Feature Reconstruction Based Channel Selection for Emotion Recognition Using EEG

J. R. Msonda[1,2], Z. He[3] and C. Lu[1]

1. Department of Computer Science, Aberystwyth University, Wales, UK
2. School of Engineering, Malawi University of Business and Applied Sciences, Blantyre, Malawi
3. Department of Psychology, Aberystwyth University, Wales, UK
{jam139, cul, zhh2}@ aber.ac.uk

Abstract— There has been a surge in the use of consumer grade wearable Electroencephalogram (EEG) devices for emotion discrimination tasks in various research laboratories in recent times. The obvious advantages carried by these compared to medical grade equipment are reduced costs and portability, which enable monitoring for a longer term and in more natural environment. Different manufacturers of consumer grade EEG devices place the electrodes at different locations. In this paper we present a novel method for determining locations of the fewest electrodes with the most emotion valence discriminative power. It starts with feature generation and selection for identifying positional features for the classification task, followed by channel selection that minimizes the feature reconstruction error. To evaluate the proposed methods, benchmarking analysis was done using leave out one subject cross validation with various machine learning models, using three public datasets. Results show with 8 electrodes AUC scores of 0.78, 0.8 and 0.67 are obtained for AMIGOS, DREAMER and DEAP datasets, respectively on emotion valence classification task. It is further observed that out of the best 8 channels selected, frontal (F8), parietal (P7), and temporal (T8 and T7) are common brain areas which are active during emotion processing across all the three datasets.

I. INTRODUCTION

The proliferation of wearable Electroencephalography (EEG) devices has given birth to an unprecedented curiosity in the study of brain signals within the scientific community. In EEG electrodes (often tiny metal discs) are placed on the scalp of an individual, to capture brain activity. These electrodes measure distribution of voltage fields at different points and changing time. Established medical grade EEG acquisition systems can use as many as 256 [1] electrodes to capture brain waves. While the use of a dense spatial resolution has potential for better localization of cerebral activities [2]–[4], such systems are difficult to set up and can only be used in laboratory settings where the subject is stationary. Moreover, high spatial resolution medical grade EEG systems are prohibitively expensive for many non-medical laboratories. Rapid technological progress made in areas of wearable sensors and machine intelligence have facilitated production of portable EEG devices with varying electrode configurations. On the higher end, there are devices such as Emotive Epoc with 14 channels while the lower end has devices with as few as a single electrode e.g. Neurosky Mindwave [5]. Even though they suffer from imbalanced distribution of electrodes and lower signal quality [6], these consumer grade devices remain attractive due to their affordability and portability benefits.

In this paper we present a simple method for determining a minimal EEG channel configuration for emotion recognition. In the method, we start by identifying suitable features for emotion recognition and thereafter, reconstruct the selected features using fewer channels. To investigate the potential issue of data source dependency and effect of product manufacturers, we apply this algorithm on up to three different benchmarking datasets. Adoption of this channel set will help in the design of cheaper wearable EEG devices for emotion recognition tasks as opposed to using general purpose EEG devices.

II. RELATED WORK

The cerebral cortex of a human brain has four lobes: frontal, parietal, temporal, and occipital, as shown in Figure 1. When recording EEG signals, it is common practice to use a 10-20 system of placing electrodes on the scalp. In this system, each position is known by its relative location and the underlying region of the brain. Accordingly, electrode sites include pre-frontal (Fp), frontal (F), temporal (T), parietal (P), occipital (O), and central (C). As the name suggests, the central (C) electrodes are located at the 'centre' of the scalp and are surrounded by the frontal, temporal, and parietal electrodes. Location is denoted by a number, with even number signifying locations on the right side of the brain while odd numbers represent the left side. The midline along the length of the brain is designated as zero (z). As an example, F3, F4 and Fz are electrode positions on the left, right and midline of the frontal lobe respectively.

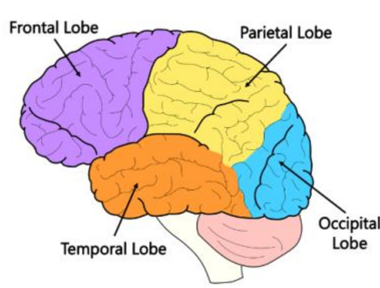

Figure 1. Image showing brain lobes - Source [32]

Although many scholars, including [7] and [8], are of the view that the frontal lobe plays an important role in the processing of emotions, there does not seem to be a general agreement on the matter. Following their study on channel selection [9], [10] observe that the frontal and temporal are the most important regions for emotion activity. Zheng et al also agree with this position by demonstrating that the best four channel positions are FT7, FT8, T7 and T8 [11]. It is argued in [12] that 'electrodes located over the parietal and central-parietal lobe are favored over occipital and frontal lobes.' On the contrary, [13] found that T7 and T8 were better at discriminating emotions than other brain areas. Goshvarpour identified FP1, C3, Cp1, P3, and Pz as being the most discriminative positions for emotion classification [14]. The observation in [15] and [16] is that frontal, parietal lobes are the ones which are associated with emotion processing. In [17], 'the channel pairs: (P7, P8) and (PO7, PO8) are selected as a most significant channel pair for detecting different emotions.'

In terms of channel selection methods, [9] used group sparse representations of EEG data samples to perform canonical correlation analysis. Different channel sets were realized for each frequency band. Similarly each frequency band got an independent set of channels after performing stepwise discriminant analysis in [18]. Maximum correlation redundancy algorithm was applied to DEAP dataset in [19] to obtain 10 best electrode positions for emotion recognition as FP1, FP2, F3, F4, T7, 78, PO3, PO4, O1 and O2. [20] took advantage of the feature and channel weights acquired through application of reliefF algorithm on DEAP dataset. This paper concluded that FP1, T7, PO4, PZ, FP2, OZ, F8, T8, P4, O1, FC5, C3 and PC2 were ideal locations for classification of emotions. Another approach that made use of DEAP dataset is described in [21]. Here, channels FC1, P3, PZ, OZ, CP2, C4, F4 and FZ were selected by exploiting normalized mutual information technique. Rizon et al made use of Fuzzy C-Means Clustering (FCM) to group EEG based on variance characteristics [17]. The challenge with the methods discussed is that a majority of them were only tested on single datasets. This makes it difficult to compare them.

Since it was proposed by Russell [22], it has become common practice for researchers to represent emotional states using a circumplex model. In this model, affective states are placed on a 2-dimensional continuous scale of valence and arousal. Valance signifies how positive or negative the feeling is whereas arousal denotes the extent of excitement or calmness that comes with a particular emotion. One implementation of the model can be seen in Figure 2.

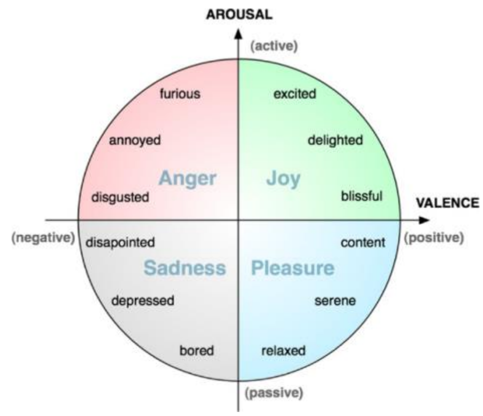

Figure 2. Allocation of different affective states - Adopted from [31]

In this work, we are particularly concerned with channel selection for binary classification of emotions into being positive or negative. We developed a novel feature-based channel selection algorithm in order to reduce the number of electrodes required for classification of emotion valence. This method consists of two phases: 1) finding an optimal subset of positional features across all channels for the classification task, and 2) finding a substitute positional feature subset using fewer positions by minimizing the reconstruction errors for the selected features. A benchmarking analysis using three public datasets has been conducted for evaluation.

III. METHODS

A. Data

In this study we analyze three different benchmarking datasets: A Database for Emotion Analysis Using Physiological Signals (DEAP) [23]; a Database for Emotion Recognition through EEG and ECG Signals (DREAMER) [24]; and a dataset for Multimodal research of affect, personality traits and mood on Individuals and GrOupS (AMIGOS) [25] .

All the three datasets contain EEG signals recorded as participants watched emotion stimulating video clips. After each session, these clips were rated using valence and arousal, amongst other measures. The sampling rate for DEAP was 512Hz while DREAMER (23 participants) and AMIGOS (40 subjects) data had a frequency of 128Hz. DEAP (32 subjects) data was collected using 32 channel Biosemi Active Two system while the other two datasets had a 14 electrode Emotiv EPOC portable device.

B. Preprocessing

We sought to classify emotion valence as either positive or negative using EEG signals. Noting that valence ratings for the AMIGOS and DEAP datasets ranged from 1 to 9 and DREAMER database had scores between 1 and 5, binary labels were generated based on those scores. For

AMIGOS and DEAP datasets, valence levels less than 3 were labelled as negative while those over 7 were categorised as positive. Less than 2 and greater than 4 were respectively classified as negative and positive valence levels in the DREAMER dataset.

To have uniform duration we extracted EEG signals captured in the final 60 seconds of each recording for analysis. Furthermore, to increase the number of training examples, we segmented the extracted signals into 5 second epochs. In addition, channel mean subtraction was done for data normalization.

C. Feature Generation

As depicted in Figure 3, feature extraction techniques were applied to four different representations of the input signal, namely: full band signal, sub-band, empirical mode and wavelet decomposed signal versions.

Sub-band Decomposition: Chebyshev type I bandpass filters were used to decompose EEG signals into delta, theta, alpha, beta, and gamma bands using ranges 0.5-4Hz, 4-8Hz, 8-13Hz, 13-30Hz, and 30-60Hz respectively.

Empirical Mode Decomposition: Empirical mode decomposition (EMD) is an adaptive technique used to split a time series signal into separate components/resolutions referred to as intrinsic mode function (IMFs). We extracted a maximum of 3 IMFs from each EEG signal.

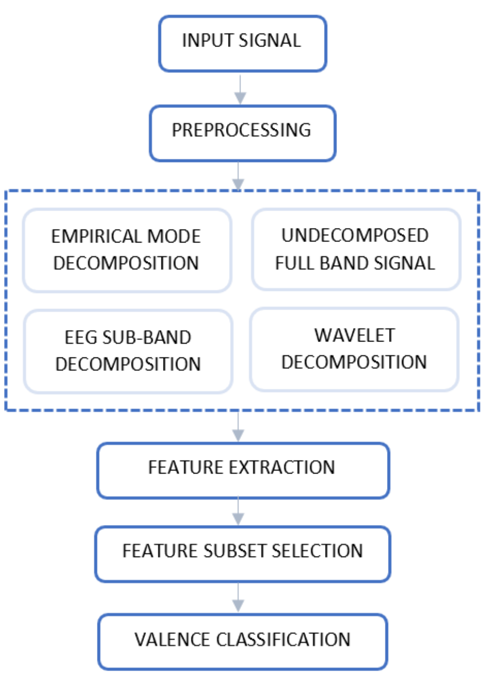

Figure 3. Emotion valence classification pipeline

Wavelet Decomposition: A wavelet is a short burst of a signal which rises from zero and quickly falls back to the zero level. In a wavelet transform, an input signal is convolved with the mother wavelet to obtain time information. We used a mother wavelet with 4 coefficients from the Daubechies family (db4).

Features: Having obtained the original signals and decomposed versions, we used packages such as PyWavelets [26], Librosa [27] and mne-features [28] Python packages to generate features for classification. Features extracted include: fractal dimension related features like Petrosian, katz, higuchi and detrended_fluctuation; entropy based features such as permutation entropy, spectral entropy, singular value decomposition entropy, approximate entropy and sample entropy; statistical features measuring skewness, kurtosis etc.; and other features like zero crossing rate, power, first difference, Teager Kaiser energy, Hjorth mobility, peak width and so on.

D. Feature Selection

Before channel selection was done, feature selection methods were used to remove noisy features and reduce the time complexity of our subsequent channel selection operation. We experimented on various feature selection methods such as filtering using ANOVA (Analysis of Variance), filtering using Random Forests feature importance scores, and the wrapper approach BorutaShap [29]. Logistic regression based recursive feature elimination technique was utilized eventually for this work to select feature subsets based on its empirical performance and computation efficiency.

E. Channel Selection

In order to select the minimal number of channels, we propose to minimise the reconstruction error for the selected discriminative positional features, i.e. mean squared error between feature values obtained using original electrode positions and those realised using substitute positions.

To illustrate the idea behind the algorithm, assume we had an EEG device with five electrode positions - F3, T8, T7, P7 and AF4. And that for each position we extracted four features power (pow), entropy (ent), fractal dimension (fd) and zero crossing rate (zr). Assume further that we performed feature selection on the 20 positional features (5 positions by 4 features) and obtained the best four positional features, namely, F3_pow, T8_ent, T8_zr and P7_fd. To get the best two electrode positions using our method, we:

1) Collect all possible combinations of two from five electrodes as [F3, T8], [F3, T7], [F3, P7], [F3,

AF4], [T8, T7], [T8, P7], [T8, AF4], [T7, P7], [T7, AF4] and [P7, AF4].

2) For each of the combinations in 1), we find the best substitute for each of the target features, by choosing the positional feature that has the smallest mean squared error, or distance to the target features, which can be considered as the reconstructed error for this target feature. E.g., for candidate combination [T7, AF4], we compute both $\| T7\_pow - F3\_pow \|2$ and $\| AF4\_pow - F3\_pow \|2$. The one with a lower value is a better substitute for F3_pow. The same procedure will be followed to get substitutes for T8_ent, T8_zr and P7_fd. Subsequently, for [T7, AF4] we could, for instance, get T7_pow, AF4_ent, T7_zr and AF4_fd as substitutes.

3) The reconstruction error between the selected channel combination and the target features can then be computed as the total mean squared errors from all substitute features to the targets (as computed in 2).

4) The best combination channel subset will then be the ones with the minimum reconstructed error or total mean square error.

We formalize the algorithm as follows: Assume that EEG data has N channels, and for each channel position we have M feature types. Then a set of F positional feature vectors, of length D for D data points, can be described as $F \in V^{M \times N}$. If out of F we have B useful features, then $B \subset F$ and $B \in V^{Q \times R}$ such that $Q \leq M$ and $R \leq N$. Figure 4 lists an algorithm that can be used to reconstitute set B using the fewest p (out of N) channels/electrode positions from a pool of $\binom{N}{p}$ possible combinations.

Algorithm 1: Channel Selection by Feature Reconstruction

REQUIRE: $\boldsymbol{B \subset F}$, $\boldsymbol{F \in V^{M \times N}}$

1: FOR each combination, $\boldsymbol{C}$ in $\binom{\boldsymbol{N}}{\boldsymbol{p}}$, **DO**

2: **FOR** each feature vector $\boldsymbol{F_{ij} \in B}$ **DO**

3: $\boldsymbol{e_i \longleftarrow Min_{k \in C}\left(\|F_{ij} - F_{ik}\|^2\right)}$,

where,

$\boldsymbol{i = feature\ type}$.

$\boldsymbol{j = original\ position}$

$\boldsymbol{k = candidate\ substitute\ position}$

$\boldsymbol{A_i = best\ substitute\ for\ feature\ type\ i}$

4: **END FOR**

5: $\boldsymbol{E_c \longleftarrow \sum_{i=1}^{Q} e_i}$

where,

$\boldsymbol{E_c = reconstruction\ error\ for\ combination\ C}$

6: END FOR

7: RETURN Channel set C with minimum $\boldsymbol{E_c}$

Figure 4. Channel selection by feature reconstruction

IV. Results and Discussion

Using Logistic regression based recursive feature elimination we obtained 20 positional features for each dataset and fed these features into our channel selection algorithm. For each number of selected channels AdaBoost, Logistic Regression, Linear Support Vector Classifier (SVC), second order polynomial SVC and Random Forest (RF) models were used. The models were implemented using default settings in python library Scikit-learn.

The performance of the selected channels and features subsets were evaluated using the Area Under the receiver operating characteristic Curve (AUC) based on leave out one subject cross validation. Before application of the electrode number reduction algorithm, baseline scores were obtained. For the DREAMER dataset, we observed the highest score of 0.83 using Linear SVC model on features extracted through recursive feature elimination method. AMIGOS and DEAP datasets had scores of 0.819 and 0.695 respectively.

Results obtained after applying our algorithm are detailed in Table 1. Generally, the table shows that as we increased the number of channels, performance also improved. With 4 channels, a mean AUC score of 0.72 was obtained in the AMIGOS dataset compared to a baseline of 0.82. In the DREAMER dataset, the mean AUC score using 4 channels was 0.76 against a baseline of 0.83. Using 3 channels in the DEAP dataset, a score of 0.61 was achieved, the baseline being 0.7. With 8 channels, the best mean AUC scores are 0.78, 0.8 and 0.67 for AMIGOS, DREAMER and DEAP datasets in that sequence.

To the best of our knowledge, [30] is the only emotion valence classification study comparable to our research basing on choice of data and evaluation technique. They too assessed their method using leave out one subject cross validation on DEAP dataset and reported an AUC score of 0.605. Thus, results realized using 8 best channels in our algorithm were slightly better that these.

Generally, scores achieved using DEAP were the lowest of the three. As has been suggested in [30], this may be due to the poor quality of the dataset. It is within reasonable expectation that if the classification models used in our study can be fine-tuned, a much more acceptable performance can be achieved.

Other than the scores, it is noteworthy that right from 4 channels, there seem to be commonality in the channels

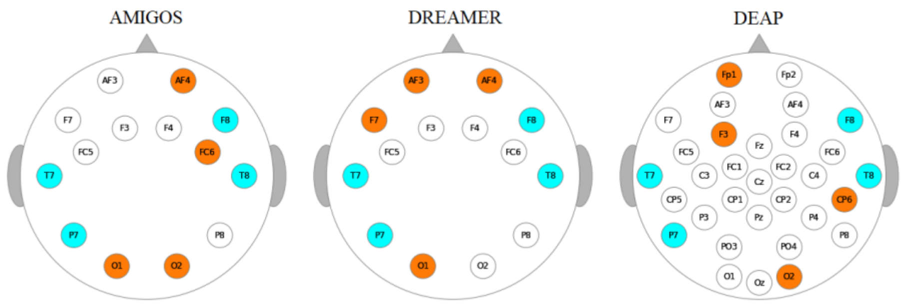

Figure 5. Topographic maps showing best 8 electrode positions in colors for AMIGOS, DREAMER and DEAP datasets. Positions highlighted in cyan are common amongst the three datasets.

Table 1. Model performance with reduced number of electrodes

Dataset	Selected Channels	Number of Channels	Leave Out One Subject Model Performance (AUC/STD)				
			AdaBoost	*LR*	*Linear SVC*	*Polynomial SVC*	*RF*
AMIGOS	All	14	0.6/0.15	**0.82/0.1**	0.79/0.12	0.78/0.12	0.73/0.13
AMIGOS	['AF4', 'F8', 'FC6', 'O1', 'O2', 'P7', 'T7', 'T8']	8	0.63/0.11	**0.78/0.09**	0.77/0.11	0.75/0.13	0.73/0.15
AMIGOS	['AF4', 'F8', 'FC6', 'O1', 'P7', 'T7', 'T8']	7	0.61/0.17	**0.78/0.1**	0.77/0.1	0.77/0.14	0.72/0.17
AMIGOS	['AF4', 'FC6', 'O1', 'P7', 'T7', 'T8']	6	0.64/0.11	0.74/0.15	**0.75/0.11**	0.73/0.13	0.74/0.16
AMIGOS	['AF4', 'FC6', 'O1', 'P7', 'T8']	5	0.66/0.14	0.73/0.16	**0.75/0.11**	0.72/0.14	0.74/0.16
AMIGOS	['FC6', 'O1', 'P7', 'T8']	4	0.68/0.14	**0.72/0.13**	0.71/0.12	0.72/0.13	0.69/0.15
AMIGOS	['FC6', 'O1', 'P7']	3	0.58/0.18	**0.62/0.21**	0.6/0.24	0.59/0.24	0.55/0.19
AMIGOS	['FC6', 'O1']	2	0.54/0.18	0.59/0.25	**0.58/0.25**	0.56/0.26	0.56/0.18
DREAMER	All	14	0.65/0.19	0.83/0.14	**0.83/0.14**	0.82/0.12	0.8/0.12
DREAMER	['AF3', 'AF4', 'F7', 'F8', 'O1', 'P7', 'T7', 'T8']	8	0.62/0.19	0.8/0.14	**0.8/0.14**	0.75/0.14	0.74/0.1
DREAMER	['AF4', 'F7', 'F8', 'FC6', 'O1', 'P7', 'T8']	7	0.63/0.17	**0.8/0.13**	0.8/0.14	0.76/0.12	0.72/0.12
DREAMER	['AF3', 'AF4', 'F7', 'F8', 'P7', 'T8']	6	0.66/0.21	0.79/0.13	**0.79/0.13**	0.72/0.18	0.72/0.11
DREAMER	['AF4', 'F7', 'F8', 'P7', 'T8']	5	0.64/0.2	**0.75/0.14**	0.75/0.15	0.7/0.21	0.7/0.13
DREAMER	['AF4', 'F8', 'P7', 'T8']	4	0.63/0.2	**0.76/0.13**	0.76/0.15	0.73/0.17	0.69/0.12
DREAMER	['AF4', 'P7', 'T8']	3	0.61/0.21	**0.75/0.14**	0.75/0.16	0.74/0.18	0.71/0.13
DREAMER	['AF4', 'P7']	2	0.52/0.18	0.59/0.18	**0.61/0.18**	0.59/0.19	0.5/0.15
DEAP	All	32	0.61/0.16	**0.7/0.15**	0.7/0.15	0.69/0.16	0.66/0.17
DEAP	['CP6', 'F3', 'F8', 'Fp1', 'O2', 'P7', 'T7', 'T8']	8	0.59/0.11	0.66/0.12	**0.67/0.15**	0.67/0.14	0.64/0.13
DEAP	['F3', 'F8', 'Fp1', 'O2', 'P7', 'T7', 'T8']	7	0.58/0.1	0.66/0.12	**0.66/0.14**	0.65/0.14	0.63/0.12
DEAP	['F3', 'Fp1', 'O2', 'P7', 'T7', 'T8']	6	0.56/0.11	**0.65/0.1**	0.65/0.12	0.64/0.13	0.61/0.14
DEAP	['F3', 'Fp1', 'O2', 'P7', 'T8']	5	0.51/0.11	0.6/0.12	**0.6/0.15**	0.59/0.15	0.59/0.13
DEAP	['F3', 'Fp1', 'P7', 'T8']	4	0.51/0.09	0.54/0.14	0.57/0.17	**0.57/0.17**	0.57/0.14
DEAP	['CP6', 'F3', 'T8']	3	0.6/0.09	0.57/0.12	0.59/0.15	0.61/0.12	**0.61/0.14**
DEAP	['CP6', 'T8']	2	0.55/0.15	0.53/0.12	0.57/0.17	**0.59/0.13**	0.58/0.17

selected. P7 and T8 are chosen in all three datasets. At 8 channel level, there are overlaps of 4 electrode positions including F8, P7, T8 and T7 as displayed in Figure 5. This suggests that these could be important electrode positions for discrimination of emotions. In fact, for AMIGOS and DREAMER in which the same EEG device was used, there are up to 6 common channels - AF4, F8, O1, P7, T7 and T8.

Potentially, O1 and O2 which lie very close to each other may be the 5th common position among the three datasets. After all, in practice when EEG headset are worn, electrodes do not rest at theorical channel spots on the scalp. This notwithstanding, results indicate that frontal(F8), parietal (P7), and temporal (T8 and T7) areas are active during emotion processing. Currently, there is no agreement in literature as different studies come up with different findings. Our results, however, include three different datasets and chances that the intersection of electrode positions found is not a mere coincidence are significantly minimised.

The possibility of including O1 and O2 however, needs further analysis as this area of the brain is associated with visual perception. Given that in all the three datasets emotion were elicited through video clips, it may be the reason why activity has been registered in this region.

While the L2 norm remains a valid choice for a reconstruction error, it tends to penalize outliers more than other methods such as L1 norm. Unfortunately, bad channels due to poor contact during EEG recording are a common occurrence, and hence L2 may not be the best method.

V. Conclusions

A study of emotions is significant as they inform decision making in various spheres of life. In this paper we have demonstrated that a simple feature reconstruction-based algorithm can be used to select channels. We have further shown that channels around positions AF4, F8, O1, P7, T7 and T8 are excited when a subject is under emotion stimulation. Future work can include further improvement of the algorithm and use of a wider range of features and datasets. An exploration of alternative measures of reconstruction error may also be warranted.

Acknowledgments

We express our gratitude to Super Computing Wales (SCW) for providing computing resources used in this research. We also thank Aberystwyth University and Malawi University of Business and Applied Science (MUBAS) for sponsoring the study.

References

[1] T. M. Lau, J. T. Gwin, and D. P. Ferris, "How Many Electrodes Are Really Needed for EEG-Based Mobile Brain Imaging?," *J. Behav. Brain Sci.*, vol. 02, no. 03, pp. 387–393, 2012, doi: 10.4236/jbbs.2012.23044.

[2] M. Odabaee, W. J. Freeman, P. B. Colditz, C. Ramon, and S. Vanhatalo, "Spatial patterning of the neonatal EEG suggests a need for a high number of electrodes," *Neuroimage*, vol. 68, pp. 229–235, 2013, doi: https://doi.org/10.1016/j.neuroimage.2012.11.062.

[3] G. Lantz, R. Grave de Peralta, L. Spinelli, M. Seeck, and C. M. Michel, "Epileptic source localization with high density EEG: how many electrodes are needed?," *Clin. Neurophysiol.*, vol. 114, no. 1, pp. 63–69, 2003, doi: https://doi.org/10.1016/S1388-2457(02)00337-1.

[4] D. M. Tucker, M. Brown, P. Luu, and M. D. Holmes, "Discharges in ventromedial frontal cortex during absence spells," *Epilepsy Behav.*, vol. 11, no. 4, pp. 546–557, 2007, doi: 10.1016/j.yebeh.2007.04.023.

[5] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affect. Comput.*, vol. 10, pp. 374–393, 2019, doi: 10.1109/TAFFC.2017.2714671.

[6] H. Rieiro et al., "Validation of Electroencephalographic Recordings Obtained with a Consumer-Grade, Single Dry Electrode, Low-Cost Device: A Comparative Study," *Sensors*, vol. 19, no. 12, 2019, doi: 10.3390/s19122808.

[7] S. Valenzi, T. Islam, P. Jurica, and A. Cichocki, "Individual Classification of Emotions Using EEG," *J. Biomed. Sci. Eng.*, vol. 07, no. 08, pp. 604–620, 2014, doi: 10.4236/jbise.2014.78061.

[8] N. Thammasan, K. Moriyama, K. ichi Fukui, and M. Numao, "Familiarity effects in EEG-based emotion recognition," *Brain Informatics*, vol. 4, no. 1, pp. 39–50, 2017, doi: 10.1007/s40708-016-0051-5.

[9] W. Zheng, "Multichannel EEG-Based Emotion Recognition via Group Sparse Canonical Correlation Analysis," *IEEE Trans. Cogn. Dev. Syst.*, vol. 9, no. 3, pp. 281–290, 2017, doi: 10.1109/TCDS.2016.2587290.

[10] M. L. Phillips, W. C. Drevets, S. L. Rauch, and R. Lane, "Neurobiology of emotion perception I: The neural basis of normal emotion perception," *Biol. Psychiatry*, vol. 54, no. 5, pp. 504–514, 2003, doi: 10.1016/S0006-3223(03)00168-9.

[11] W. L. Zheng and B. L. Lu, "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks," *IEEE Trans. Auton. Ment. Dev.*, vol. 7, no. 3, pp. 162–175, 2015, doi: 10.1109/TAMD.2015.2431497.

[12] R. Jenke, A. Peer, and M. Buss, "Feature Extraction and Selection for Emotion Recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 327–339, 2014, doi: 10.1109/TAFFC.2014.2339834.

[13] N. Jatupaiboon, S. Pan-Ngum, and P. Israsena, "Real-time EEG-based happiness detection system," *Sci. World J.*, vol. 2013, 2013, doi: 10.1155/2013/618649.

[14] A. Goshvarpour and A. Goshvarpour, "A Novel Approach for EEG Electrode Selection in Automated Emotion Recognition Based on Lagged Poincare's Indices and sLORETA," *Cognit. Comput.*, vol. 12, no. 3, pp. 602–618, 2020, doi: 10.1007/s12559-019-09699-z.

[15] J. Zhang, M. Chen, S. Zhao, S. Hu, Z. Shi, and Y. Cao, "ReliefF-Based EEG Sensor Selection Methods for Emotion Recognition," *Sensors*, vol. 16, no. 10, 2016, doi: 10.3390/s16101558.

[16] A. Chatchinarat, K. W. Wong, and C. C. Fung, "A comparison study on the relationship between the selection of EEG electrode channels and frequency bands used in classification for emotion recognition," in *2016 International Conference on Machine Learning and Cybernetics (ICMLC)*, 2016, vol. 1, pp. 251–256, doi: 10.1109/ICMLC.2016.7860909.

[17] M. Rizon, M. Murugappan, R. Nagarajan, and S. Yaacob, "Asymmetric ratio and FCM based salient channel selection for human emotion detection using EEG," *WSEAS Trans. Signal Process.*, vol. 4, no. 10, pp. 596–603, 2008, doi: 10.1.1.558.4201.

[18] E. S. Pane, A. D. Wibawa, and M. H. Pumomo, "Channel Selection of EEG Emotion Recognition using Stepwise Discriminant Analysis," in *2018 International Conference on Computer Engineering, Network and Intelligent Multimedia, CENIM 2018 - Proceeding*, 2018, pp. 14–19, doi: 10.1109/CENIM.2018.8711196.

[19] Q. Meng, J. Yan, and H. Xu, "Research on EEG signal recognition based on channel selection," *Proc. - 2017 Chinese Autom. Congr. CAC*, 2017, pp. 6413–6417, 2017, doi: 10.1109/CAC.2017.8243933.

[20] L. Tong, J. Zhao, and W. Fu, "Emotion Recognition and Channel Selection Based on EEG Signal," in *2018 11th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, 2018, pp. 101–105, doi: 10.1109/ICICTA.2018.00031.

[21] Z. M. Wang, S. Y. Hu, and H. Song, "Channel Selection Method for EEG Emotion Recognition Using Normalized Mutual Information," *IEEE Access*, vol. 7, pp. 143303–143311, 2019, doi: 10.1109/ACCESS.2019.2944273.

[22] J. A. Russell, "A circumplex model of affect.," *J. Pers. Soc. Psychol.*, vol. 39, no. 6, pp. 1161–1178, 1980, doi: 10.1037/h0077714.

[23] S. Koelstra et al., "DEAP: A Database for Emotion Analysis ;Using Physiological Signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, 2012, doi: 10.1109/T-AFFC.2011.15.

[24] S. Katsigiannis and N. Ramzan, "DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices," *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 1, pp. 98–107, 2018, doi: 10.1109/JBHI.2017.2688239.

[25] J. A. Miranda Correa, M. K. Abadi, N. Sebe, and I. Patras, "AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups," *IEEE Trans. Affect. Comput.*, p. 1, 2018, doi: 10.1109/TAFFC.2018.2884461.

[26] G. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary, "PyWavelets: A Python package for wavelet analysis,"

J. Open Source Softw., vol. 4, no. 36, p. 1237, 2019, doi: 10.21105/joss.01237.

[27] B. McFee et al., "librosa: Audio and Music Signal Analysis in Python," in *Proceedings of the 14th Python in Science Conference*, 2015, pp. 18–24, doi: 10.25080/majora-7b98e3ed-003.

[28] J.-B. Schiratti, J.-E. Le Douget, M. Le Van Quyen, S. Essid, and A. Gramfort, "An Ensemble Learning Approach to Detect Epileptic Seizures from Long Intracranial EEG Recordings," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 856–860, doi: 10.1109/ICASSP.2018.8461489.

[29] E. Keany, "BorutaShap : A wrapper feature selection method which combines the Boruta feature selection algorithm with Shapley values." *Zenodo, 2020, doi: 10.5281/zenodo.4247610*.

[30] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring EEG Features in Cross-Subject Emotion Recognition.," *Front. Neurosci.*, vol. 12, p. 162, 2018, doi: 10.3389/fnins.2018.00162.

[31] A. Yazdani, E. Skodras, N. Fakotakis, and T. Ebrahimi, "Multimedia content analysis for emotional characterization of music video clips," *Eurasip J. Image Video Process.*, vol. 2013, 2013, doi: 10.1186/1687-5281-2013-26.

[32] K. Guenther, "Brain Lobes." *https://brainframe-kids.com/brain/facts-lobes.htm* (accessed Apr. 09, 2021).