction, specializing in brain-computer interfaces and data science. He hopes to one day become a tenured computer science professor at a research institution. Vinh Le is a PhD student under the advisement of Dr. Sergiu Dascalu and Dr. Frederick Harris, Jr., at the University of Nevada, Reno (UNR). His interests lie primarily in human-computer interaction, software systems, and web development. Vinh defended his master's thesis on the topic of microservice architecture for envirosensing projects. He has also worked professionally as a senior developer working on cyberinfrastructure in state licensing for over 20 states, including Nevada, Texas, Louisiana, and California. Currently, Vinh is one of three capstone instructors for the Computer Science Department at UNR. Vinh has a great passion for teaching and research and aspires to one day be a professor at a research university. Frederick C. Harris, Jr., is a Foundation Professor in the Department of Computer Science and Engineering and associate dean in the College of Engineering at the University of Nevada, Reno. He received his BS and MS degrees in mathematics and educational administration from Bob Jones University and went on to receive his MS and PhD degrees in computer science from Clemson University. He is co-director of the Software Systems Lab at UNR. He is also the Nevada State EPSCoR director and the project director for Nevada NSF EPSCoR. He has published more than 300 peer-reviewed journal and conference papers along with several book chapters and has edited or co-edited 14 books. He has had 14 PhD students and 81 MS thesis students finish under his supervision. His research interests are in parallel computation, simulation, computer graphics, and virtual reality. He is a senior member of the ACM. Sergiu Dascalu is a professor in the Department of Computer Science and Engineering at the University of Nevada, Reno. He received a master's degree in automatic control and computers from the Polytechnic of Bucharest, Romania, (1982) and a PhD in computer science from Dalhousie University, Canada (2001). He is the co-director of the Software Systems Lab at UNR. His main interests are in software engineering, human-computer interaction, and data science. He has worked on many research projects funded by federal agencies such as NSF, NASA, and DoD-ONR. He has published over 250 peer-reviewed papers. Also, he received several awards, including the 2011 UNR F. Donald Tibbitts Distinguished Teacher of the Year Award and the 2019 UNR Vada Trimble Outstanding Graduate Mentor Award. He has been a panelist for several NSF program solicitations and a reviewer for over 15 journals. He is a senior member of the ACM.

Vinh Le is a PhD student under the advisement of Dr. Sergiu Dascalu and Dr. Frederick Harris, Jr. at the University of Nevada, Reno (UNR). His interests lie primarily in Human-Computer Interaction, Software Systems, and Web Development. Vinh defended his Master's Thesis on the topic of Microservice Architecture for Envirosensing Projects. He has also worked professionally as a Senior Developer working on cyberinfrastructure in state licensing for over 20 states, including Nevada, Texas, Louisiana, and California. Currently, Vinh is one of three Capstone Instructors for the Computer Science Department at UNR. Vinh has great passion for teaching and research, and aspires to one day be a professor at a research university.

Frederick C. Harris, Jr. is a Foundation Professor in the Department of Computer Science and Engineering and Associate Dean in the College of Engineering at the University of Nevada, Reno. He received his BS and MS degrees in Mathematics and Educational Administration from Bob Jones University, and went on and received his MS and Ph.D. degrees in Computer Science from Clemson University. He is co-Director of the Software Systems Lab at UNR. He is also the Nevada State EPSCoR Director and the Project Director for Nevada NSF EPSCoR. He has published more than 300 peer-reviewed journal and conference papers along with several book chapters and has edited or co-edited 14 books. He has had 14 PhD students and 81 MS Thesis students finish under his supervision. His research interests are in parallel computation, simulation, computer graphics, and virtual reality. He is a Senior Member of the ACM.

Sergiu Dascalu is a Professor in the Department of Computer Science and Engineering at the University of Nevada, Reno. He received a Master's degree in Automatic Control and Computers from the Polytechnic of Bucharest, Romania (1982) and a PhD in Computer Science from Dalhousie University, Canada (2001). He is the co-Director of the Software Systems Lab at UNR. His main interests are in software engineering, human-computer interaction, and data science. He has worked on many research projects funded by federal agencies such as NSF, NASA, and DoD-ONR. He has published over 250 peer-reviewed papers. Also, he received several awards, including the 2011 UNR F. Donald Tibbitts Distinguished Teacher of the Year Award and the 2019 UNR Vada Trimble Outstanding Graduate Mentor Award. He has been a panelist for several NSF program solicitations as well as reviewer for over 15 journals. He is a senior member of the ACM.