2018 International Conference on Intelligent Systems (IS)

A Study on Mental State Classification using EEG-based Brain-Machine Interface

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

Luis J. Manso
School of Engineering & Applied Science
Aston University
Birmingham, UK
l.manso@aston.ac.uk

Eduardo P. Ribeiro
Department of Electrical Engineering
Federal University of Parana
Curitiba, Brazil
edu@eletrica.ufpr.br

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

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

***Abstract*—This work aims to find discriminative EEG-based features and appropriate classification methods that can categorise brainwave patterns based on their level of activity or frequency for mental state recognition useful for human-machine interaction. By using the Muse headband with four EEG sensors (TP9, AF7, AF8, TP10), we categorised three possible states such as relaxing, neutral and concentrating based on a few states of mind defined by cognitive behavioural studies. We have created a dataset with five individuals and sessions lasting one minute for each class of mental state in order to train and test different methods. Given the proposed set of features extracted from the EEG headband five signals (alpha, beta, theta, delta, gamma), we have tested a combination of different features selection algorithms and classifier models to compare their performance in terms of recognition accuracy and number of features needed. Different tests such as 10-fold cross validation were performed. Results show that only 44 features from a set of over 2100 features are necessary when used with classical classifiers such as Bayesian Networks, Support Vector Machines and Random Forests, attaining an overall accuracy over 87%.**

Keywords — EEG, brain-machine interface, machine learning, mental states classification

I. Introduction

The ability to autonomously detect mental states, whether cognitive or affective, is useful for multiple purposes in many domains such as robotics, health care, education, neuroscience, etc. The importance of efficient human-machine interaction mechanisms increases with the number of real life scenarios where smart devices, including autonomous robots, can be applied. One of the many alternatives that can be used to interact with machines is through superficial brain activity signals. These signals, called electroencephalograms or EEG for short, convey information regarding the voltage measured by electrodes (dry or wet) placed around the scalp of an individual. In addition to regular non-invasive electroencephalography there can also be found invasive alternatives which can monitor brain activity placing the electrodes directly inside the skull of the subject [35]. This technique is known as intracranial electroencephalography (iEEG). Despite iEEG can yield better signal acquisition, it is invasive and therefore more complex to apply. Extracranial electroencephalography techniques include wearable and non-wearable technologies. The fact that extracranial devices used to acquire EEG signals are non-invasive, are becoming easier to wear, and their price is decreasing widens the range of applications for which they are suitable.

A major challenge in brain-machine interface applications is inferring how momentary mental states are mapped into a particular pattern of brain activity. One of the main issues of classifying EEG signals is the amount of data needed to properly describe the different states, since the signals are complex, non-linear, non-stationary, and random in nature. The signals are considered stationary only within short intervals, that is why the best practice is to apply short-time windowing technique in order to detect local discriminative features to meet this requirement. The paper at hand focuses on selecting a subset of highly discriminative features and comparing to state-of-the-art classification methods that can categorise EEG signals into different mental states, taking into consideration the performance in terms of accuracy and computational cost. The application considered herein is to distinguish among three different mental states (e.g. relaxed, neutral and highly concentrated) of an individual using an EEG device with dry electrodes that can interface a range of applications, such as to control the movement of a robot.

The remainder of the paper proceeds as follows. Related works are summarised in section II. The experimental setup, including information regarding the device used, and details about the data acquisition are described in section III. The methods tested to perform feature selection and the criteria used to compare the different classifiers are presented in section IV. Preliminary results are presented in section V. A discussion on the conclusions drawn from the experimental results is provided in section VI.

II. Related Work

Statistical features derived from EEG data are commonly used alongside machine learning techniques to classify mental states [18], [19]. These nominal states can then be used for

978-1-5386-7097-2/18/$31.00 ©2018 IEEE

finite points of control as a Brain-Computer Interface. A *Muse* headband has been recognised by neuroscientists for its effectiveness and relatively low cost as well as its accuracy when classified with Bayesian methods [8]. Through signals, two tasks were recognised with 95% accuracy, though it is worth noting that tasks were classified rather than mental states, and said tasks were in binary distinction to one another. Using a *Muse* headband, researchers accurately measured a user's enjoyment [11], [12] of an activity from brain signals alone using the stimuli of two videogames, one measurably more enjoyable than the other. With the use of a high resolution 32-channel EEG and statistical feature extraction, a model was developed to control a robot's movement [9]. Using statistics focused on the signals produced by the motor cortex which is thought to control muscles for movement [10], researchers classified various states which successfully resulted in a model that could direct a robot's movement. EEG data has been used extensively to detect abnormal brain activity related to ill-health such as stroke [13] specifically when ischemia is present in the brain, brain activity points to abnormalities prior to the stroke occurring. As well as stroke detection, neuroscientists found that upper extremities in motor function post-stroke could be rehabilitated using EEG data with robotics feedback [14] in the form of a brain-machine interface. Results were promising in terms of the effectiveness of the system's ability to rehabilitate. Also studied extensively is the ability to use EEG data to detect seizures both in adults suffering with epilepsy [15] and notably in new-born infants [16]. A Spiking Neural Network was developed to classify seizure detection based on statistics extracted from EEG streams with a high accuracy of 92.5% [17]. Random Forest classification of extracted EEG features was used to identify mental states during stages of sleep with a high accuracy of 82% [20], a Bayesian classifier was trained on more general awake, sleep and REM sleep states with accuracies ranging between 92-97% in both humans and rats [21]. Neural Networks have been observed to have an accuracy of 64% when classifying emotional states based on EEG data [7].

Differently from the aforementioned works, this work focuses on a study on features selection and classification models given a set of proposed features such as statistical, entropy-based, derivatives and time-frequency features from short temporal lapses of EEG data to then generate multiple data sets of the same data points with original contribution in their differing selections of attributes, which in turn are selected by various machine learning models. The primary goal is to find a suitable model that can categorise mental states based on EEG data from the TP9, AF7, AF8 and TP10 electrodes.

III. Experimental Setup and Dataset

A. EEG Data Acquisition

The sensor *Muse Headband* was used for data collection. The *Muse* is a commercial EEG sensing device with five dry-application sensors, one used as a reference point (NZ) and four (TP9, AF7, AF8, TP10) to record brain wave activity.

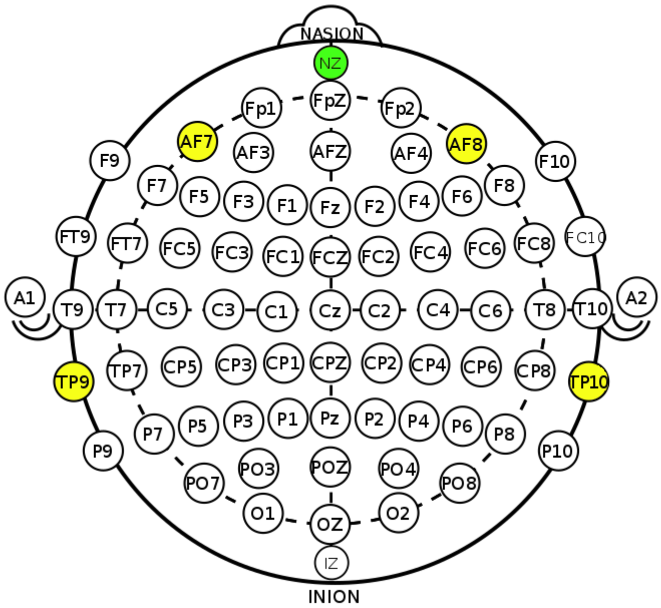

Fig. 1. The International 10-20 EEG Electrode Placement Standard [4] Highlighted in yellow are the sensors of the *Muse Headband*. The NZ placement (green) is used as a reference point for calibration.

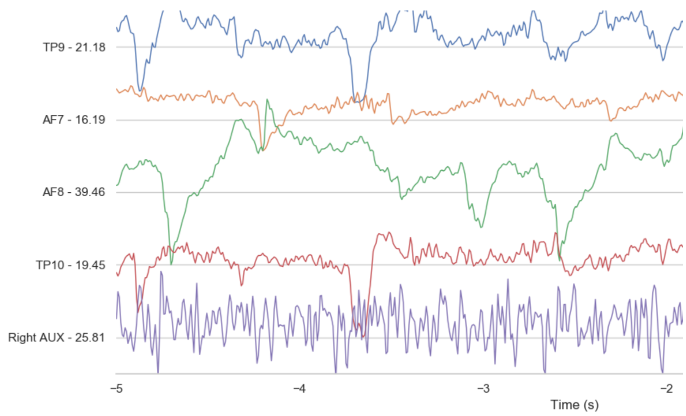

Fig. 2. Example of a live EEG stream of the four *Muse* sensors, Right AUX did not have a device and was discarded due to it simply being noise. This live feed graph has a Y-Axis of measured microvolts at t=0 on each sensor, and an X-axis detailing the time reading.

To prevent the interference of electromyographic signals, nonverbal tasks that required little to no movement were set. Blinking, though providing interference to the AF7 and AF8 sensors, was neither encouraged nor discouraged to retain a natural state. This was due to the dynamicity of blink rate being linked to tasks requiring differing levels of concentration [1], and as such the classification algorithms would take these patterns of signal spikes into account. In addition, subjects were asked not to close their eyes during any of the tasks. Three stimuli were devised to cover the three mental states available from the *Muse Headband* - *relaxed*, *neutral*, and *concentrating*. The relaxed task had the subjects listening to low-tempo music and sound effects designed to aid in meditation whilst being instructed on relaxing their muscles and resting. For a neutral mental, a similar test was carried out, but with no stimulus at all, this test was carried out prior to any others to prevent lasting effects of a relaxed or concentrative mental state. Finally, for concentration, the subjects were instructed to follow the “shell game” in which a ball was hidden under one of three cups, which were then switched, the task was to try and follow which cup hid the ball. Future work arises in the implementation of a standard experiment for each state, for proper comparison to similar experiment. After a short amount of time into the stimulus starting, as to not gather data with an inaccurate class, the EEG data from the *Muse Headband* was automatically recorded for sixty seconds. The data was observed to be streaming at a variable frequency within the range of 150 - 270 Hz. *BlueMuse* [5] was used for interfacing the device to a computer, and *Muselsl* [6] was used to convert the Muse signals to *MicroVolts* and record the data into a preliminary dataset ready for feature extraction. Fig 2. shows a live stream of EEG data, blinking can be seen in the troughs of TP9 and TP10 (forehead sensors). At each point in the data stream (150 - 270 Hz), all signals were recorded along with a UNIX timestamp which was further used for down sampling the data to produce a uniform stream frequency. The measured voltages on the graph can be mapped to the EEG placements seen in Fig 1. Before the features extraction we have down sampled the data. The sampling rate was decimated to 200 Hz based on fast Fourier transformations along a given axis. The resampled signal starts at the same value as *x*, but it is sampled with a spacing of len(*x*) / *num* * (*spacing* of *x*). Because a Fourier method is used, the signal is assumed to be periodic. This is a realistic down-sampling as the dominant energy is concentrated in the range of 20 - 500Hz, even though the frequency range of the EEG sensor is superior.

IV. METHODS

A. Proposed Set of Features for EEG signals

Feature extraction and classification of EEG signals are core issues in brain computer interface (BCI) applications. One challenging problem when it comes to EEG feature extraction is the complexity of the signal, since it is non-linear, non-stationary, and random in nature. The signals are considered stationary only within short intervals, that is why the best practice is to apply short-time windowing technique to meet this requirement. However, it is still considered an assumption that holds during a normal brain condition. Non-stationary signals can be observed during the change in alertness and wakefulness, during eye blinking, and also during transitions of mental states. Thus, this subsection describes the set of features considered in this work to adequately discriminate different classes of mental states. These features rely on statistical techniques, time-frequency based on fast Fourier transform (FFT), Shannon entropy, max-min features in temporal sequences, log-covariance and others. All features proposed to classify the mental states are computed in terms of the temporal distribution of the signal in a given time window. This slide window is defined as a period of 1 second at 250 Hz, i.e. all features are computed within this time instant. An overlap of 0.5 second is used when moving the window, i.e. the temporal window 1 (w1) starts at 0 sec. and finishes at 1 sec.; w2 starts at 1.5 sec. and finishes at 2.5 sec.; w3 starts at 2 sec. and finishes at 3 sec.; w4 starts at 2.5 sec. and finishes at 3.5 sec., and so on. Another important point to compute the features is the signals from the EEG Muse headband. Since it returns five types of signal frequencies $\{\alpha, \beta, \theta, \delta, \gamma\}$, then we compute all proposed set of features for each signal. Thus, the total number of feature values extracted from these signals is 2147 values.

Statistical Features: In order to have a compact representation of the raw sensor data in a given time range, we are using a set of classical statistical features since they are useful with proven efficiency to complement set of multiples features in order to recognise patterns in time series. The statistical features are: (i) given a set of data values $\{x_1, x_2, ...x_N\}$ acquired in each temporal window, the mean value $\mu = \frac{1}{N}\sum_i^N x_i$ of that sequence is computed; (ii) the standard deviation $\sigma = \sqrt{\frac{1}{N}\sum_i^N (x_i - \mu)^2}$; (iii) statistical moments of 3rd and 4th order, which gives us the skewness to measure the asymmetry of the data, and also the kurtosis to measure the peakedness of the probability distribution of the data, respectively. The statistical moments employed are computed as follows:

$$y = \frac{\mu^k}{\sigma^k}, \tag{1}$$

$$\mu^k = \frac{1}{N}\sum_i^N (x_i - \mu)^k \,, \tag{2}$$

where μ^k is the k = {3rd, 4th} moment about the mean and y = {skewness, k = 3; kurtosis, k = 4}. Another type of statistical features computed was the autocorrelation of the signals at each time window for each of the five signals from the EEG. The correlation of the signal with a delayed copy of itself as a function of delay was employed similarly to [22] and [23], where the implementation details and parameters are described.

Max, Min and Derivatives: Given a time window of 1 sec., the maximum and minimum values are computed to increase the diversity of the features types. Derivatives are also computed as temporal features. For each time window, we split the time window by 2, such *w*/2 = 0.5 sec. and *w* = *1 sec.*, resulting in two sequences of data at ~125 Hz, then we compute:

$$uv = \frac{\mu^w - \mu^{w/2}}{2}, \tag{3}$$

where *w* and *w*/2 indicates the first and second half of the sequence of data in a time window of 1 sec. The same strategy is employed to get the derivative given the *max* and *min* features in sub time windows:

$$max_t = \frac{max^w - max^{w/2}}{2}, \tag{4}$$

$$min_t = \frac{min^w - min^{w/2}}{2}. \tag{5}$$

The next temporal features are extracted after splitting the initial time window of one second into 4 batches of 0.25 sec. each. Then we computed the *mean*, *max* and *min* values of each batch, {μ_1, μ_2, μ_3, μ_4}, {max_1, max_2, max_3, max_4} and {min_1, min_2, min_3, min_4}. Then we compute the 1D Euclidean distance among all mean values, $\delta_{\mu 12} = |\mu_1 - \mu_2|$, $\delta_{\mu 13} = |\mu_1 - \mu_3|$, $\delta_{\mu 14} = |\mu_1 - \mu_4|$, $\delta_{\mu 23} = |\mu_2 - \mu_3|$, $\delta_{\mu 24} = |\mu_2 - \mu_4|$, $\delta_{\mu 34} = |\mu_3 - \mu_4|$, the same for the minimum and maximum values, so that in the end we got 18 features based on distances. Using the four *mean* values, and the four *max* and four *min* values, and adding the previous 18, we got 30 features for each signal in the short time window, so that counting the 5 signals we have 150 temporal features per second.

Log-covariance features: Given the previous 150 temporal features, we then discard the last 6 features in order to attain 144 features, so that we could build a 12×12 square matrix to compute the log-covariance as follows:

$$lcM = U(logm(cov(M))), \tag{6}$$

where *lcM* is a resulting vector containing the upper triangular elements (78 features) of the matrix after computing the matrix logarithm over the covariance matrix *M*; U(.) is a function to return the upper triangular elements; logm(.) is the matrix logarithm function; and the covariance matrix is given by $cov(M) = cov_{ij} = 1/N \sum_k^N (x_{ik} - \mu_i)(x_{kj} - \mu_j)$. The rationale behind of log-covariance is the mapping of the convex cone of a covariance matrix to the vector space by using the matrix logarithm so that it does not lie in Euclidean space, i.e., the covariance matrix space is not closed under multiplication with negative scalars.

Shannon entropy and log-energy entropy: non-linear analysis such as Shannon entropy has proven its efficiency in signal processing and time series since randomness of non-linear data is well embodied by calculating entropies over the time series. Entropy is an uncertainty measure and in brain-machine interface applications, it is used to measure the level of chaos of the system, since it is a non-linear measure quantifying the degree of complexity of the data. In information theory, the Shannon entropy is given by:

$$h = -\sum_j S_j \times log(S_j), \tag{7}$$

where *h* is a feature computed in every time window of 1 sec. and S_j is each element (normalized) of this temporal window. Then, given the same time window, we split into two to compute the log-energy entropy as follows:

$$loge = \sum_i log(S_i^2) + \sum_j log(S_j^2), \tag{8}$$

where *i* represents an index for the elements of the first sub window (0 - 0.5 sec.) and *j* represents an index for the second sub window (0.5 - 1 sec.).

Frequency domain: The FFT is an advantageous method to analyse the spectrum of a given time-series. At every time window we compute it as follows:

$$X_k = \sum_{n=0}^{N-1} S_n^t \, e^{-i2\pi k \frac{n}{N}}, \; k = 0, \ldots, N-1. \tag{9}$$

Accumulative features as energy model: An accumulative value was obtained frame-by-frame given a time window, for each individual feature, duplicating the number of features. We compute the difference between the values of the current frame to the previous frame and accumulate it over time as follows:

$$y_{icum}^t = \begin{cases} y_i^t, & t = 1. \\ \sum_{z=2}^t (y_i^z - y_i^{z-1})^2, & t > 1 \end{cases}, \tag{10}$$

where y_{icum}^t is the resulting energy model for the current time instant given a specific type of feature y_i^k, i = {1, ... , N} at a time instant *z* representing a specific frame within a time window.

B. Feature Selection Algorithms

Feature selection aims to remove data which has no useful application and only serves to unneededly increase the demand for resources. Five datasets were generated using different algorithms. Each retained the same data points, but which had a reduced number of attributes selected by the algorithm. The evaluators used were as follows:

1. *OneR*: calculates error rate of each prediction based on one rule and selects the lowest risk classification [24].
2. *Information Gain*: assigns a worth to each individual attribute by measuring the information gain with respect to the class (difference of entropy) [25].
3. *Correlation*: measures the correlation of the attribute and class via their *Pearson's* coefficient which is used to rank attributes' worth comparable to all others. [26].
4. *Symmetrical Uncertainty*: measures the uncertainty of an attribute with respect to the class and bases selection on lower uncertainties [27].
5. *Evolutionary Algorithm*: creates a population of attribute subsets and ranks their effectiveness with a fitness function to measure their predictive ability of the class. At each generation, solutions are bred to create offspring, and weakest solutions are killed off in a tournament of fitness [34].

C. Machine Learning Algorithms

As a benchmark, a *ZeroR* classifier was first run on each dataset. This simplistic classifier chooses one single class to apply to all of the data to reduce inaccurate classifications, it is expected that an accuracy of one third is achieved with a fair distribution of the three mental states. Two models were trained on Bayes Theorem, a formula of conditional probability based on hypothesis *H* and evidence *E*. The theorem states that the probability of the hypothesis being true before evidence *P(H)* is related to the probability of the hypothesis after reading the evidence *P(H | E)* and is given as follows [29]:

$$P(H \mid E) = \frac{P(E \mid H)\, P(H)}{\sum_j P(E|H)P(H)}. \tag{11}$$

Naivety arises due to the unverified assumption of non-consideration of the relationships between the absence of attributes. A Bayesian Network (*Bayes Net*) model was also trained. This method generates a probabilistic graphical model via representing probabilities of variables to classes on a Directed Acyclic Graph (DAG) [28] as follows:

$$P(C^{t-1:t-T} \mid X^{t:t-T}) = \frac{1}{\beta} \prod_{k=t}^{T-t} P(X^k|C^k)P(C^k). \tag{12}$$

TABLE I. TABLE TO SHOW ACCURACY OF TRAINED MODELS

Dataset	**Model Accuracy % (2dp)**						
	Naive Bayes	***Bayes Net***	***J48***	***Random Tree***	***Random Forest***	***MLP***	***SVM***
OneR	56.21	73.67	80	76.21	**87.16**	74.27	61.18
Information Gain	54.2	71.64	76.85	65.02	78.02	72.22	64.1
Correlation	56.3	72.69	77.05	75.85	84.17	80.82	75.24
Symmetrical Uncertainty	51.49	71.41	76.29	74.35	82.96	72.25	60.1
Evolutionary Algorithm	55.04	70.31	80.65	72.62	85.29	80.85	67.65

TABLE II. NUMBER OF ATTRIBUTES SELECTED BY FIVE EVALUATORS OF THE ORIGINAL 2147 STATISTICAL ATTRIBUTES

Attribute Selection Evaluator	**Ranker Search Cut-off**	**No. of attributes selected**
OneR	60.0	44
Information Gain	0.5	31
Correlation	0.3	26
Symmetrical Uncertainty	0.25	36
Evolutionary Algorithm	N/A	99

The goal is to infer the current time value of C^t given the data $X^{t:t-T} = \{X^t, X^{t-1},...,X^{t-T}\}$ and the prior knowledge of the class, which is attained by the a-posteriori probability $P(C^t \mid C^{t-1:t-T}, X^{t:t-T})$. The superscript notation denotes the set of values over a time interval.

Three decision trees were developed. Generated by the *C4.5* algorithm [2], a *J48* tree splits each decision based on information gain, due to the measure of entropy in a leaf node.

A Random Tree is generated through a stochastic process that will consider a random number of attributes at each node. A Random Forest is the process of generating multiple Random Trees [3]. A Multilayer Perceptron (MLP) model was generated, a feedforward Neural Network in that cycles are not formed by neurons. An MLP was implemented due to its ability to classify data points that are not linearly separable in Euclidean space [30]. A model was also trained using a Support Vector Machine (SVM), which classifies labelled data through a process of supervised learning, where examples are mapped out in space and classification is performed by the closest area in which the unknown class data falls [31]. In particular, an improved version of *Platt's Sequential Minimal Optimization* (SMO) was used to train the SVM [32], [33].

V. PRELIMINARY RESULTS

The five generated sets from the original dataset are shown in Table I. Five different algorithms were chosen, and their results ranked by their individual scores. Arbitrary cut off points were implemented where the scores closed in on either 0 or the lowest score present if there were no zero values. The values given are incomparable between algorithms due to their unique methods of giving score. The MLP was given 2000 epochs to train with the number of nodes on layers set to the default "a" setting, dynamically calculated by $n = (attributes + classes)/2$ for each dataset it was trained on. A *Zero Rules* classifier was run as a benchmark, and with close to equally distributed data, set a model accuracy of 33.36% on all datasets for comparison. We can observe from when compared to all other classifiers which are not naive. The most effective model was a Random Forest classifier along with the dataset created by the *OneR Attribute Selector*, which had a high accuracy of 87.16% when classifying the data into one of the three mental states. Preliminary results for each of the datasets and their trained models are presented in Table II. For each test, 10-fold cross validation was used to train the model. All random seeds were set to their default value of 1. Table II that all of the models far outperformed the benchmarks set by the *Zero Rules* classifier, the lowest being 51.49% (Symmetrical Uncertainty dataset with a Naive Bayes classifier). It is reasonable to assume that the naivety in not considering attribute relationships has led to poorer results.

VI. CONCLUSION

This paper presented a study on mental state classification based on EEG signals, it proposed a set of features using a short-term windowing extracted from five signals from an

EEG sensor to categorise three different states: *neutral*, *relaxed* and *concentrated*. A dataset was created using data from five individuals in sessions lasting one minute for each state. The primary goal of this work was to find appropriate set of features by testing multiple feature selection algorithms and classification models that provide acceptable accuracy performance on the dataset that can be useful for human-machine interaction. From the multiple feature sets and models produced, the most accurate is a Random Forest classifier on an attribute selected by the *OneR* ruleset, with a prediction accuracy of 87.16%. Future work will be focused on comparing our best results with deep learning strategies and implementing a real-time application to: (i) control devices, such as robots; and (ii) detect positive and negative moods useful for applications in mental health care.

REFERENCES

[1] Himebaugh, N.L., Begley, C.G., Bradley, A. and Wilkinson, J.A., 2009. Blinking and tear break-up during four visual tasks. Optometry and Vision Science, 86(2), pp. E106-E114.

[2] Quinlan, R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA.

[3] Breiman, L., 2001. Random forests. Machine learning, 45(1), pp.5-32..

[4] Jasper, Herbert H. 1958. "The ten-twenty electrode system of the International Federation." Electroenceph. Clin. Neurophysiol. 370-375.

[5] Kowaleski, J. (2017). BlueMuse.

[6] Barachant, A. (2017). Muselsl.

[7] Bos, D.O., 2006. EEG-based emotion recognition. The Influence of Visual and Auditory Stimuli, 56(3), pp.1-17.

[8] Krigolson, O.E., Williams, C.C., Norton, A., Hassall, C.D. and Colino, F.L., 2017. Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research. Frontiers in neuroscience, 11, p.109.

[9] Li, W., Jaramillo, C. and Li, Y., 2012, January. Development of mind control system for humanoid robot through a brain computer interface. In 2012 International Conference on Intelligent System Design and Engineering Application (pp. 679-682). IEEE.

[10] Rosenzweig, M.R., Breedlove, S.M. and Leiman, A.L., 2002. Biological psychology: An introduction to behavioral, cognitive, and clinical neuroscience. Sinauer Associates.

[11] Abujelala, M., Abellanoza, C., Sharma, A. and Makedon, F., 2016, June. Brain-ee: 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.

[12] Plotnikov, A., Stakheika, N., De Gloria, A., Schatten, C., Bellotti, F., Berta, R., Fiorini, C. and Ansovini, F., 2012, July. Exploiting real-time EEG analysis for assessing flow in games. In 2012 IEEE 12th International Conference on Advanced Learning Technologies (pp. 688-689). IEEE.

[13] Jordan, K.G., 2004. Emergency EEG and continuous EEG monitoring in acute ischemic stroke. J. of Clinical Neurophys., 21(5), pp.341-352.

[14] Ang, K.K., Guan, C., Chua, K.S.G., Ang, B.T., Kuah, C., Wang, C., Phua, K.S., Chin, Z.Y. and Zhang, H., 2010, August. Clinical study of neurorehabilitation in stroke using EEG-based motor imagery brain-computer interface with robotic feedback. 2010 Annual International Conference of the IEEE (pp. 5549-5552).

[15] Tzallas, A.T., Tsipouras, M.G. and Fotiadis, D.I., 2009. Epileptic seizure detection in EEGs using time–frequency analysis. IEEE transactions on information technology in biomedicine, 13(5), pp.703-710

[16] Aarabi, A., Grebe, R. and Wallois, F., 2007. A multistage knowledge-based system for EEG seizure detection in newborn infants. Clinical Neurophysiology, 118(12), pp.2781-2797.

[17] Ghosh-Dastidar, S. and Adeli, H., 2007. Improved spiking neural networks for EEG classification and epilepsy and seizure detection. Integrated Computer-Aided Engineering, 14(3), pp.187-212.

[18] Chai, T.Y., Woo, S.S., Rizon, M. and Tan, C.S., 2010. Classification of human emotions from EEG signals using statistical features and neural network. In International (Vol. 1, No. 3, pp. 1-6). Penerbit UTHM.

[19] Tanaka, H., Hayashi, M. and Hori, T., 1996. Statistical features of hypnagogic EEG measured by a new scoring system. Sleep, 19(9), pp.731-738.

[20] Fraiwan, L., Lweesy, K., Khasawneh, N., Wenz, H. and Dickhaus, H., 2012. Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier. Computer methods and programs in biomedicine, 108(1), pp.10-19.

[21] Rytkönen, K.M., Zitting, J. and Porkka-Heiskanen, T., 2011. Automated sleep scoring in rats and mice using the naive Bayes classifier. Journal of neuroscience methods, 202(1), pp.60-64.

[22] Vital, J.P., Faria, D.R., Dias, G., Couceiro, M.S., Coutinho, F. and Ferreira, N.M., 2017. Combining discriminative spatiotemporal features for daily life activity recognition using wearable motion sensing suit. Pattern Analysis and Applications, 20(4), pp.1179-1194.

[23] Faria, D.R., Vieira, M., Premebida, C. and Nunes, U., 2015, August. 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.

[24] University of Waikato. 2018. OneR. [online] Weka.sourceforge.net. Available at: http://weka.sourceforge.net/doc.dev/weka/classifiers/rules/OneR.html [Accessed 9 Aug. 2018].

[25] University of Waikato. 2018. InfoGainAttributeEval. [online] Weka.sourceforge.net. Available at: http://weka.sourceforge.net/doc.dev/weka/attributeSelection/InfoGainAttributeEval.html [Accessed 9 Aug. 2018].

[26] Pearson, K., 1895. Note on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London, 58, pp.240-242.

[27] Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J., 2016. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

[28] Pearl, Judea 2000. Causality: Models, Reasoning, and Inference. Cambridge University Press. ISBN 0-521-77362-8.

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

[30] Rosenblatt, F., 1961. Principles of neurodynamics. perceptrons and the theory of brain mechanisms (No. VG-1196-G-8). CORNELL AERONAUTICAL LAB INC BUFFALO NY.

[31] Cortes, C. and Vapnik, V., 1995. Support-vector networks. Machine learning, 20(3), pp.273-297.

[32] Platt, J.C., 1999. 12 fast training of support vector machines using sequential minimal optimization. Adv. in kernel methods, pp.185-208.

[33] Keerthi, S.S., Shevade, S.K., Bhattacharyya, C. and Murthy, K.R.K., 2001. Improvements to Platt's SMO algorithm for SVM classifier design. Neural computation, 13(3), pp.637-649.

[34] Back, T., 1996. Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms. Oxford university press.

[35] Shenoy, P; Miller, KJ; Ojemann, JG; Rao, RPN (2007). Generalized features for electrocorticographic BCIs. IEEE Transactions on Biomedical Eng. 55 (1), pp. 273–80.