On the benefits of using Hidden Markov Models to predict emotions

Yuyan Wu
yuyan.wu@uv.es
Universitat de València
Valencia, Spain

Miguel Arevalillo-Herráez
miguel.arevalillo@uv.es
Universitat de València
Valencia, Spain

Stamos Katsigiannis
stamos.katsigiannis@durham.ac.uk
Durham University
Durham, UK

Naeem Ramzan
naeem.ramzan@uws.ac.uk
University of the West of Scotland
Paisley, Scotland, UK

ABSTRACT

The availability of low-cost wireless physiological sensors has allowed the use of emotion recognition technologies in various applications. In this work, we describe a technique to predict emotional states in Russell's two-dimensional emotion space (valence and arousal), using electroencephalography (EEG), electrocardiography (ECG), and electromyography (EMG) signals. For each of the two dimensions, the proposed method uses a classification scheme based on two Hidden Markov Models (HMMs), with the first one trained using positive samples, and the second one using negative samples. The class of new unseen samples is then decided based on which model returns the highest score. The proposed approach was validated on a recently published dataset that contained physiological signals recordings (EEG, ECG, EMG) acquired during a human-horse interaction experiment. The experimental results demonstrate that this approach achieves a better performance than the published baseline methods, achieving an F1-score of 0.940 for valence and 0.783 for arousal, an improvement of more than +0.12 in both cases.

CCS CONCEPTS

• **Information systems** → **Sentiment analysis**; • **Mathematics of computing** → *Kalman filters and hidden Markov models*; • **Computing methodologies** → **Supervised learning by classification**.

KEYWORDS

Emotion recognition, User modelling, Physiological signal

ACM Reference Format:
Yuyan Wu, Miguel Arevalillo-Herráez, Stamos Katsigiannis, and Naeem Ramzan. 2022. On the benefits of using Hidden Markov Models to predict emotions. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '22), July 4–7, 2022, Barcelona, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3503252.3531323

UMAP '22, July 4–7, 2022, Barcelona, Spain

ACM ISBN 978-1-4503-9207-5/22/07…$15.00
https://doi.org/10.1145/3503252.3531323

1 INTRODUCTION

Affective computing is an emerging research field that aims to endow machines with the ability to recognise human emotions and achieve a more harmonious human-computer interaction (HCI) environment. Emotion recognition may be based on facial expressions, body gesture and movement, physiological signals, and/or other modalities [17, 19]; but physiological signals have the advantages of objectivity, continuity, and real-time data acquisition. Although special devices are generally required to obtain physiological signals, recent advancements have contributed to the development of consumer-grade inexpensive devices that facilitate data acquisition. Nowadays, portable devices are gradually entering human life, and various types of smart bracelets can already detect human heart rate, body temperature, and other data; which can be processed in place or transferred to a cloud-based service in real-time. Accordingly, in recent years, physiological signals have been widely used by researchers in the field of emotion recognition. Among these, electroencephalography (EEG), electrocardiography (ECG), and electromyography (EMG) signals, which capture the electrical activity of the brain, heart, and muscles respectively, have been studied extensively for this task [1, 3, 11].

Hidden Markov Models (HMMs) have been widely used to predict emotions from speech signals [20]; to model the EEG signal's temporal evolution for each user in biometrics tasks; and to estimate concentration heart rate, breath rate, skin conductance and skin temperature [16]. In contrast, there are not many studies that use HMMs for emotion recognition based on physiological signals. Despite the few works that use HMM to classify physiological signals according to emotions showing that performance is not worse than others that use different classifiers [4, 8, 10], most works use other machine learning methods that ignore the temporal evolution of the signal, such as SVM or k-NN [1, 11, 18] or use deep learning methods which generally require large amounts of data for training [2].

In this work, we evaluate the effectiveness of HMM at detecting emotions from physiological signals, in comparison to other classification methods used in the literature. We use the dataset from [1], which contains EEG, ECG, and EMG signals recorded from subjects in a human-horse interaction experiment, in the context of equine assisted therapy (EAT). This dataset was selected because it was collected in a challenging outdoors environment, in contrast to most other datasets, which were captured in a more typical controlled laboratory setting. To conduct a fair comparison, we use the same feature extraction mechanisms as in the original publication [1], but use a novel method based on HMM to locate the physiological signals in Russell's two-dimensional emotion space model. The proposed approach relies on the creation of two HMM models for each of the valence and arousal dimensions, one trained with positive samples and one with negative ones. The class of a new unknown sample is then based on the probability that it belongs to each of the two models. Results show that the proposed technique outperforms the baseline approaches proposed in the original publication.

2 DATASET

The work in [1] demonstrated the effectiveness of using supervised classification methods to predict self-reported emotional states of human subjects in terms of valence and arousal, using EEG, ECG, and EMG signals captured outdoors to evaluate the emotional response of humans when interacting with horses. For the purpose of this research, we have used the same dataset and performed a fair evaluation to determine the performance of HMM as compared to the multiple classification mechanisms used in [1], namely k-Nearest Neighbours (k-NN) for k = 1, 3, 5, Linear Support Vector Machines (LSVM), SVM using a Radial Basis Function kernel (SVM-RBF), Decision Trees (DT), and Linear Discriminant Analysis (LDA).

The authors of [1] designed a collection experiment to acquire human physiological signals while they were interacting with horses within a sand arena, under three different emotional stimulus conditions, namely:

(1) Looking. The subjects stayed with the horse in the same place, for a time period of 4 minutes. The subjects were sat on a chair while the horse could move freely.
(2) Grooming. The subjects used a brush to groom the horse's hair for 2 minutes.
(3) Leading. The subjects led the horse to walk on a predetermined route on the sand arena for 4 minutes.

Two horses of ages 8 and 12 were used in this experiment, and each participant had to perform the three activities mentioned above with both horses. During the human-horse interaction, ECG, EEG, and EMG signals of the subjects were simultaneously recorded. Researchers used SHIMMER^{TM} v2 [7] wireless sensors to capture ECG and EMG signals. ECG signals were captured by using 4 standard electrodes placed on the two ribs and the clavicle. EMG signals were captured by using 3 standard electrodes placed on the trapezius muscle. An Emotiv EPOC+ wireless headset [5] was used for capturing 14-channel EEG signals (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4). All physiological signals were captured at 256 Hz, and were recorded along with timestamps with millisecond precision.

A total of 19 subjects participated in the experiment, 12 male and 7 female. Their age ranged from 19 to 64 years, with an average of 38.05 and a standard deviation of 13.14. 8 of the subjects did not have any prior experience with horses, 5 had prior experience with horses but were unfamiliar with the horses used in the experiment, and the remaining 6 subjects had prior experience with horses and have had contact with the horses used in the experiment. All subjects were required to fill a questionnaire regarding the emotions they felt before, during and after each activity, marking the emotional label for the corresponding physiological signals. This emotional label was transformed into a tuple (valence, arousal) according to Russell's circumplex model of affect [15]. As a result, 70.2% of the samples were associated with high arousal (HA), 29.8% with low arousal (LA), 87.7% with positive valence (PV), and 12.3% with negative valence (NV).

Table 1: Number of features per modality

Modality	No. of features
ECG	84
EMG	21
EEG-PSDavg	70
EEG-Spectral	280
EEG-MFCC [4-40]	168
EEG-MFCC [0.5-40]	168
EEG-MFCC [4-30]	168
EEG-MFCC [0.5-30]	168

3 METHODOLOGY

3.1 Data pre-processing

To allow for a fair comparison of the results and ensure that differences in performance can be attributed to the proposed classification scheme, we followed the same pre-processing approach as in [1].

To this end, ECG signals were first filtered using a median filter with a window of 200 ms to remove the noise of the baseline wander from the collected raw ECG signal. Then, a median filter with a window of 600 ms was applied and the raw signal was subtracted from the filtered one to obtain the denoised result. Finally, a bandpass filter between 0.7-20 Hz was applied to reduce noise.

The pre-processing of the EMG signal consisted of first cutting the peaks with values within 3% of the lowest or highest values, then using a third-order Butterworth FIR low-pass filter with a cut-off frequency of 0.4 Hz, and finally normalising the results in the range $[0, 1]$.

The pre-processing of the EEG signals consisted of first the application of a Butterworth bandpass filter between 0.4 and 65 Hz and then the application of the PREP pre-processing pipeline [6] using the EEGLAB toolbox [9].

3.2 Feature extraction

After pre-processing, 5 different feature sets were extracted from the resulting signals. These features are described below. Similar to the pre-processing step, the same feature extraction approach as in [1] was followed, in order to ensure a fair performance comparison. Table 1 provides a summary of the number of features computed for each modality. However, the smaller frame used in our case (20 s) avoided the computation of some of the features in this extensive list. This happened for a total of 39 features, which were discarded for all samples. The concrete features used are as follows:

- **ECG-based features**. Among the ECG signals, the heart rate and heart rate variability features are the most commonly used. The QRS complexes and R-peaks of ECG signals were detected through the Pan-Tompkins QRS detection algorithm. Then, the Augsburg Biosignal Toolbox (AUBT) [21] was used to extract features from each part of the PQRST complexes. A total of 84 statistical features were extracted, consisting of the maxima, minima, mean, median, standard deviation and range from the raw signal and the derivative of PQ, QS and ST complexes, the number of intervals with latency greater than 50 ms from HRV, the Power Spectral Density (PSD) from HRV between the intervals [0 , 0.2], [0.2 , 0.4], [0.4 , 0.6] and [0.6 , 0.8], and the maxima, minima, mean, median, standard deviation and range from the HRV histogram.
- **EMG-based features**. Using AUBT [21], 21 statistical features were extracted from the available EMG signals. The computed features consisted of the mean, median, standard deviation, minima, maxima, and the number of times per time unit that the signal reached the minima and the maxima. These features were extracted from the raw EMG signal, its first derivative, and its second derivative.
- **EEG-based average PSD features**. In this case, the logarithm of the average PSD was computed from each of the 4-8 Hz (theta - θ), 8-10 Hz (low alpha - $\tilde{\alpha}$), 8-13 Hz (alpha - α), 13-30 Hz (beta - β), and 30-60 Hz (gamma - γ) frequency bands from each of the 14 channels of EEG signals. They were computed using Welch's estimate of spectral power by averaging across the components belonging to the frequency band, resulting to a total of 70 features (5 frequency bands' PSDs $\times$ 14 channels).
- **EEG-based Spectral features**. The methods described by Monge-Álvarez *et al.* [13] were applied to extract the following five spectral features: the Spectral Bandwidth (SB), the Spectral Crest Factor (SCF), the Spectral Flatness (SF), the Spectral Rolloff (SRO), and the Ratio f50 vs f90 (R5090), which were computed for each of the θ, α, β, and γ bands of each of the 14 channels of the EEG signals, resulting to a total of 280 features (5 features $\times$ 4 frequency bands $\times$ 14 channels).
- **EEG-based MFCC features**. Mel Frequency Cepstral Coefficients (MFCCs) were computed for each channel of the EEG signal using 18 filterbanks, leading to 12 cepstral coefficients per channel, as proposed by Piciucco *et al.* [14]. This yielded 168 features (12 cepstral coefficients $\times$ 14 channels).

3.3 Classification using Hidden Markov Models

In the original publication [1], for each classification algorithm tested, one model was trained for valence and another for arousal, by extracting the feature vectors described in section 3.2 from last 30 s of each signal and using the available positive/negative labels. In this work, the feature vectors were converted into time series, according to the following procedure: We first applied the same pre-processing described in section 3.1. Then, we split the signal using a moving frame of 20 s, with a 10 s overlap. Finally, we built a time series by computing a predefined set of features $\mathcal{F}$ for each time frame, with $\mathcal{F}$ a subset of the features presented in section 3.2. This yielded a time series for each signal whose dimension was $f \times t$, with f the number of features in $\mathcal{F}$ and t the number of time frames.

Once the original signals were converted into time series, arousal and valence were used as the target emotion labels and two HMM models were trained for each modality (two for arousal and two for valence). The first one was trained with all time series corresponding to samples labeled as positive, whereas all time series associated with negative samples were used to build the second model. It must be noted that in the case of arousal, positive samples refer to high arousal and negative samples to low arousal. To classify an unseen unlabeled sample, we first translated it into a time series representation, using the same method as for the training data, and then we computed the probability that the time series belonged to each of the two trained models (positive and negative) for each label (valence and arousal). The model providing the highest probability value determined the class returned as the classification result.

Results strongly depend on the set of features $\mathcal{F}$, which needs to be sufficiently small so that the classification problem can be computationally handled. Reported results in section 4 correspond to the best combination of features obtained by using the Sequential Forward Selection (SFS) algorithm [12, 22]. Starting from an empty set of features $\mathcal{F}$, the yet unused feature that yielded the best results was iteratively added to $\mathcal{F}$ until performance started decreasing, according to the F1-score of a leave-one-out classification scheme on the whole 114 samples set reported in [1] (19 subjects $\times$ 2 horses $\times$ 3 activities). This means that feature selection has used all available data and hence results reported for the HMM models should be considered as an upper limit on this dataset. It is also worth noting that a relatively small number of tests (below 3% of the total samples) had to be discarded because of numerical stability problems when building the HMM models.

When using other classifiers, such as the ones used in [1], the time variations of the signals were taken into account by using frequency-based features. One major expected advantage of HMMs, compared to those approaches, is their inherent ability to seamlessly learn the meaning of changes in the input signal along the time axis. Given that the same features as in the original work were used and that some of these were frequency based, the proposed HMM models are expected to be able to naturally capture frequency variations of the signals and boost their accuracy.

HMMs have a strong statistical foundation and are able to seamlessly handle variable length signals. In our case, the use of HMM models is supported by the assumption that the probability of a feature vector depends on the valence and arousal values. If this is the case, the construction of independent models for positive and negative arousal and valence should facilitate the identification of the arousal or valence state in terms of what model fits the sequence of samples better.

4 EXPERIMENTAL RESULTS

To assess the performance of the proposed method, the experiment described above was repeated for each of the feature sets in Table 1. As the problem is highly unbalanced (for valence, 87.7% of the samples were positive and only 12.3% were negative, whereas for arousal the distribution was 70.2% and 29.8% respectively), both classification accuracy and the F1-score were used to assess the predictive power of our alternative classification mechanism, as the accuracy metric can be biased towards the majority class. Classification accuracy is defined as the number of correct predictions divided by the total number of predictions; and the F1-score is defined as the harmonic mean between precision and recall, which can be mathematically expressed as:

$$\text{F1-score} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{1}$$

In this formula, *Precision* is the number of correct positive predictions divided by the total number of positive predictions; and *Recall* is the number of correct positive predictions divided by the number of positive samples in the test data.

A summary of the experimental results of the proposed approach (HMM) in terms of accuracy and F1-score is provided in Tables 2 and 3 for valence and arousal, respectively, in comparison to the ones reported in [1] for a series of classification algorithms, namely k-Nearest Neighbour (kNN) for k = 1, 3, 5, Linear Support Vector Machines (LSVM), Support Vector Machines using a Radial Basis Function kernel (SVM-RBF), Decision Trees (DT), and Linear Discriminant Analysis (LDA).

For each feature set of physiological signals, the best HMM results are shown in the first 3 numeric columns of Tables 2 and 3. These columns include the accuracy, the F1-score, and the number of features finally used according to the SFS algorithm. In addition, the last three columns correspond to the best results reported in [1], and the classifier that led to them.

Results for the proposed HMM approach are consistent with those reported in [1] on the single modality case, with regard to the features that offer the highest prediction power. The features that provide the best performance are EEG-based features and, among them, the EEG-based MFCC features offer the best results. For the prediction of valence, the MFCC features extracted from the frequency band [4-40 Hz] offered the highest recognition rate, with an accuracy as high as 0.987 and a F1-score of **0.940**. For arousal, the recognition rate obtained with MFCC features extracted from the frequency band [4-30 Hz] was the highest, with an accuracy of 0.818 and a F1-score as high as **0.783**.

The use of the proposed HMM approach led to consistent improvements in the arousal and valence dimensions for MFCC features across all four different frequency bands that were examined. Considerable gains were obtained in all these cases, as it can be observed in the last 4 rows of Tables 2 and 3. Nevertheless, performance decreased for the prediction of valence when using ECG, EEG-based spectral or PSD features, whereas results for arousal showed improvements in all features except for EEG-based spectral features. Nevertheless, the F1-score obtained for all cases other than the MFCC features was quite poor, always below 0.62, with the proposed approach achieving a considerably higher overall best performance compared to [1], increasing the highest F1-score for valence to 0.940 from 0.782, and for arousal to 0.783 from 0.654, as shown in Tables 2 and 3.

The consistency of the SFS algorithm for feature selection was also analysed by studying the performance variations in terms of the F1-score as the number of features is increased. Figures 1 and 2 depict a plot for the best performing cases (EEG-MFCC features for different frequency bands), for valence and arousal respectively. As expected, the performance initially grows as more features are added, but at some point the effect of the curse of dimensionality interferes with the amount of information added by the new feature and the performance starts to decrease. It is important to note that, even if the number of features was the same, the set of actual features selected was different for valence and arousal.

During the experimental evaluation, the number of hidden states in the HMM models was also varied in order to choose the one that yielded the best results. However, the experimental results showed that the highest F1-score was always obtained for a single state model. This was expected, as the activity performed was constant across any given sample and there was no apparent reason that justified a different signal evolution at different periods of time.

To summarise, the proposed HMM models have significantly boosted the performance of the models when using EEG-based MFCC features and this improvement holds across both the valence and arousal dimensions, as well as the four examined frequency bands.

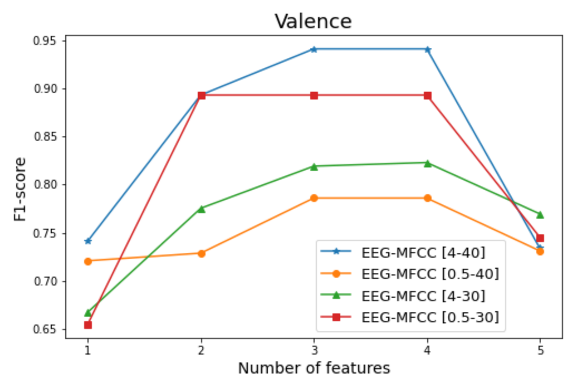

Figure 1: F1-score as a function of the number of features used to build the valence model.

5 CONCLUSION

The main aim of this work was to evaluate the effectiveness of HMMs based on physiological signals for emotion recognition. To this end, we used the physiological signals and the associated emotion labels obtained through a data collection experiment performed outdoors [1], based on human-horse interaction. The raw data included ECG, EEG, and EMG signals, and authors also provided the results of some initial experiments at predicting valence and arousal as a performance baseline.

In this work, we aimed to improve their baseline results, not only by using a different classification scheme, but also by selecting some specific features. In our approach, we used HMM models to generate two models for each of the valence and arousal dimensions. The first one was trained with samples with positive labels, and the second by using samples with negative labels. During validation,

Table 2: Classification performance for valence, in term of accuracy and F1-score, for each set of features

Features	HMM results			Results from [1]		
	# features	Accuracy	F1-score	Accuracy	F1-score	Classifier
ECG	1	0.674	0.508	0.842	0.580	LSVM
EMG	1	0.863	0.587	0.868	0.523	LDA
EEG-PSDavg	5	0.798	0.658	0.903	**0.782**	1NN
EEG-Spectral	1	0.948	0.653	0.877	0.715	1NN
EEG-MFCC [4-40]	3	0.987	**0.940**	0.868	0.684	DT
EEG-MFCC [0.5-40]	3	0.948	0.786	0.868	0.684	LSVM
EEG-MFCC [4-30]	4	0.961	0.822	0.807	0.552	LSVM
EEG-MFCC [0.5-30]	2	0.974	0.893	0.850	0.682	DT

Table 3: Classification performance for arousal, in term of accuracy and F1-score, for each set of features

Features	HMM results			Results from [1]		
	# features	Accuracy	F1-score	Accuracy	F1-score	Classifier
ECG	1	0.639	0.593	0.561	0.446	1NN
EMG	2	0.579	0.554	0.657	0.507	5NN
EEG-PSDavg	4	0.623	0.619	0.631	0.567	DT
EEG-Spectral	1	0.558	0.558	0.701	0.643	DT
EEG-MFCC [4-40]	4	0.805	0.770	0.701	0.637	LSVM
EEG-MFCC [0.5-40]	4	0.779	0.745	0.701	**0.654**	LDA
EEG-MFCC [4-30]	4	0.818	**0.783**	0.605	0.558	LDA
EEG-MFCC [0.5-30]	3	0.779	0.759	0.666	0.619	LDA

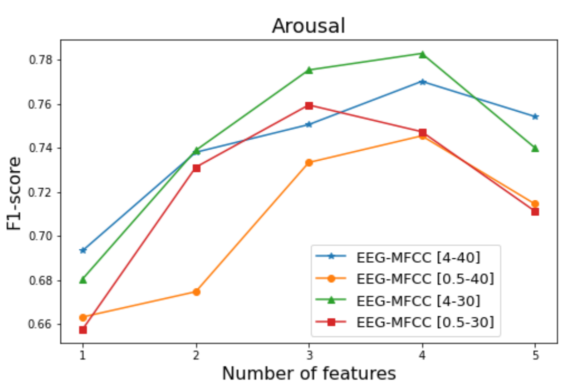

Figure 2: F1-score as a function of the number of features used to build the arousal model.

previously unseen samples were fed into both models, and the predicted class was the one leading to the highest score.

Experimental results showed that using the HMMs for emotion recognition of physiological signals can significantly improve recognition rates. Using a leave-one-out cross-validation procedure we have demonstrated the high accuracy of the proposed approach, especially when using EEG-based MFCC features. An F1-score of 0.941 was reached for valence using MFCC features on the [4-40 Hz] frequency band, whereas an F1-score of 0.783 was reached for arousal using the [4-30 Hz] frequency band. This is in contrast to the lower F1-scores of 0.782 and 0.654 reported in the original publication for valence and arousal, respectively. Consequently, it is evident that the proposed approach has significantly boosted the performance of the models when using the EEG-based MFCC features.

The results reported are in support of using the two-model HMM scheme to predict valence and arousal based on physiological signals. However, we shall also point out some limitations that need to be further explored. The first one relates to the statistical significance of the results that has not been tested in this work. Classification results obtained in [1] have been extracted from the original paper, and a significance analysis would imply a series of repeated experiments involving all methods compared. The second one concerns the feature selection process and the generalisation of the results when the set of best performing features is applied to new unseen samples. Whether these features remain constant along different datasets is a key aspect that should be further studied.

In reference to possible improvements, we have only used a single modality in all cases. However, we believe the results achieved can be enhanced by simultaneously considering features coming from different modalities and hence it is our intention to explore multi-modality fusion approaches in the context of this problem. Another line of improvement consists of validating our conclusions by replicating the experiments on other existing datasets and trying other classifiers that are not based on the analysis of the time sequence, to determine whether our achievements are related to the classifier or to the features we have used.

ACKNOWLEDGMENTS

We thank reviewers for their helpful comments and suggestions. This paper is part of projects PGC2018-096463-B-I00, financed by MCIN/AEI/10.13039/501100011033 and FEDER Una manera de hacer Europa; and AICO/2021/019, from Valencian Regional Government (Spain). Yuyan Wu's research is also funded by the Valencian Regional Government through grant ACIF/2021/439.

REFERENCES

[1] Turke Althobaiti, Stamos Katsigiannis, Daune West, and Naeem Ramzan. 2019. Examining human-horse interaction by means of affect recognition via physiological signals. *IEEE Access* 7 (2019), 77857–77867.

[2] Pablo Arnau-González, Stamos Katsigiannis, Naeem Ramzan, Debbie Tolson, and Miguel Arevalillo-Herráez. 2017. ES1D: A Deep Network for EEG-Based Subject Identification. In *17th IEEE International Conference on Bioinformatics and Bioengineering, BIBE 2017, Washington, DC, USA, October 23-25, 2017*. IEEE Computer Society, 81–85. https://doi.org/10.1109/BIBE.2017.00-74

[3] Pablo Arnau-González, Miguel Arevalillo-Herráez, and Naeem Ramzan. 2017. Fusing highly dimensional energy and connectivity features to identify affective states from EEG signals. *Neurocomputing* 244 (2017), 81–89. https://doi.org/10.1016/j.neucom.2017.03.027

[4] Pablo Arnau-González, Stamos Katsigiannis, Miguel Arevalillo-Herráez, and Naeem Ramzan. 2021. BED: A New Data Set for EEG-Based Biometrics. *IEEE Internet of Things Journal* 8, 15 (2021), 12219–12230. https://doi.org/10.1109/JIOT.2021.3061727

[5] Badcock, Nicholas A and Mousikou, Petroula and Mahajan, Yatin and De Lissa, Peter and Thie, Johnson and McArthur, Genevieve. 2013. Validation of the Emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs. *PeerJ* 1 (2013), e38.

[6] Nima Bigdely-Shamlo, Tim Mullen, Christian Kothe, Kyung-Min Su, and Kay A Robbins. 2015. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics* 9 (2015), 16.

[7] Burns, Adrian and Greene, Barry R and McGrath, Michael J and O'Shea, Terrance J and Kuris, Benjamin and Ayer, Steven M and Stroiescu, Florin and Cionca, Victor. 2010. SHIMMER™–A wireless sensor platform for noninvasive biomedical research. *IEEE Sensors Journal* 10, 9 (2010), 1527–1534.

[8] Chaurasiya, Rahul Kumar and Shukla, Shourya and Sahu, Tirath Prasad. 2019. A Sequential Study of Emotions through EEG using HMM. In *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*. IEEE, 104–109.

[9] Arnaud Delorme and Scott Makeig. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* 134, 1 (2004), 9–21.

[10] J. Healey and R. Picard. 1998. Digital processing of affective signals. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*, Vol. 6. 3749–3752.

[11] Katsigiannis, Stamos and Ramzan, Naeem. 2018. DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. *IEEE Journal of Biomedical and Health Informatics* 22, 1 (2018), 98–107.

[12] Pat Langley and Stephanie Sage. 1994. Induction of selective Bayesian classifiers. In *Uncertainty Proceedings 1994*. Elsevier, 399–406.

[13] Monge-Álvarez, Jesús and Hoyos-Barceló, Carlos and San-José-Revuelta, Luis Miguel and Casaseca-de-la-Higuera, Pablo. 2018. A machine hearing system for robust cough detection based on a high-level representation of band-specific audio features. *IEEE Transactions on Biomedical Engineering* 66, 8 (2018), 2319–2330.

[14] Piciucco, Emanuela and Maiorana, Emanuele and Falzon, Owen and Camilleri, Kenneth P and Campisi, Patrizio. 2017. Steady-state visual evoked potentials for EEG-based biometric identification. In *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 1–5.

[15] Russell, James A. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39, 6 (1980), 1161.

[16] Ana Serrano-Mamolar, Miguel Arevalillo-Herráez, Guillermo Chicote-Huete, and Jesus G. Boticario. 2021. An Intra-Subject Approach Based on the Application of HMM to Predict Concentration in Educational Contexts from Nonintrusive Physiological Signals in Real-World Situations. *Sensors* 21, 5 (2021). https://doi.org/10.3390/s21051777

[17] Shu, Lin and Xie, Jinyan and Yang, Mingyue and Li, Ziyi and Li, Zhenqi and Liao, Dan and Xu, Xiangmin and Yang, Xinyi. 2018. A review of emotion recognition using physiological signals. *Sensors* 18, 7 (2018), 2074.

[18] Soleymani, Mohammad and Lichtenauer, Jeroen and Pun, Thierry and Pantic, Maja. 2011. A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing* 3, 1 (2011), 42–55.

[19] Tao, Jianhua and Tan, Tieniu. 2005. Affective computing: A review. In *International Conference on Affective Computing and Intelligent Interaction*. Springer, 981–995.

[20] Torres-Valencia, Cristian A and Garcia-Arias, Hernan F and Lopez, Mauricio A Alvarez and Orozco-Gutiérrez, Alvaro A. 2014. Comparative analysis of physiological signals and electroencephalogram (EEG) for multimodal emotion recognition using generative models. In *2014 XIX Symposium on Image, Signal Processing and Artificial Vision*. IEEE, 1–5.

[21] Wagner, Johannes. 2005. Augsburg Biosignal Toolbox (AUBT). *University of Augsburg* (2005).

[22] Cen Wan. 2019. Feature selection paradigms. In *Hierarchical Feature Selection for Knowledge Discovery*. Springer, 17–23.