Mental Emotional Sentiment Classification with an EEG-based Brain-machine Interface

Conference Paper · April 2019

4 authors, including:

Jordan J. Bird
Nottingham Trent University
88 PUBLICATIONS 2,387 CITATIONS

Aniko Ekart
Aston University
8 PUBLICATIONS 566 CITATIONS

Diego R. Faria
University of Hertfordshire
123 PUBLICATIONS 3,096 CITATIONS

Mental Emotional Sentiment Classification with an EEG-based Brain-Machine Interface

Jordan J. Bird
School of Engineering and Applied Science
Aston University
Birmingham, UK
birdj1@aston.ac.uk

Anikó Ekárt
School of Engineering and Applied Science
Aston University
Birmingham, UK
a.ekart@aston.ac.uk

Christopher D. Buckingham
School of Engineering and Applied Science
Aston University
Birmingham, UK
c.d.buckingham@aston.ac.uk

Diego R. Faria
School of Engineering and Applied Science
Aston University
Birmingham, UK
d.faria@aston.ac.uk

ABSTRACT

This paper explores single and ensemble methods to classify emotional experiences based on EEG brainwave data. A commercial MUSE EEG headband is used with a resolution of four (TP9, AF7, AF8, TP10) electrodes. Positive and negative emotional states are invoked using film clips with an obvious valence, and neutral resting data is also recorded with no stimuli involved, all for one minute per session. Statistical extraction of the alpha, beta, theta, delta and gamma brainwaves is performed to generate a large dataset that is then reduced to smaller datasets by feature selection using scores from OneR, Bayes Network, Information Gain, and Symmetrical Uncertainty. Of the set of 2548 features, a subset of 63 selected by their Information Gain values were found to be best when used with ensemble classifiers such as Random Forest. They attained an overall accuracy of around 97.89%, outperforming the current state of the art by 2.99 percentage points. The best single classifier was a deep neural network with an accuracy of 94.89%.

Keywords

Emotion Classification, Brain-Machine Interface, Machine Learning.

1. INTRODUCTION

The proceedings are the records of the IAPE'18 conference. We ask that authors follow some simple guidelines. In essence, we ask you to make your paper look exactly like this document. The easiest way to do this is simply to replace the content with your own material.

Autonomous non-invasive detection of emotional states is potentially useful in multiple domains such as human robot interaction and mental healthcare. It can provide an extra dimension of interaction between user and device, as well as enabling tangible information to be derived that does not depend on verbal communication [1]. With the increasing availability of low-cost electroencephalography (EEG) devices, brainwave data is becoming affordable for the consumer industry as well as for research, introducing the need for autonomous classification without the requirement of an expert on hand.

Due to the complexity, randomness, and non-stationary aspects of brainwave data, classification is very difficult with a raw EEG stream. For this reason, stationary techniques such as time windowing must be introduced alongside feature extraction of the data within a window. There are many statistics that can be derived from such EEG windows, each of which has varying classification efficacy depending on the goal. Feature selection must be performed to identify useful statistics and reduce the complexity of the model generation process, saving both time and computational resources during the training and classification processes.

The main contributions of this work are as follows:

- Exploration of single and ensemble methods for the classification of emotions.
- A high performing data mining strategy reaching 97.89% accuracy.
- The inclusion of facial EMG signals as part of the classification process.
- A resolution of three emotional classes (positive, neutral, negative) to allow for real world on mental states that are not defined by prominent emotions.
- One Rule classification demonstrating how accurately the AF7 electrode's mean value classifies mental states.

The remainder of this paper will explore related state-of-the-art research and provide the main inspiration and influences for the study. It will explain the methodology of data collection, feature generation, feature selection and prediction methods. The results will be presented and discussed alongside comparable work, followed by conclusions and future work.

2. RELATED WORK

Statistics derived from a time-windowing technique with feature selection have been found to be effective for classifying mental states such as relaxed, neutral, and concentrating [2]. An ensemble method of Random Forest had an observed classification accuracy of 87% when performed with a dataset which was pre-processed with the OneR classifier as a feature selector. These promising results suggested a study on classification of emotional states using a similar exploration method would be similarly successful.

The best current state-of-the-art solution for classification of emotional EEG data from a low-resolution, low-cost EEG setup used Fisher's Discriminant Analysis to produce an accuracy of 95% [3]. The study tried to prevent participants from becoming tense and discourage blinking but the previous study [2] found that EMG data from these activities helped classification because blink rates are a factor in concentration for example. Hence the new study described in this paper will explore classification of emotions in EEG data when unconscious movements are neither encouraged nor discouraged. Conscious extraneous movements such as taking a sip of water will not be allowed because they just form outlying or masking points in the data. For example, if the people experiencing positive emotions are also drinking water, the model will simply classify the electrical data that has been generated by those movements. Stimuli to evoke emotions for EEG-based studies are often found to be best with music [4] and film [5]. This paper thus focuses on film clips that have audio tracks (speech and/or music) to evoke emotions, similarly to a related study that used music videos [6].

Common Spatial Patterns have proved extremely effective for emotion classification, attaining an overall best solution at 93.5% [7]. A MUSE EEG headband was successfully used to classify high resolutions of valence through differing levels of enjoyment during a certain task [8]. Deep Belief Network (DBN), Artificial Neural Network (ANN), and Support Vector Machine (SVM) methods have all been able to classify emotions from EEG data was also found to be very effective with when considering binary classes of positive and negative [9]. This study will build on all these results using similar methods as well as an ensemble, to exploit their differing strengths and weaknesses. The study also supports the usage of a neutral class, for transition into real-world use, to provide a platform for emotional classification where emotions are not prominent. It adds valence or perceived sentiment because this was previously found to be helpful in the learning processes for a web-based chatbot [10].

3. BACKGROUND

3.1 Electroencephalography

Electroencephalography is the process using applied electrodes to derive electrophysiological data and signals produced by the brain [11] [12]. Electrodes can be subdural [13] ie. under the skull, placed on and within the brain itself. Noninvasive techniques require either wet or dry electrodes to be placed around the cranium [14]. Raw electrical data is measured in Microvolts (uV) at observed time t producing wave patterns from *t* to *t+n*.

3.2 Human Emotion

Human emotions are varied and complex but can be generalized into positive and negative categories [15]. Some emotions overlap

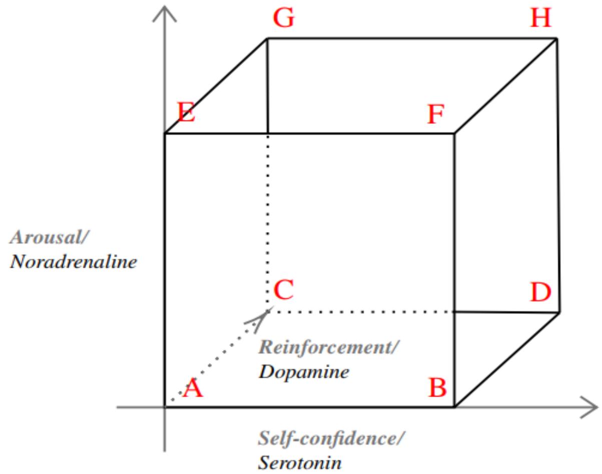

Figure 1. Diagram to show Lövheim's Cube of Emotional Categorization

such as 'hope' and 'anguish', which are considered positive and negative respectively but that are often experienced

Table 1. Table to show Lövheim categories and their encapsulated emotions with a valence label

Emotion Category	Emotion/Valence
A	Shame (Negative) Humiliation (Negative)
B	Contempt (Negative) Disgust (Negative)
C	Fear (Negative) Terror (Negative)
D	Enjoyment (Positive) Joy (Positive)
E	Distress (Negative) Anguish (Negative)
F	Surprise (Negative) (Lack of Dopamine)
G	Anger (Negative) Rage (Negative)
H	Interest (Positive) Excitement (Positive)

contemporaneously: e.g. the clearly doomed hope and accompanying anguish for a character's survival in a film. This study will concentrate on those emotions that do not overlap, to help correctly classify what is and is not a positive experience.

Lövheim's three-dimensional emotional model maps brain chemical composition to generalised states of positive and negative valence [16]. This is shown in Fig. 1 with emotion categories A-H from each of the model's vertices, further detailed in Table I. Various chemical compositions can be mapped to emotions with positive and negative classes. Furthermore, studies show that chemical composition influences nervous oscillation and thus the generation of electrical brainwaves [17]. Since emotions are encoded within chemical composition that directly influence electrical brain activity, this study proposes that they can be classed using statistical features of the produced brainwaves.

3.3 Machine Learning Algorithms

The study in this paper applies a number of machine learning algorithms. One Rule (OneR) classification is a simplistic probabilistic process of selecting one attribute from the dataset and generating logical rules based upon it. For example:

"IF temperature LESS THAN 5.56 THEN December"

and

"IF temperature MORE THAN 23.43 THEN July"

are rules generated based on a temperature attribute to predict the month (class). This model will identify the strongest attribute within the dataset for classifying emotions

Decision Trees follow a linear process of conditional control statements based on attributes, through a tree-like structure where each node is a rule based decision that will further lead to other nodes. Finally, an end node is reached, and a class is given to the data object. The level of randomness or entropy on all end nodes is used to measure the classification ability of the tree. The calculation of entropy is given as:

$$E(S) = -\sum_{i=1}^{c} P_i \times log_2(P_i) \ . \tag{1}$$

Entropic models are compared by their difference in entropy which is information gain. A positive value would be a better model, whereas a negative value shows information loss versus the comparative model. This is given as:

$$GAIN(T,X) = E(T) - E(T,X), \tag{2}$$

where E is the entropy calculated by Equation 1.

Support Vector Machines (SVM) classify data points by generating and optimising a hyperplane to separate them and classifying based on their position in comparison to the hyperplane [18]. A model is considered optimised when the average margins between points and the separator is at its maximum value. Sequential Minimal Optimisation (SMO) is a high-performing algorithm to generate and implement an SVM classifier [19]. The large optimisation problem is broken down into smaller subproblems, that can then be solved linearly.

Bayes' Theorem [20] uses conditional probabilities to determine the likelihood of Class A based on Evidence, B, as follows:

$$P(A|B) = \frac{P(\mathrm{B}|\mathrm{A})\mathrm{P}(\mathrm{A})}{P(B)}. \tag{3}$$

For this study, evidence consists of attribute values (EEG time-window statistics) and ground-truth training for determining their most likely classes. A simpler version is known as Naive Bayes, which assumes independence of attribute values whether or not they are really unrelated. Classification of Naive Bayes is adapted from Equation 3 as follows:

$$\hat{y} = k \in (1, \ldots, k)\, p(C_k) \prod_{i=1}^{n} p(x_i|C_k)\,, \tag{4}$$

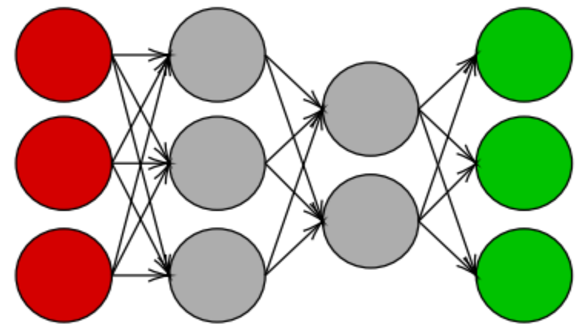

Figure 2. A simplified diagram of a fully-connected feed forward deep neural network.

where y is the class and k is the data object (row) that is being classified.

Logistic Regression is a symmetric statistical model used for mapping a numerical value to a probability, ie. hours of study to predict a student's exam grade [21]. For a binary classification problem with i attributes, and β model parameters, the log odds l is given as $l = \beta_0 + \sum_{i=0}^{x} \beta_i + x_i$ and thus the corresponding odds of outcome are therefore given as $o = b^{\beta_0 + \sum_{i=0}^{x} \beta_i + x_i}$ which can be used to predict a model outcome based on previous data.

A Multilayer Perceptron is a type of Artificial Neural Network (ANN) that predicts a class by taking input parameters and computing them through a series of hidden layers to one or more nodes on the final output layer. More than one hidden layer forms a *deep neural network* and output layers can be different classes or, if there is just one, a regression output. A simplified diagram of a fully connected feed forward deep neural network can be seen in Fig. 2. Learning is performed for a defined time and follows the process of backpropagation [22], which is the process of deriving a gradient that is further used to calculate weights for each node (neuron) in the network. Training is based on reducing the error rate given by the error function ie. the performance of a network in terms of correct and incorrect classifications or total Euclidean distance from the real numerical values. An error is calculated at output and fed backwards from outputs to inputs.

3.4 Model Ensemble Methods

An ensemble combines two or more prediction models into a single process. A method of fusion takes place to increase the success rate of a prediction process by treating the models as a sum of their parts.

Voting is a simple ensemble process of combining models and allowing them to vote through a democratic or elitist process. Each of the models are trained, and then for prediction, they award vote *v* to class(es) via a specified method:

- Average of probabilities; v = confidence
- Majority vote; v = 1
- Min/Max probability v = average confidence of all models

Following the selected process, a democracy will produce an outcome prediction as that of the class that has received the strongest vote or set of votes.

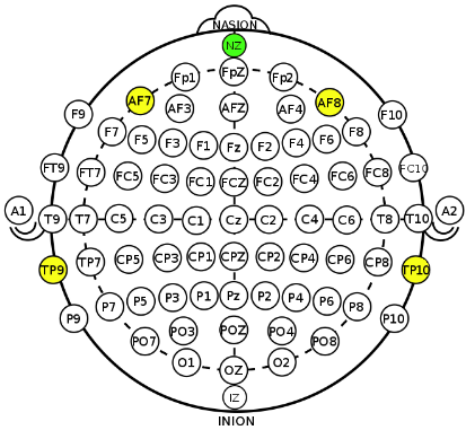

Figure 3. EEG sensors TP9, AF7, AF8 and TP10 of the Muse headband on the international standard EEG placement system [26]

Random Forest forms a voting ensemble from Decision Trees [23]. Multiple trees are generated on randomly generated subsets of the input data (Bootstrap Aggregation) and then those trees, the random forest, will all vote on their predicted outcome and a prediction is derived. Adaptive Boosting is the process of creating multiple unique instances of one type of model prediction to effectively improve the model in situations where selected parameters may prove ineffective [24]. Classification predictions are combined and weighted after a process of using a random data subset to improve on a previous iteration of a model. Combination is given as:

$$F_T(x) = \sum_{t=1}^{T} f_t(x)\,, \tag{5}$$

where F is the set of t models and x is the data object with an unknown class [25].

4. METHOD

The study employs four dry extra-cranial electrodes via a commercially available MUSE EEG headband. Microvoltage measurements are recorded from the TP9, AF7, AF8, and TP10 electrodes, as seen in figure 3. Sixty seconds of data were recorded from two subjects *(1 male, 1 female, aged 20-22)* for each of the 6 film clips found in Table II producing 12 minutes (720 seconds) of brain activity data (6 minutes for each emotional state). Six minutes of neutral brainwave data were also collected resulting in a grand total of 36 minutes of EEG data recorded from subjects. With a variable frequency resampled to 150Hz, this resulted in a dataset of 324,000 data points collected from the waves produced by the brain. Activities were exclusively stimuli that would evoke emotional responses from the set of emotions found in Table I and were considered by their valence labels of *positive* and *negative* rather than the emotions themselves. Neutral data were also collected, without stimuli and before any of the emotions data (to avoid contamination by the latter), for a third class that would be the resting emotional state of the subject. Three minutes of data were collected per day to reduce the interference of a resting emotional state.

Table 2. Source of Film Clips used as Stimuli for EEG Brainwave Data Collection

Stimulus	Valence	Studio	Year
Marley and Me	Neg	Twentieth Century Fox, etc.	2008
Up	Neg	Walt Disney Pictures, etc.	2009
My Girl	Neg	Imagine Entertainment, etc.	1991
La La Land	Pos	Summit Entertainment, etc.	2016
Slow Life	Pos	BioQuest Studios	2014
Funny Dogs	Pos	MashupZone	2015

Table 3. Attribute Evaluation Methods used to Generate Datasets for Model Training

Evaluator	Ranker Cutoff	No. Attributes
OneR	0.4	52
BayesNet	0.4	67
InfoGain	0.75	63
Symmetrical Uncertainty	0.4	72

Participants were asked to watch the film without making any conscious movements (eg. drinking coffee) to prevent the influence of Electromyographic (EMG) signals on the data due to their prominence over brainwaves in terms of signal strength. A previous study that suggested blinking patterns are useful for classifying mental states [2] d blinking patterns are useful for classifying mental states [2] inspired this study to neither encourage nor discourage unconscious movements. Observations of the experiment showed a participant smile for a short few seconds during the 'funny dogs' compilation clip, as well as become visibly upset during the 'Marley and Me' film clip (death scene). These facial expressions will influence the recorded data but are factored into the classification model because they accurately reflect behaviour in the real world, where these emotional responses would also occur. Hence, to accurately model realistic situations, both EEG and facial EMG signals are considered as informative. To generate a dataset of statistical features, an effective methodology from a previous study [2] was used to extract 2400 features through a sliding window of 1 second beginning at t=0 and t=0.5. Downsampling was set to the minimum observed frequency of 150Hz.

Feature selection algorithms were run to generate a reduced dataset from the 2,549 source attributes. Chosen methods ranked attributes based on their effectiveness when used in classification, and a manual cutoff point was tuned where the score began to drop off, therefore retaining only the strongest attributes. Details of attribute numbers generated by each method can be seen in Table III. The reduced dimensionality makes the classification the classification experiments more tractable and within the remit of given computational resources.

Table 4. Classification Accuracy of Single and Ensemble Methods on the Four Generated Datasets

Dataset	Single Model Accuracy							Ensemble Model Accuracy		
	OneR	***RT***	***SMO***	***NB***	***BN***	***LR***	***MLP***	***RF***	***Vote***	***AB(RF)***
OneR	85.18	91.18	89.49	66.56	91.18	91.84	<u>92.07</u>	95.26	92.68	<u>95.59</u>
BayesNet	85.27	93.05	89.49	60.69	91.23	91.93	<u>93.81</u>	97.14	93.39	<u>97.23</u>
InfoGain	85.27	94.18	89.82	60.98	91.46	92.35	**<u>94.89</u>**	**<u>97.89</u>**	94.04	97.84
Symmetrical Uncertainty	85.27	<u>94.15</u>	89.54	69.66	92.03	91.93	94.18	97.56	94.32	<u>97.65</u>

Normalised mean value of the AF7 electrode:

< -460.0 -> ***NEGATIVE***

< -436.5 -> ***POSITIVE***

< -101.5 -> ***NEGATIVE***

< 25.45 -> ***POSITIVE***

< 25.85 -> ***NEUTRAL***

< 26.25 -> ***POSITIVE***

< 37.7 -> ***NEUTRAL***

< 39.05 -> ***POSITIVE***

< 43.599999999999994 -> ***NEUTRAL***

< 63.95 -> ***POSITIVE***

< 97.7 -> ***NEUTRAL***

< 423.0 -> ***POSITIVE***

>= 423.0 -> ***NEGATIVE***

Figure 4. The most effective single rule for classification.

5. PRELIMINARY RESULTS

Model training for each method was performed on every dataset generated by the four methods shown in Table III. The parameters, where required, were set to the following:

- 10-fold cross validation for training models (average of 10 models on 10 folds of data).
- A manually tuned deep neural network of two layers, 30 and 20 neurons on each layer respectively. Backward propagation of errors. 500 epoch training time.
- All random numbers generated by the Java Virtual Machine with a seed of 0.
- Ensemble voting based on Average Probabilities.

After downsampling, there were slightly more datapoints for the neutral state, and thus to benchmark a Zero Rules ('most common class') classifier would classify all points as neutral. This was 33.58% and therefore any result above this shows useful rule generation.

Models for ensemble were selected manually based on best performance. Voting was performed on average probabilities using the Random Tree, SMO, BayesNet, Logistic Regression, and MLP models. Random Forests, due to their impressive

Table 5. An Indirect Comparison of this Study to Similar Works Performed on Different Datasets

Study	Method	Accuracy
This study	InfoGain, RandomForest	**97.89**
Bos, et al. [3]	Fisher's Discriminant	94.9
This study	InfoGain, MLP	94.89
Li, et al. [7]	Common Spatial Patterns	93.5
Li, et al.	Linear SVM	93
Zheng, et al. [9]	Deep Belief Network	87.62
Koelstra, et al. [6]	Common Spatial Patterns	58.8

classification ability was attempted to be optimized by the AdaBoost Algorithm.

Results of both single and ensemble classifiers can be seen in Table IV. The best model, a Random Forest with the Infogain dataset, achieved a high accuracy of **97.89%**. The small amount of classification errors came from a short few seconds of the half an hour dataset, meaning that errors could be almost completely mitigated when classifying in real time due to the sliding window technique used for small timeframes *t-n*. Adaptive boosting was promising for all Random Forest models but could not achieve a score higher, pointing towards the possibility of outlying points. For single classification, the multilayer perceptron was the most consistently best model, showing the effectiveness of neural networks for this particular problem.

The effectiveness of OneR classification showed that a certain best attribute (*mean value of AF7*) existed that alone had a classification ability of 85.27%. The rule is specified in Fig. 4. The normalised mean value of the time windows extracted from the AF7 electrode when observed show that minimum and maximum values most commonly map to negative emotions, whereas positive and neutral are very closely related, having rules overlapping one another. One Rule classification improved over the Zero Rule benchmark by over 50 points, and therefore would have been an effective attribute to consider over others when it came to utilising more than one of the attributes in the other methods.

The two best models in our study are compared to the state of the art alternatives in Table V. The method of generating attributes, attribute selection via info gain and finally classification with a

Random Forest outperforms an FDA model by 2.99 points. Further work should be carried out to identify whether this improved result was due to the methods chosen or the attribute generation and selection, or possibly both.

6. DISCUSSION

The high performance of simple multilayer perceptrons suggests neural network models can be effective, especially more complex ones such as Convolutional Neural Networks (CNNs) that have performed well in various classification experiments [27]. Similarly, ensemble and and Bayesian models are promising avenues that could perform better with more advanced models, such as Dynamic Bayesian Mixture Models (DBMM) [28] that have previously been applied to statistical data extracted from EEG brainwave signals.

Being able to recognise emotions autonomously would be valuable for mental-health decision support systems such as GRiST which is a risk and safety management system used by mental-health practitioners and by people for assessing themselves [29], [30]. Evaluations of emotions independent of self-reporting would help calibrate the advice as well as guiding more sensitive interactions. The measurement of brainwaves used in this paper is too intrusive but would be useful for providing a benchmark for finding other more appropriate methods.

7. CONCLUSION

This paper explored the application of single and ensemble methods of classification to take windowed data from four points on the scalp and quantify that data into an emotional representation of what the participant was feeling at that time. The methods showed that using a low resolution, commercially available EEG headband can be effective for classifying a participant's emotional state. There is considerable potential for producing classification algorithms that have practical value for real-world decision support systems. Responding to emotional states can improve interaction and, for mental-health systems, contribute to the overall assessment of issues and how to resolve them.

ACKNOWLEDGEMENT

This work was partially supported by the European Commission through the H2020 project EXCELL (https://www.excell-project.eu/), grant number 691829 (A. Ekart) and by the EIT Health GRaCEAGE grant number 18429 awarded to C. D. Buckingham.

REFERENCES

[1] M. S. El-Nasr, J. Yen, and T. R. Ioerger, "Flame - fuzzy logic adaptivemodel of emotions," Autonomous Agents and Multi-agent systems, vol. 3, no. 3, pp. 219–257, 2000.

[2] Ding, W. and Marchionini, G. 1997. *A Study on Video Browsing Strategies*. Technical Report. University of Maryland at College Park.

[3] J. J. Bird, L. J. Manso, E. P. Ribiero, A. Ekart, and D. R. Faria, "A study on mental state classification using eeg-based brain-machine interface," in 9th International Conference on Intelligent Systems, IEEE, 2018.

[4] D. O. Bos et al., "EEG-based emotion recognition," The Influence of Visual and Auditory Stimuli, pp. 1–17, 2006.

[5] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "EEG-based emotion recognition in music listening," IEEE Transactions on Biomedical Engineering, vol. 57, no. 7, pp. 1798–1806, 2010.

[6] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from eeg data using machine learning approach," Neurocomputing, vol. 129, pp. 94–106, 2014.

[7] S. Koelstra, A. Yazdani, M. Soleymani, C. Mühl, J.-S. Lee, A. Nijholt, T. Pun, T. Ebrahimi, and I. Patras, "Single trial classification of eeg and peripheral physiological signals for recognition of emotions induced by music videos," in Int. Conf. on Brain Informatics, pp. 89– 100, Springer, 2010.

[8] M. Li and B.-L. Lu, "Emotion classification based on gamma-band eeg," in Engineering in medicine and biology society, 2009. EMBC 2009. Annual international conference of the IEEE, pp. 1223–1226, IEEE, 2009.

[9] M. Abujelala, C. Abellanoza, A. Sharma, and F. Makedon, "Brainee: Brain enjoyment evaluation using commercial eeg headband," in Proceedings of the 9th acm international conference on pervasive technologies related to assistive environments, p. 33, ACM, 2016.

[10] W.-L. Zheng, J.-Y. Zhu, Y. Peng, and B.-L. Lu, "Eeg-based emotion classification using deep belief networks," in Multimedia and Expo (ICME), 2014 IEEE International Conference on, pp. 1–6, IEEE, 2014.

[11] J. J. Bird, A. Ekárt, and D. R. Faria, "Learning from interaction: An intelligent networked-based human-bot and bot-bot chatbot system," in UK Workshop on Computational Intelligence, pp. 179–190, Springer, 2018.

[12] B. E. Swartz, "The advantages of digital over analog recording techniques," Electroencephalography and clinical neurophysiology, vol. 106, no. 2, pp. 113–117, 1998.

[13] A. Coenen, E. Fine, and O. Zayachkivska, "Adolf beck: A forgotten pioneer in electroencephalography," Journal of the History of the Neurosciences, vol. 23, pp. 276–286, 2014.

[14] A. K. Shah and S. Mittal, "Invasive electroencephalography monitoring: Indications and presurgical planning," Annals of Indian Academy of Neurology, vol. 17, pp. S89, 2014.

[15] B. A. Taheri, R. T. Knight, and R. L. Smith, "A dry electrode for eeg recording," Electroencephalography and clinical neurophysiology, vol. 90, no. 5, pp. 376–383, 1994.

[16] K. Oatley and J. M. Jenkins, Understanding emotions. Blackwell publishing, 1996.

[17] H. Lövheim, "A new three-dimensional model for emotions and monoamine neurotransmitters," Medical hypotheses, vol. 78, no. 2, pp. 341–348, 2012

[18] J. Gruzelier, "A theory of alpha/theta neurofeedback, creative performance enhancement, long distance functional connectivity and psychological integration," Cognitive processing, vol. 10, no. 1, pp. 101–109, 2009.

[19] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, no. 3, pp. 273–297, 1995.

[20] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," 1998.

[21] T. Bayes, R. Price, and J. Canton, “An essay towards solving a problem in the doctrine of chances,” 1763.

[22] S. H. Walker and D. B. Duncan, “Estimation of the probability of an event as a function of several independent variables,” Biometrika, vol. 54, no. 1-2, pp. 167–179, 1967.

[23] Y. Bengio, I. J. Goodfellow, and A. Courville, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[24] T. K. Ho, “Random decision forests,” in Document analysis and recognition, 1995., proceedings of the third international conference on, vol. 1, pp. 278–282, IEEE, 1995.

[25] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” Journal of computer and system sciences, vol. 55, no. 1, pp. 119–139, 1997.

[26] R. Rojas, “Adaboost and the super bowl of classifiers a tutorial introduction to adaptive boosting,” Freie University, Berlin, Tech. Rep, 2009

[27] H. H. Jasper, “The ten-twenty electrode system of the international federation,” Electroencephalogr. Clin. Neurophysiol., vol. 10, pp. 370– 375, 1958.

[28] M. Hussain, J. J. Bird, and D. R. Faria, “A study on cnn transfer learning for image classification,” in UK Workshop on Computational Intelligence, pp. 191–202, Springer, 2018.

[29] D. R. Faria, M. Vieira, C. Premebida, and U. Nunes, “Probabilistic human daily activity recognition towards robot-assisted living,” in Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on, pp. 582–587, IEEE, 2015.

[30] C. D. Buckingham, A. Ahmed, and A. Adams, “Designing multiple user perspectives and functionality for clinical decision support systems,” pp. 211–218, 2013.

[31] C. D. Buckingham, A. Adams, L. Vail, A. Kumar, A. Ahmed, A. Whelan, and E. Karasouli, “Integrating service user and practitioner expertise within a web-based system for collaborative mental-health risk and safety management,” Patient Education and Counseling, pp. 1189-1196, 2015.