

Classification of Affective States via EEG and Deep Learning

Jason Teo, Lin Hou Chew, Jia Tian Chia, James Mountstephens

Faculty of Computing and Informatics

Universiti Malaysia Sabah

Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia

Abstract—Human emotions play a key role in numerous decision-making processes. The ability to correctly identify likes and dislikes as well as excitement and boredom would facilitate novel applications in neuromarketing, affective entertainment, virtual rehabilitation and forensic neuroscience that leverage on sub-conscious human affective states. In this neuroinformatics investigation, we seek to recognize human preferences and excitement passively through the use of electroencephalography (EEG) when a subject is presented with some 3D visual stimuli. Our approach employs the use of machine learning in the form of deep neural networks to classify brain signals acquired using a brain-computer interface (BCI). In the first part of our study, we attempt to improve upon our previous work, which has shown that EEG preference classification is possible although accuracy rates remain relatively low at 61%-67% using conventional deep learning neural architectures, where the challenge mainly lies in the accurate classification of unseen data from a cohort-wide sample that introduces inter-subject variability on top of the existing intra-subject variability. Such an approach is significantly more challenging and is known as subject-independent EEG classification as opposed to the more commonly adopted but more time-consuming and less general approach of subject-dependent EEG classification. In this new study, we employ deep networks that allow dropouts to occur in the architecture of the neural network. The results obtained through this simple feature modification achieved a classification accuracy of up to 79%. Therefore, this study has shown that the use of a deep learning classifier was able to achieve an increase in emotion classification accuracy of between 13% and 18% through the simple adoption of the use of dropouts compared to a conventional deep learner for EEG preference classification. In the second part of our study, users are exposed to a roller-coaster experience as the emotional stimuli which are expected to evoke the emotion of excitement, while simultaneously wearing virtual reality goggles, which delivers the virtual reality experience of excitement, and an EEG headset, acquires the raw brain signals detected when exposed to this excitement stimuli. Here, a deep learning approach is used to improve the excitement detection rate to well above the 90% accuracy level. In a prior similar study, the use of conventional machine learning approaches involving k-Nearest Neighbour (kNN) classifiers and Support Vector Machines (SVM) only achieved prediction accuracy rates of between 65% and 89%. Using a deep learning approach here, rates of 78%-96% were achieved. This demonstrates the superiority of adopting a deep learning approach over other machine learning approaches for detecting human excitement when immersed in an immersive virtual reality environment.

Keywords—*Neuroinformatics; emotion classification; preference classification; excitement classification;*

electroencephalography (EEG); deep learning; virtual reality; dropouts.

I. INTRODUCTION

We have conducted a number of prior investigations into the use of electroencephalography (EEG) as a method for passively monitoring the brainwaves of users as they are exposed to 3D visual stimuli as well as immersive stimuli and then using different machine learning algorithms to predict their preferences among the various visual stimuli [1], [2]. In the first part our study, we focus on human preference classification. The ability to passively identify the preferences of users as they are being presented with different stimuli will have novel and significant applications in various choice-based domains such as neuromarketing, affective entertainment, virtual rehabilitation and forensic neuroscience.

In our early work with a small set of five test subjects, good classification rates of up to 80% were attained using simple k-nearest neighbor (kNN) classifiers [1]. However, when the number of test subjects was increased to 16, the noise arising from inter-subject variability became a substantial factor which made the classification process significantly more challenging [2]. While most studies generally deal only with intra-subject variability where for each user, retraining is required before classification testing. We attempt a cohort-wide classification to enable direct applications to new users without the need for per-person pre-training before classification usage. In the our expanded study, classification rates for the large majority of conventional classifiers such as kNN, support vector machines, Naïve Bayes, Random Forest, C4.5 and other rule-based classifiers were only between 56-60%. The best classification result obtained from this comparative study was using deep neural networks at 64% [2].

The second part of our study focuses on excitement detection in immersive environments since much less is known about human emotion recognition in fully immersive environments such as virtual reality (VR). VR environments provide an arguably more effective emotion stimulating environment since users are fully immersed in the stimulus environment without any distracting views and/or other stimuli such as those present when using conventional displays such as computer and TV screens. Furthermore, users are free to move their heads to fully view their VR environments, which is more akin to their real-world viewing experience, hence suggesting the possibility of greater emotional response correlation with real life experiences. Additionally, as VR continues to garner

widespread adoption among everyday consumers, the ability to incorporate an effective emotion recognition system for VR applications will open up a wealth of novel interactions between the user and the VR experience particularly in the video gaming, live events, video entertainment, retail, real estate, healthcare, education, military and engineering domains [3].

As such, the main objective of the study is to investigate the various architectural tuning of the deep neural networks for improving the classification rates of our EEG-based preference classification as well as excitement classification task. Section II presents the background on emotion classification. Section III presents our approach to EEG-based preference and excitement classification using visual stimuli. Section IV presents the results of our investigations and Section V concludes the paper with some future avenues for expanding upon the current work.

II. BACKGROUND

A. Emotion Modeling and Classification

Emotion classification entails the use of various physiological signals and markers in an attempt to identify different emotions such as the user being in a state of anger, disgust, happiness, sadness, fear, anxiety, excitement and surprise among other [4], [5]. Some commonly measured bio-signals include the heart rate, skin conductance, pupil dilation, respiration rate and also brainwaves, which is also known as EEG [6], [7].

EEG-based emotion classification typically involves the measurement of the millivolt-range electrical signals through the placement of a number of electrodes on the scalp of the user, the waveforms of which are then spectrally transformed into features used by machine learning algorithms trained on labelled data to predict the emotion currently being sensed. Numerous studies have shown that classifications for various emotions can be reliably obtained using EEG.

B. Emotion Classification of Preferences

Preference classification can be considered a sub-task of emotion classification. This more specific task entails the identification of a user's like or dislike when presented with a stimulus. Preference classification is generally considered to be more challenging to classify compared to other emotions that are more strongly evoked such as anger or sadness.

The very large majority of EEG-based preference classification has been conducted using music as the stimulus [8], [9]. There have been very limited studies done using 2D images [10], [11] whereas our earlier studies were the first to implement rotating virtual 3D images as the stimulus [1], [2]. Furthermore, preference classification, which is already more challenging compared to other forms of emotion classification due to its comparatively weaker evocation, is rarely studied as a cohort-wide classification task. EEG-based emotion classification with large-sized cohorts will typically yield significantly lower accuracy rates due to inter-subject [12] and as well as intra-subject variability [13]. Doing so requires the classifier to be able to overcome inter-subject variability in addition to intra-subject variability of the users' EEG signal.

Consequently, the weak signal evocation and inter-subject variability make EEG preference classification a very challenging classification task.

C. Extraction of Features from EEG Signals

Emotion modeling using machine learning approaches can be categorized into three broad domain classes: (i) time, (ii) frequency, and (iii) time and frequency combination. Time-based emotion modeling employs the detection of event-related potentials (ERPs). Of these, they can be further divided into groups that are detected based on whether they are having short, medium or long post-latency exposures after stimuli presentation. Emotion classification for valence and arousal produced accuracy rates of 55.7% for arousal and 58.8% for valence [14] when using these ERP-based methods.

The classification of emotions based on the frequency domain is achieved through the learning of features obtained power spectrum analysis, producing the canonical delta, theta, alpha, beta and gamma frequency bands. Emotion classification for the preference of music produced an accuracy of 74.8% with linear support vector machines (SVMs) using the preprocessed features obtained through the Common Spatial Patterns (CSPs) method [15]. Emotion classification for the preference of music via preprocessed features obtained from a using a conventional Fast Fourier Transform (FFT) produced a classification accuracy rate of 85.7% using SVMs [16]. Radial SVMs were used in the only published emotion classification of preferences not using music stimuli, in this case for 2D image preferences using power spectrum analysis where the classification outcome produced an accuracy of 88.5% [17].

From the perspective of using a combination of time and frequency (TF) leverages on the power spectrum analysis at predefined time periods that encompass the whole duration of the post stimuli period for measuring brain activity. Several conventional machine learning algorithms were used to conduct emotion classification tasks employing three distinctly different TF analysis methods were studied to identify the preference for music. Here, it was observed that the k-Nearest Neighbors (kNN) machine learning approach produced the overall best outcome with an accuracy of 86.5% [18]. The same group of researchers then conducted a follow-on investigation utilizing a much finer-grained approach which attempted to categorize the emotion stimuli into two groups: (i) familiar versus, (ii) unfamiliar music. In this later study, using a kNN machine learning approach, they managed to produce a much higher emotion classification accuracy of 91.0% [19]. Emotion modeling for the preference of music using TF approaches used a Short-Time Fourier Transform (STFT) and using a kNN machine learning approach produced emotion classification accuracy rates of 98.0% [20].

D. Preference Classification using Deep Learning Approaches

The preferences of 32 participants for the viewing of music video clips was attempted using deep learning via the Deep Belief Networks (DBNs) approach [21]. DBNs accomplish deep learning through the stacking of various Restricted Boltzmann Machines (RBMs) on top of each other. In this method of deep learning, the output obtained from a lower-

level RBM is subsequently utilized to serve as the input to a higher-level RBM. This process is continued progressively through deeper and deeper layers thus forming a multi-layer stacking of these so-called RBMs. An average emotion classification accuracy of 77.8% was obtained where this method performed significantly better than various types of different SVMs as well as standard non-stacked RBMs.

A limited study involving only 6 subjects was reported for the emotion classification of participants when presented with the stimuli of viewing a number of short video clips for the elicitation of emotions with positive or negative valences [22]. In a novel approach for the emotion classification task which utilizes only the top five EEG recording electrodes, the investigation produced emotion classification accuracies of 87.6% using DBN's with this novel critical feature channel selection method. These results were observed to perform better than Extreme Learning Machines (ELMs) as well as SVMs and at the same time was observed to perform significantly better than the kNN machine learning approach. However in both of these two reported studies, it is important to point out that the training and classification prediction tasks were accomplished on a per-subject basis and not over the entire cohort of participants, which means that this only caters for intra-subject variability and not inter-subject variability. In other words, these two studies utilized an approach that requires the retraining of machine learning classifiers during the training phase whenever there is a new participant before the emotion classification prediction task can be performed. Essentially what this intra-subject or subject-dependent method employs is an approach that bypasses the difficulty of handling inter-subject variability and only caters for intra-subject variability, which means that it will not work for subject-independent classification tasks.

From the literature survey, there was only one paper found in which the deep learning approach was used in emotion modeling to classify preferences in a subject-independent methodology. Here it was reported that using a combination of unsupervised learning employing stacked autoencoders (AEs) in conjunction with the supervised learning of softmax machine learning classifiers was able to perform prediction of the emotional states for 32 participants for valence and arousal. Nonetheless, this paper reported the requirement of utilizing an extremely large number of hidden neurons in the deep learning classifier. It is interesting to note that the authors themselves alluded to the fact that an extended amount of computational time was utilized during the training phase with such an approach. Subsequently the authors hybridized this approach with feature preprocessing routines employing Principal Component Analysis (PCA) as well as Covariate Shift Adaptation (CSA) during the pre-learning process. However, even with the extended processing time and numerous augmentations with supplementary preprocessing, the emotion modeling was only able to produce very low prediction accuracy rates of 53.4% and 52.0% for valence and arousal classification, respectively from this subject-independent approach using leave-one-out cross-validation (LOOCV) [23]. What this study clearly demonstrates is the fact that inter-subject variability very significantly and critically adds tremendous difficulty to the classification of emotions based on

preferences when compared against the much more common and significantly easier prediction task of subject-dependent studies that only caters for intra-subject variability in the learning of the EEG-based emotion modeling.

E. Classification of Affective States in Virtual Reality and Mixed Reality Environments

There have been very few studies that have conducted human emotion recognition that have used virtual reality environments as the stimulus. To the best of our knowledge, there has yet to be any study that uses solely EEG to detect human emotions using purely VR stimulus.

Wu et al. [24] used a Virtual Reality Stroop Task (VRST) from the Virtual Reality Cognitive Performance Assessment Test (VRCPAT) to detect arousal levels in their attempt to identify various affective/cognitive states. A number of VR stimuli were presentations with various levels of arousal were selected from the VRST. It was shown that a relatively high classification accuracy rate of 96.5% using support vector machines (SVM) could be achieved through VR stimuli. However, the study used an elaborate and involved sensor setup with a wide range of psychophysiological responses which included skin conductance level, respiration, ECG, as well as EEG were used to conduct the emotion recognition task. As such, it remains unknown if a much simpler setup involving EEG alone would be feasible in achieving successful emotion recognition.

Massari et al. [25] and Kovacevic et al. [26], respectively used mixed reality stimuli to conduct brain state recognition based solely on EEG signals as input to the classification system. Massari et al. utilized their proprietary eXperience Induction Machine (XIM) as the mixed reality stimuli system to classify different brain states for spatial navigation, reading and calculation, achieving the best results of 86% using linear discriminant analysis [25]. Kovacevic et al. implemented an EEG-based mental state recognition system as part of an immersive and interactive multi-media science-art installation using the recognition of relaxation and concentration mental states of its participants to determine the audio-visual output of a dome-based artistic installation comprising video animations that were projected on to the 360° surface of the semitransparent dome as well as the generation of soundscapes based on pre-recorded sound libraries and live improvisations [26]. Although both these studies utilized EEG solely as the feature input, these studies were not specifically classifying emotional states and both were utilizing mixed reality stimuli rather than pure virtual reality stimuli. As such, it remains unknown whether a purely VR-based stimulus system could be successfully used for emotion recognition.

III. METHODOLOGY

A. Experimental Setup for Preference Classification

Emotion classification entails the use of various physiological signals and markers in an attempt to identify different emotions such as the user being in a state of anger. In this, first of this investigation for preference classification, 16 subjects (8 female and 8 male, mean age = 22.44) were involved where all the participants had corrected-to-normal or normal vision. Furthermore, they were asked and confirmed

to be free of any known history of psychiatric illnesses prior to the participation in the study. The participants were briefed on what to expect in terms of the BCI equipment that was to be used during the data acquisition phase before the actual experimentation was to proceed. The EEG acquisition device was a brain-computer interface (BCI) headset called the ABM B-Alert X10, which has nine active electrodes, namely the POz, Fz, Cz, C3, C4, F3, F4, P3 and P4 channels according to the standard 10-20 naming convention where a subject participant wearing the said BCI headset is depicted in Fig. 1. MATLAB, Java and R were the three programming languages used. The visual stimuli were developed and displayed using the Java programming language. Integration between the visual stimuli and the BCI headset was accomplished by implementing the MATLAB programming language with the B-Alert X10's SDK. Finally, the statistical programming language R was used for the signal preprocessing phases, feature extraction, and finally for the training and prediction classification tasks.

The data acquisition processes experienced by the participants are as shown in Fig. 2 where during the commencement of the data acquisition process, a blank screen of three seconds is shown to the participant to obtain the base resting brain signal in order to avoid any brain activities related to the previous stimuli during the actual emotion modeling trial phase. After this blank screen, there will be between five to fifteen of actual viewing time for the 3D stimuli where the minimum viewing time and maximum viewing time is set between five and fifteen seconds respectively. The participant is allowed to commence to the following rating state based on their own choosing after the minimum viewing duration time of five seconds while once the maximum viewing duration time is up, the system will proceed by default to the next rating state. The purpose of implementing this particular method of the data acquisition process flow is to allow the participant to decide on their own accord during the stimuli viewing time so as to mitigate the possibility of boredom from setting in and making the participant fatigued while viewing the stimuli during the data acquisition process since requiring the participant to continuously view only at fixed intervals in a repetitive manner for the purposes of rating the stimuli could possibly cause the participant to experience boredom which will subsequently lead to further fatigue towards the end of the data acquisition process. As such, since the participant is no longer required to just wait until the maximum set and fixed time in order to conduct the rating, this essentially provides the participant with the freedom and ability to shift to the following visual stimuli, which will potentially save some overall viewing time and at the same time prevent the participant from fatiguing. A rating system containing a discrete scale of 1-5, where 1 represents like very much; 2 represents like; 3 represents undecided; 4 represents do not like; and finally 5 represents do not like at all, is shown to the participant at the conclusion of the visual shape stimuli viewing period.



Fig. 1. The medical-grade 9-channel EEG acquisition device is shown being worn by a participant in the study.

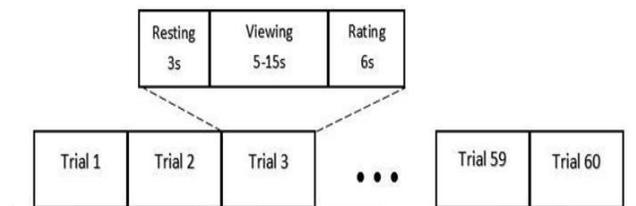


Fig. 2. The flow of the data acquisition process as experienced by the participants during the experimentation.

$$r(\theta) = \frac{1}{n1 \sqrt{\left(\left(\frac{1}{a} \cos\left(\frac{m}{4} \theta\right)\right)\right)^{n2} + \left(\left(\frac{1}{b} \sin\left(\frac{m}{4} \theta\right)\right)\right)^{n3}} \quad (1)$$

The Gielis Superformula is used to generate three-dimensional shapes which were used as the visual stimuli in this study and had the visual appearances of a bracelet-like virtual generated [27], the mathematical formula of which is as shown in (1). Our main reason for choosing this shape as the three dimensional visual stimuli for evoking emotions is to determine the aesthetic quality of jewelry-type objects since visual aesthetic quality is primarily the key motivating factor when one decides whether or not make a purchase of such an item. By modifying the various superformula parameters, the generation of different and myriad natural three dimensional virtual shapes can be generated.

Sixty different bracelet-like shapes generated and used in this study is as shown in Fig. 3, which were generated by utilizing different parameters with randomly generated values in the superformula. Through preliminary testing, different ranges of suitable parameter values were chosen to synthesize virtual three dimensional shapes that possess visual characteristics of a bracelet-like shape. These three dimensional bracelet-like shapes were then shown to the participants virtually on a computer. The visual system allowed the presentation of the three dimensional virtual shapes with rotations on different axes of the presented stimuli so that it could be viewed at different angles in order for the participant to be able to fully visualize the generated three dimensional bracelet-like shapes.

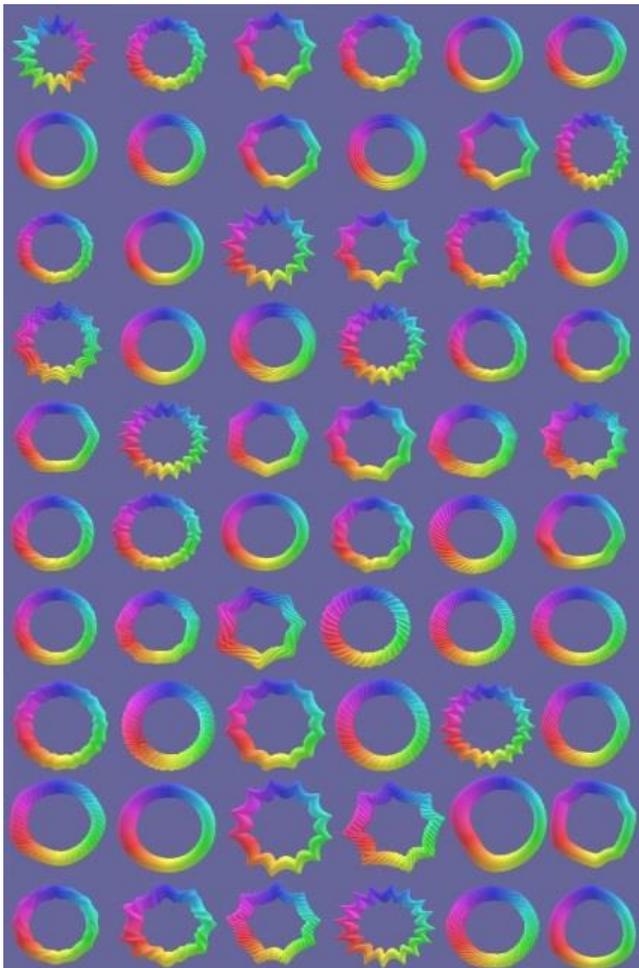


Fig. 3. The Gielis Superformula used to produce 60 bracelet-like shapes.

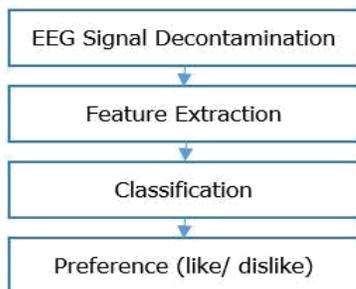


Fig. 4. Summary of the signal processing process flow.

The major processes for preference classification are as shown in Fig. 4. Firstly, environmental and physiological artifacts are always present in EEG signal recordings and require decontamination. The SDK in the MATLAB programming language provided by ABM for the B-Alert X10 BCI headset automatically provides this decontamination function. A 50Hz notch filter removes environmental artifacts while five physiological artifacts comprising electromyography (EMG), eye blinks, excursions, saturations, and spikes are similarly removed automatically in real-time. The eye excursions, saturations, and spikes are replaced by zero values where they are later filled in using spline interpolation.

Subsequently, a Short-Time Fourier Transform (STFT) is then used to transform the decontaminated EEG signals into the TF domain where it decomposes each of the nine BCI channels into five spectral bands, which are the delta 1-3Hz, theta 4-6Hz, alpha 7-12 Hz, beta 13-30 Hz, and gamma 31-64Hz bands. These five bands across the nine channels thereby provides a total of forty-five input features. The brainwave recordings from the 16 participants where each viewed the sixty 3D visual stimuli of the bracelet-like shapes generated 960 observations altogether. However, only 208 observations were used during the training and prediction classification process. These were the strongest ratings on the ratings scale of 1, which represented like very much, and 5, which represented do not like at all, respectively. A final dataset matrix comprising forty-seven feature columns consisting of the observation ID reference, participant rating, and each of the forty-five TF features, over two hundred and eight rows of selected observations served as the training and testing data for the respective machine learning classifiers. Moreover, the subjects' baseline readings acquired while in the resting state were subtracted from the stimuli viewing state values before the values were utilized in the prediction classification process.

The deep neural networks utilized were set to two hundred hidden neurons within each of the two hidden layers using the uniform adaptive method [28] for weight matrix initialization. Preliminary experimentation showed that this setup with the number of hidden layers as well as the number of hidden neurons per layer provided the optimal settings for this preference prediction task. Cross-entropy [29] was used as the error function during the 10-fold cross-validation, which was conducted for 10 epochs in each of the cross-validation steps.

B. Experimental Setup for Excitement Classification

Fig. 5 describes the overall approach adopted in conducting this second study of excitement classification in virtual reality consisting of a number of distinct phases, each of which will be explained in the following subsections below.

1) Experiment Stimuli

In this project, the immersive stimuli were created using Google's Cardboard VR technology and a 360° video available on YouTube.com. The selected video was the experience video of the 360-degree ride on a roller coaster that is promising in eliciting excitement emotion with providing sensations of a roller coaster ride which drops from high peaks and high speed 360-degrees turns. Screenshots of these stimuli is as shown in Fig. 6 and 7.

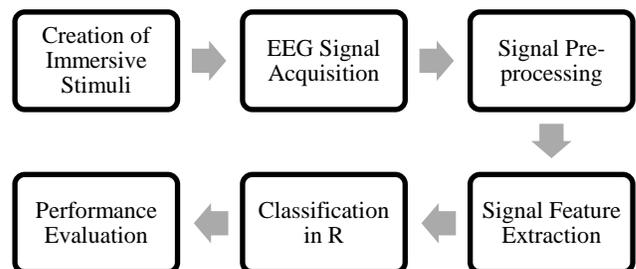


Fig. 5. Overall process of human mental states classification.



Fig. 6. Scenes in the stimuli video.



Fig. 7. Scenes in the stimuli video.

2) Experimental Test Subjects

A total of 24 human subjects (12 females, 12 males) participated in this study and had normal or corrected-to-normal vision with no history of psychiatric illness. The age of the subjects was in the range of 20 to 28 years old. During the experimental session, subjects were advised to sit comfortably on the chair without any restriction to head movements which being immersed in the virtual reality stimuli. An image of a test session in progress with the test subject wearing the VR headset and EEG headband is as shown in Fig. 8.



Fig. 8. Experiment setup using human test subject.

3) Data Acquisition Device

To increase the applicability of EEG-based predictive analytics in human mental state classification, the Muse brain sensing headband from Interaxon was used as the data acquisition device since it is trivial to set up and comes at a much lower cost compared to medical-grade conventional EEG devices. Conventional EEG devices such as the B-Alert X10 by ABM that uses adhesive sponge discs with the requirement of applying electrode gels require a significantly longer set up time for individual electrodes and such a setup often limits the behavioural freedom on the participants since there are numerous connecting wires to connect to the electrodes which severely restricts head movements as necessary in immersive VR environments. The Muse headband is extremely accessible as it is wireless, lightweight, flexible, adjustable and easy application. The wearer that puts on the Muse headband will not experience any limitations on their mobility as the Muse headband is connected through wireless Bluetooth technology for data transmission. The Muse headband has four dry electrode channels at international standard 10-20 coordinates of TP9, AF7, AF8 and TP10 as illustrated in Fig. 9. The earpieces of Muse are adjustable but the headband area with channel electrodes AF7, AF8 and reference channel Fpz are not flexible. Although the Muse headband is still new in the market as a commercialised EEG device, there have been other studies that have reported on its potential to be used as a research tool despite its limited numbers of electrodes and low signal resolution [30], [31].

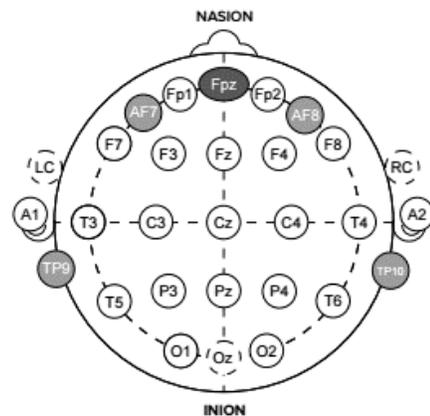


Fig. 9. Muse electrode locations by 10-20 International Standards.

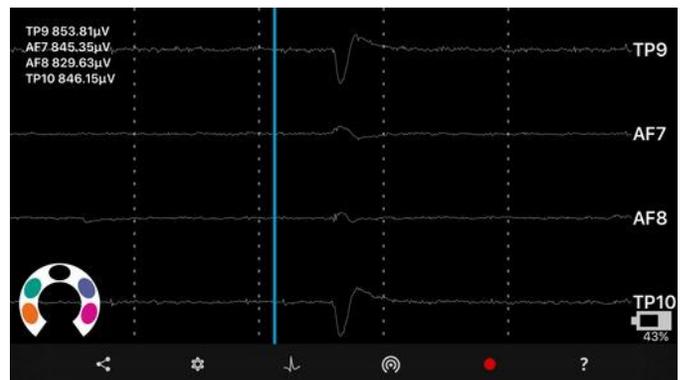


Fig. 10. Signal acquisition and recording screen of the Muse Monitor app.

4) EEG Recording

The EEG recordings were acquired using the Muse Monitor app available from Google’s Android Play Store. The EEG recordings were exported in CSV file format from Muse Monitor. The real-time EEG signal is recorded with an interval of 0.5 seconds, providing 256 data points per second for raw values to capture the minor changes of the brain rhythms. Fig. 10 shows the recording screen of the Muse Monitor app with the view of raw EEG values captured from each of the sensors (TP9, AF7, AF8, TP10) in microvolts (μV). In this study, two recordings were recorded from each experimental test subjects. One recording is for the “Rest” state where there is no video stimulus and subjects were asked to keep calm and breathe normally. The second recording is for the “Excited” state where subjects were wearing VR headset being immersed and stimulated by the 360° roller coaster video experience.

C. Signal Pre-processing and Feature Extraction

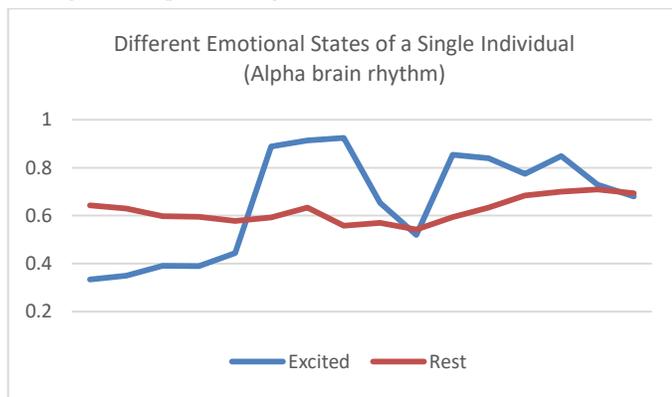


Fig. 11. Different emotional states of a single individual in Alpha brain rhythm representation.

Recorded EEG signals are always subject to artefacts and noise during acquisition. The common artefacts found are electromyography (EMG), eye-blinks, excursions, saturations, and muscle spikes. It is important to perform signal pre-processing to enhance the signal-noise power ratio [32]. The band of interest in this study are the frequency bands: delta (δ), theta (θ), alpha (α), beta (β) and gamma (γ), each reflecting different brain states of human experimental subject. Fast Fourier Transform (FFT) was used to convert the obtained EEG signal to a representation in the frequency domain based on Butterworth’s 4th Order Filter with different cut-off frequency thresholds to extract the five frequency bands [32] as in (2):

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}} \quad (2)$$

where $k = 0, 1, 2, \dots, N-1$, X_k is the FFT coefficients, N is the total number of input EEG samples, n is the total number of points in FFT. There are two EEG recordings per subject, one is the “Rest” state, and the other one is the “Excited” state. Fig. 11 shows the two Alpha brain rhythms of different states from a single individual. In the “Rest” state recordings, a length of 16 data points was extracted for classification. While in the “Excited” state recordings, two sets of 16 data points were extracted in accordance with the two excitement eliciting

events in the video stimulus: (1) the drop of the roller coaster from the highest peak and (2) the high speed 360° degree turns of the roller coaster. In conclusion, 3 sets of data points were extracted from each experimental subject, giving a total of 72 objects for classification. The extracted data was then tabulated according to each band as features of interest for classification.

The classification work in this project was performed in the R environment as R is the leading statistical analysis tool that includes a large collection of packages that provides a wide variety of linear and non-linear modelling, classification function and etc. The main package used to build classification models was the ‘caret’ package as it has a consistent syntax for various machine learning methods. Additionally, the ‘caret’ package also provides an easy implementation to perform the 10-fold cross validation on the classification model. Since the ‘caret’ package made it easy to expand the range of tuning parameters of the machine learning methods, this experiment had systematically investigated various parameter settings for each classifier used.

1) *K-Nearest Neighbour (KNN)*: KNN is a simple and intuitive method of classifier used in many research works typically for classifying signals and images. KNN classifies objects based on the similarity between two instances to locate the nearest neighbour. The classifier will compare a newly labelled sample with the baseline data. The decision rule applied will vote where the new labelled sample will be assigned based on the class of the majority of the k-nearest neighbours.

2) *Support Vector Machine (SVM)*: SVM function attempts to find a hyperplane in between the groups of objects to classify them. The SVM operates by minimising the loss function as in (3):

$$\min_w \frac{1}{2} w^T w + C \sum_{i=1}^l \max(0, 1 - y_i w^T \phi(x_i)) \quad (3)$$

where w is the vector of weights, C is cost parameter, $\phi(x)$ is a kernel function applied on the input data.

The Radial Basis Function (RBF) kernel will be applied with SVM to enable operations performed in the input space rather than the potentially higher dimensional feature space [33] as in (4):

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (4)$$

where $\|x - x'\|^2$ is the square of the Euclidean distance between the two vectors, γ is the kernel parameter, equivalent to $\frac{1}{2\sigma^2}$ where σ is a free parameter: the inverse kernel width for RBF kernel. There are two tuneable parameters in this function used: C and σ .

3) *Random Forest (RF)*: RF is an ensemble classifier that operates by constructing a multitude of decision trees [34]. The final predicted class for a test example is obtained by combining the predictions of all individual trees. The decision tree with controlled variance was constructed through a combination of bootstrap aggregation (bagging) and random feature selection. Each node in RF is split using the best

among a subset of predictors that are randomly chosen at the node. This strategy makes the classifier perform better compared to other classifiers such as SVM, neural networks and Linear Discriminant Analysis (LDA) and is robust against over-fitting [35]. There is only one tuning parameter for RF: mtry (number of variables randomly sampled as candidates at each split).

4) *Feed-forward Neural Network (NN)*: NN is the first and simplest form of artificial neural network developed. The network's information only moves in one direction, from the input nodes through the hidden nodes (if any) and to the output nodes in forward directions. There exist no cycles or loops in the network. The simplest design of NN is a single-layer perceptron network that consists only one layer of output nodes. The inputs are fed directly to the outputs via a series of weights and the output units are of the same form but with an output function [36]:

$$y_k = \phi_0 \left(\alpha_k + \sum_h \omega_{hk} \phi_h \left(\alpha_h + \sum_i \omega_{ih} x_i \right) \right) \quad (4)$$

The activation functions ϕ_0 and ϕ_h are taken to be the logistic function:

$$l(z) = \frac{\exp(z)}{1 + \exp(z)} \quad (5)$$

There are two tuning parameters in NN: size (number of hidden units) and decay (weight decay).

5) *C5.0 Decision Tree & Rule-based Model (C5.0)*: This algorithm was developed based on the C4.5 algorithm. C5.0 can be applied for classification as a decision tree or rule-based model. It supports boosting with any number of trials and can automatically winnow the attributes to remove those attributes that may be obstructive. For high-dimensional applications, this winnow features can lead to smaller classifiers and higher predictive accuracy while minimising the time required to generate rule sets. C5.0 has three tuning parameters: model (choose between decision tree and rule-based model), winnow (decision on whether predictor winnowing should be used) and trials (number of boosting iterations).

IV. RESULTS AND DISCUSSION

A. Preference Classification Result

Four distinct deep net architectures were tested, which were the standard deep nets, deep nets with dropouts only, deep nets with L1 regularizations only and finally deep nets with both dropouts and L1 regularizations. In L1 regularizations, λ is set at 10^{-5} . For dropouts, we set the hidden layer dropout probability at 0.5. For each of these architectures, we also

paired them with different activation functions for the hidden layers, which were the tanh, maxout and rectified linear unit (ReLU) activation functions. The rectified linear activation function [37] was used with an adaptive learning rate method [38]. The results of this specific part of the study have been previously published [39].

Table I presents the 10-fold cross-validation results obtained from using the various deep net architectures as well as with dropouts and L1 regularization terms. The best classification was obtained using the deep net with dropout architecture using rectified linear units for activation at 79.76%. The second best classification result was also obtained using the deep net with dropout architecture but using the tanh activation at 74.38%. This was followed next with the deep net architecture using both dropouts and L1 regularization with the rectified linear unit and tanh activations, respectively at 72.44% and 72.43%. The lowest classification obtained was 54.92% using the deep net with L1 regularization and maxout activation. As can be seen from Fig. 12, a very significant improvement in classification accuracy was attained using the deep net with dropouts compared to the earlier work which did not make use of any dropouts and/or regularization, which was only between 61.15-67.68%. This is an improvement of over 10% and clearly shows the benefits of using dropouts to improve the generalization ability of deep nets.

TABLE I. EEG PREFERENCE CLASSIFICATION RESULTS

Deep Net Architecture	Hidden Layer Activation Function	Classification Accuracy (%)
Standard Deep Net	Tanh	67.68
Standard Deep Net	Maxout	61.15
Standard Deep Net	ReLU	63.99
Deep Net with Dropout	Tanh	74.38
Deep Net with Dropout	Maxout	67.71
Deep Net with Dropout	ReLU	79.76
Deep Net with L1 Regularization	Tanh	71.86
Deep Net with L1 Regularization	Maxout	54.92
Deep Net with L1 Regularization	ReLU	63.02
Deep Net with Dropout and L1 Regularization	Tanh	72.43
Deep Net with Dropout and L1 Regularization	Maxout	67.16
Deep Net with Dropout and L1 Regularization	ReLU	72.44

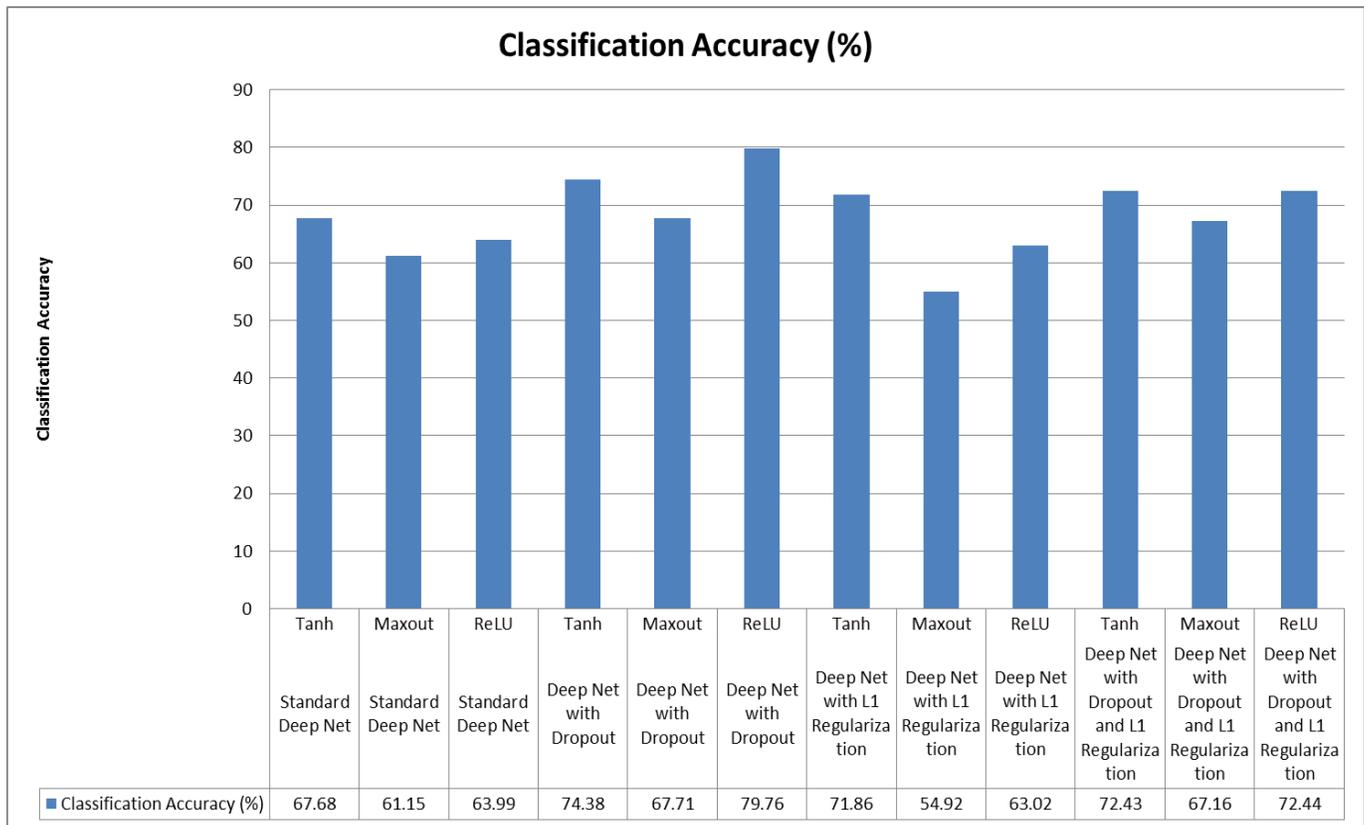


Fig. 12. Summary comparison of various deep learning architectures used.

B. Excitement Classification Results

Based on results shown in Table II, the SVM classifier achieved the overall best accuracy result of 89.36% using the Alpha band, while the second highest was 84.82%, achieved by the KNN classifier using the Theta band. Taking into account all datasets, the KNN classifier had the best performance as it held the most number of highest accuracy from 4 datasets (Delta, Theta, Beta, Gamma). However, the SVM classifier had a better average performance (SVM: 80.58%, KNN: 79.41%, RF: 78.92%, NN: 77.77%, C5.0: 75.54%).

Alpha band had shown the highest classification accuracy from two different classifiers (SVM and RF). This suggests that the Alpha band that represents the relaxed awareness of human contains some features that are useful to be used to classify the human emotion of “Excitement”. Moreover, the Theta band also showed a similar behaviour as the Alpha band. The Theta band tops the accuracy results on KNN and C5.0 classifiers and it represents the emotional stress, drowsiness and sleeps in adults.

In contrast as shown in Table III, the Gamma band had the worst overall results across all of the classifiers except NN classifier. This suggests that Gamma band that represents consciousness is not suitable to be used to classify human emotion of “Excitement” when the subject id immersed into the virtual stimuli.

For deep learning, preliminary testing yielded deep neural networks that performed best for this excitement classification

task using three hidden layers with 200 nodes each with weights initialized using the uniform adaptive method [29]. The deep neural networks were run using 10-fold cross-validation for 10 epochs each time using cross-entropy [29] as the error function and having a softmax output layer. Six different deep neural network architectures with different activation functions were tested, namely, tanh, maxout, and rectified linear (ReLU) [33], with and without dropout respectively, with dropout set at 0.5 and an adaptive learning rate method [34] applied when ReLU was used. The results obtained are tabulated below in Table IV.

TABLE II. SUMMARY OF TOP RESULTS OF 5 CLASSIFIERS

Classifier	Best Accuracy (%)	Band of Interest
SVM	89.36	Alpha
KNN	84.82	Theta
RF	81.96	Alpha
RF	81.96	All
NN	81.07	Beta
C5.0	80.89	Theta

TABLE III. SUMMARY OF BEST RESULTS OF ALL BANDS

Classifier	Accuracy of Band of Interest (%)					Combined
	Delta	Theta	Alpha	Beta	Gamma	
KNN	81.25	84.82	79.11	82.32	76.79	72.14
SVM	78.04	82.14	89.36	82.32	73.93	77.68
RF	79.62	79.29	81.96	78.21	72.50	81.96
NN	71.25	80.71	79.11	81.07	74.11	80.36
C5.0	72.50	80.89	80.36	75.36	65.36	78.75

TABLE IV. EXCITEMENT CLASSIFICATION RESULTS (%)

Activation	Delta	Theta	Alpha	Beta	Gamma	Combined
Tanh	89.77	93.71	82.57	88.21	86.21	92.07
Maxout	88.27	79.95	88.60	93.41	93.67	80.70
ReLU	85.44	88.18	88.67	87.82	86.44	91.01
Tanh with dropout	77.94	80.42	90.11	92.57	91.39	92.41
Maxout with dropout	83.68	83.12	82.82	88.96	88.71	91.42
ReLU with dropout	80.95	91.87	88.17	87.88	84.58	95.55

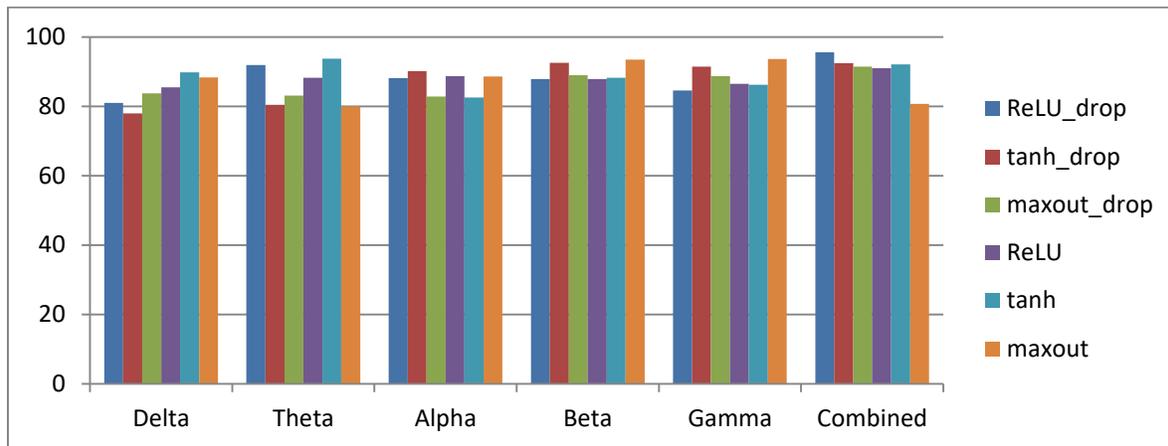


Fig. 13. Classification results grouped according to spectral band(s).

The best classification result of 95.55% was obtained using the ReLU with dropout deep neural network architecture using the combination of all of the available spectral bands. The next best result of 93.71% was obtained using the tanh with the theta band as the only input feature. The worst result of 77.94% was given by the tanh with dropout using the delta band. From the results, it appeared that there were no clear trends in terms of the architecture used but in terms of the spectral bands used, the combined approach appeared to provide an advantage whereby five out of six results yielded more than 90% accuracy as shown in Fig. 13. This suggests that, at least in terms of the excitement emotion, detecting this emotion benefits from looking at all spectral bands and not just at one or two specific bands such as alpha and beta which are commonly adopted for classifying EEG signals during active cognition.

V. CONCLUSION AND FUTURE WORK

Firstly, this study has comprehensively tested dropout and L1 regularization approaches to deep net architectures in an effort to improve the classification performance of deep learning neural networks in EEG-based preference classification. We have shown that using a deep net with dropouts using rectified linear units for activation was able to achieve a gain of more than 13%-18% at 79.76% accuracy compared to standard deep nets without such approaches at only between 61.15%-67.68% using various activations.

Secondly, this study has also investigated the use of deep learning for the detection of excitement while being immersed in virtual reality stimuli. To the best of our knowledge, this represents the first reported work that uses EEG solely as the input feature for the classification with the stimuli being virtual reality. It has been shown that a relatively high classification accuracy can be achieved with the best result yielding close to

96% accuracy. The results also suggest that using a combination of all EEG spectral bands as the input features provided more reliable classification results in general compared to using any other single EEG spectral alone.

For future work, due to the significant noise typically encountered in inter-subject EEG variations, we intend to investigate the use of autoencoders to pre-train the features extracted in order to further improve classification accuracy. Also, with the significant improvement in classification accuracy obtained through this study, we also plan to embark on application-based investigations into the use of EEG-based preference classification to guide automated generation of affective entertainment content in games, music and storytelling. It would be worthwhile to expand this line of work to include other emotions such as fear, boredom, frustration among others in view of expanding the potential applications of this EEG-based emotion classification in virtual reality approach particularly in the field of affective entertainment.

VI. ACKNOWLEDGEMENTS

This project is supported by the FRGS research grant scheme ref: FRG0349 & FRG0435 from the Ministry of Higher Education, Malaysia.

REFERENCES

- [1] L.H. Chew, J. Teo, and J. Mountstephens, "Aesthetic preference recognition of 3D shapes using EEG", *Cognitive Neurodynamics*, 10(2), pp.165-173, 2016.
- [2] J. Teo, L.H. Chew, and J. Mountstephens, "Deep learning for EEG-based preference classification", *International Conference on Applied Science and Technology (ICAST 2017)*, IEEE, April 2017.
- [3] H. Bellini, W. Chan, M. Sugiyama, M. Shin, S. Alam, and D. Takayama, excerpt from "Virtual & Augmented Reality: Understanding the Race for the Next Computing Platform", Profiles in Innovation, Goldman

- Sachs Equity Research, pp. 1-30, Feb. 2016, www.goldmansachs.com/our-thinking/pages/technology-driving-innovation-folder/virtual-and-augmented-reality/report.pdf, accessed 24 October 2017 (2016).
- [4] X.W. Wang, D. Nie, and B.L. Lu, "Emotional state classification from EEG data using machine learning approach", *Neurocomputing*, 129, pp.94-106, 2014.
- [5] A. Dhall, R. Goecke, J. Joshi, K. Sikka, and T. Gedeon, "Emotion recognition in the wild challenge 2014: Baseline, data and protocol", *Proceedings of the 16th International Conference on Multimodal Interaction*, pp.461-466, ACM, 2014.
- [6] G.K. Verma and U.S. Tiwary, "Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals", *NeuroImage*, 102, pp.162-172, 2014.
- [7] E.H. Jang, B.J. Park, S.H. Kim, M.A. Chung, M.S. Park, and J.H. Sohn, "Emotion classification based on bio-signals emotion recognition using machine learning algorithms", *2014 International Conference on Information Science, Electronics and Electrical Engineering (ISEEE)*, IEEE, vol. 3, pp.1373-1376, 2014.
- [8] S.K. Hadjidimitriou, A.I. Zacharakis, P.C. Doulgeris, K.J. Panoulas, L.J. Hadjileontiadis and S.M. Panas, "Revealing action representation processes in audio perception using fractal EEG analysis", *IEEE Transactions on Biomedical Engineering*, 58(4):1120-1129, 2011.
- [9] D.A. Adamos, S.I. Dimitriadis, and N.A. Laskaris, "Towards the bio-personalization of music recommendation systems: A single-sensor EEG biomarker of subjective music preference", *Information Sciences*, 343, pp.94-108, 2016.
- [10] M. Yadava, P. Kumar, R. Saini, P.P. Roy, and D.P. Dogra, "Analysis of EEG signals and its application to neuromarketing", *Multimedia Tools and Applications*, pp.1-25, 2017.
- [11] L.C. Chen, P. Sandmann, J.D. Thorne, C.S. Herrmann, and S. Debener, "Association of concurrent fNIRS and EEG signatures in response to auditory and visual stimuli", *Brain Topography*, 28(5), pp.710-725, 2015.
- [12] S. Goncalves, J. De Munck, P. Pouwels, R. Schoonhoven, J. Kuijter, N. Maurits, J. Hoogduin, E. Van Someren, R. Heethaar, F. L. Da Silva, "Correlating the alpha rhythm to bold using simultaneous EEG/fMRI: inter-subject variability", *Neuroimage* 30(1):203-213, 2006.
- [13] G. Pfurtscheller, C. Brunner, A. Schlogl, F. L. Da Silva, "Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks", *Neuroimage* 31(1):153-159, 2006.
- [14] A. Yazdani, J. S. Lee, J.-M. Vesin, T. Ebrahimi, "A ECT recognition based on physiological changes during the watching of music video", *ACM Transactions on Interactive Intelligent Systems 2 (EPFL-ARTICLE-177741)*, pp. 1-26, 2012.
- [15] Y. Pan, C. Guan, J. Yu, K. K. Ang, T. E. Chan, "Common frequency pattern for music preference identification using frontal EEG", in: *Neural Engineering (NER)*, 2013 6th International IEEE/EMBS Conference on, IEEE, pp. 505-508, 2013.
- [16] K. C. Tseng, B.-S. Lin, C.-M. Han, P.-S. Wang, "Emotion recognition of EEG underlying favourite music by support vector machine", in: *Orange Technologies (ICOT)*, 2013 International Conference on, IEEE, pp. 155-158, 2013.
- [17] Y. Kim, K. Kang, H. Lee, C. Bae, "Preference measurement using user response electroencephalogram", in: *Computer Science and its Applications*, Springer, pp. 1315-1324, 2015.
- [18] S. K. Hadjidimitriou, L. J. Hadjileontiadis, "Toward an EEG-based recognition of music liking using time-frequency analysis", *IEEE Transactions on Biomedical Engineering* 59(12):3498-3510, 2012.
- [19] S. K. Hadjidimitriou, L. J. Hadjileontiadis, "EEG-based classification of music appraisal responses using time-frequency analysis and familiarity ratings", *IEEE Transactions on Affective Computing* 4(2):161-172, 2013.
- [20] J. Moon, Y. Kim, H. Lee, C. Bae, W. C. Yoon, "Extraction of user preference for video stimuli using EEG-based user responses", *ETRI Journal* 35(6):1105-1114, 2013.
- [21] K. Li, X. Li, Y. Zhang, A. Zhang, "Affective state recognition from EEG with deep belief networks", in: *Bioinformatics and Biomedicine (BIBM)*, 2013 IEEE International Conference on, IEEE, pp. 305-310, 2013.
- [22] W.-L. Zheng, J.-Y. Zhu, Y. Peng, B.-L. Lu, "EEG-based emotion classification using deep belief networks", in: *2014 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, pp. 1-6, 2014.
- [23] S. Jirayucharoensak, S. Pan-Ngum, P. Israsena, "EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation", *The Scientific World Journal*, 2014.
- [24] D. Wu, C. Courtney, B. Lance, S. Narayanan, M. Dawson, K. Oie, and T. Parsons, "Optimal Arousal Identification and Classification for Affective Computing Using Physiological Signals: Virtual Reality Stroop Task", *IEEE Transactions on Affective Computing*, 1(2): 109-118 (2010).
- [25] D. Massari, D. Pacheco, R. Malekshahi, A. Betella, P. Verschure, N. Birbaumer, and A. Caria, "Fast Mental States Decoding in Mixed Reality," *Frontiers in Behavioral Neuroscience*, 8:415, doi:10.3389/fnbeh.2014.00415 (2014).
- [26] N. Kovacevic, P. Ritter, W. Tays, S. Moreno, and A. McIntosh, "My Virtual Dream: Collective Neurofeedback in an Immersive Art Environment", *PLoS ONE*, 10(7): e0130129. doi.org/10.1371/journal.pone.0130129 (2015).
- [27] J. Gielis, "A generic geometric transformation that unifies a wide range of natural and abstract shapes", *American Journal of Botany*, 90:3, pp.333-338, 2003.
- [28] D. Nguyen, B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights", in: *Neural Networks, 1990. 1990 IJCNN International Joint Conference on, IEEE*, pp. 21-26, 1990.
- [29] P.T. De Boer, D.P. Kroese, S. Mannor, and R.Y. Rubinstein, "A tutorial on the cross-entropy method", *Annals of operations research*, 134(1):19-67, 2005.
- [30] P. Bashivan, I. Rish and S. Heisig, "Mental State Recognition via Wearable EEG," *arXiv preprint arXiv: 1602.00985* (2016).
- [31] D. Surangsrirat and A. Intarapanich, "Analysis of the meditation brainwave from consumer EEG device," *SoutheastCon 2015*, pp. 1-6 (2015).
- [32] M. Murugappan and S. Murugappan, "Human Emotion Recognition Through Short Time Electroencephalogram (EEG) Signals Using Fast Fourier Transform (FFT)," *IEEE 9th International Colloquium on Signal Processing and its Applications*, pp. 289-294 (2013).
- [33] S. Gunn, "Support Vector Machines for Classification and Regression," 1998.
- [34] S. Zainudin, S. D. Jasim and A. A. Bakar, "Comparative Analysis of Data Mining Techniques for Malaysian Rainfall Prediction," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, no. 6, pp. 1148-1153, 2016.
- [35] C. Lehmann, "Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG)," *Journal of Neuroscience Methods* 161, pp. 342-350, 2007.
- [36] M. Soleymani, M. Pantic and P. Thierry, "Multi-Model Emotion Recognition in Response to Videos," *IEEE transactions on affective computing*, vol. 3, no. 2, pp. 211-223, 2011.
- [37] V. Nair, G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines", in: *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 807-814, 2010.
- [38] M. D. Zeiler, "Adadelta: an adaptive learning rate method", *arXiv preprint arXiv:1212.5701*.
- [39] J. Teo, L.H. Chew, and J. Mountstephens, "Improving Subject-Independent EEG Preference Classification using Deep Learning Architectures with Dropouts", *Future of Information and Communication Conference (FICC 2018)*, IEEE, April 2018.