Biomedical signal processing methods for neuromarketing: A comparative study

Suzan SABAN[1], Eda DAĞDEVİR *[2]

[1] Department of Electrical and Electronics Engineering, Graduate School, Kayseri University, Turkey
[2] Department of Electronic and Automation, Vocational School of Technical Science, Kayseri University, Turkey

*(edadagdevir@kayseri.edu.tr) Email of the corresponding author

Abstract – Neuromarketing involves the integration of neuropsychology into marketing research, focusing on analyzing consumer sensory-motor actions, including cognitive and emotional responses to marketing stimuli, using advanced technologies. It represents one of the latest strategies in marketing research and has the potential to shape the future of the field. Numerous studies have already been conducted in this domain to enhance research outcomes. Nevertheless, the literature indicates that there are still opportunities for further advancements and improvements. A literature review was conducted in this article to explore the availability of an openly accessible dataset widely utilized by researchers in the field of neuromarketing. We examined the signal preprocessing, feature selection, feature extraction, and classification methods employed in studies utilizing the dataset created by Yadava et al.

Keywords – Neuromarketing, EEG, Signal Processing, Classification, Data Mining

I. INTRODUCTION

Yadava et al. obtained EEG signals from 25 male participants using a 14-channel Emotiv EPOC+ device. EEG sensors were placed on the participants' heads, and they were instructed to view shopping items as shown in Figure 1, where a user was viewing an item on a computer screen. During the recording phase, EEG signals were simultaneously recorded while the user viewed a product. After the viewing process, the user was asked to indicate their preference for the product, classifying it as either liking or disliking. Subsequently, the signals underwent specific signal preprocessing and feature extraction steps, followed by classification using HMM, SVM, RF, and NN classifiers [1].

Amin et al. have described their proposed method using a block diagram in Figure 2. The initial step of their methodology involved collecting the EEG dataset. Subsequently, signal preprocessing was conducted, and the features were divided into training and testing sets for classification purposes.

A model was developed based on the training set, and class labels were assigned to the testing set. Performance metrics were then calculated using the actual and predicted class labels. In their study, they employed the averaging method for signal preprocessing. Feature extraction was performed using the DWT method. They utilized KNN, DA, NB, DT, SVM, and RF classifiers, achieving the highest accuracy rate of 93% with the DT classifier [2].

Fig. 1. System setup where a user watching consumer products during experiment [1]

In their study, Aldayel et al. extracted true hidden preferences from EEG signals by employing two methods to determine preference labels. They utilized a self-assessment questionnaire to identify subjective preferences and employed valence indicators to determine objective preferences [3].

Alimardani et al., in their study, compared the accuracy rates obtained from machine learning models and deep learning models. In the machine learning model, they utilized SVM, RF, and Logistic Regression classifiers and achieved results based on ensemble classifiers. On the other hand, in the deep learning model, they employed the convolutional neural network (CNN) method, with the Stochastic Gradient Descent (SGD) optimizer [4].
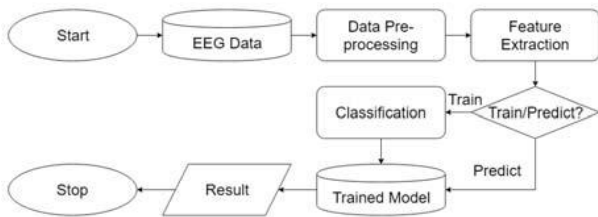

Fig. 2 A block diagram of the proposed method for consumer behaviour analysis [2].

In the study conducted by Al-Nafjan, the significance of feature selection is emphasized. Classification was performed by applying feature selection using PCA, Relief, mRMR, and RFE methods. The accuracy rates of the classifier algorithms used for each feature selection method were compared. Ultimately, it was determined that feature selection improved the performance of all classifiers [5].

When comparing all classifiers, the best results were obtained using mRMR and ReliefF feature selection methods. The KNN, LDA, and SVM classifiers achieved accuracy rates of 94%, 95%, and 97%, respectively, while DNN and RF attained the highest accuracies of 99% and 100%, respectively. Similar outcomes were achieved using the feature importance method, except for DNN, where all classifiers, except DNN, yielded better accuracy rates when the number of selected features was set to 10. However, DNN performed better with 30 features, achieving an accuracy rate of 98%. This can be attributed to the fact that DNN works more effectively with larger datasets. On the other hand, PCA did not yield improved accuracy compared to other feature selection methods [5].

The primary objective in the study conducted by Kumar et al. was to enhance the rating prediction performance of the system using a multimodal framework. The proposed approaches are illustrated in Figure 3. The ABC optimization technique was employed to optimize the ratings obtained from three data sources: EEG signals, sentiment scores from customer reviews on the product's brand, and the product itself. The Random Forest regression technique was utilized to compute the multimodal ratings [6].
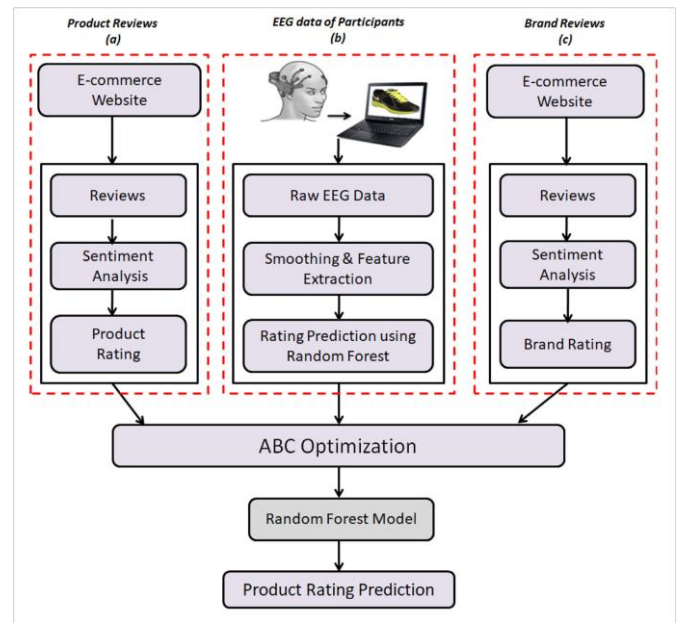

Fig. 3. Proposed multimodal framework for rating prediction using EEG signals and customer reviews [6]

In their study, Aldayel and Ykhlef et al. examined the probability of determining two affective levels, namely "liking" and "disliking," using different feature combinations of EEG indices and various approaches of feature extraction and classification algorithms [7].

In their study, they defined four indicators based on EEG signals as preference indicators: approach-withdrawal (AW) indicator, valence indicator, choice indicator, and effort indicator. These indicators were identified to assist in understanding consumers' responses to products [7].

II. MATERIALS AND METHOD

In this study, the place of biomedical signal processing in neuromarketing applications were presented comparatively in Yadava et. al. dataset. A systematic review of studies using this dataset was conducted. In addition, Figure 3 showed that the signal processing methods in these studies which

used the dataset were examined as preprocessing, feature extraction, feature selection, and classification, respectively.
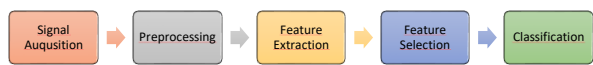

Fig. 4 Signal Processing Stages

A. *Dataset Description*

The dataset provided by Yadava et al [1]. included EEG recordings obtained from 25 participants, aged between 18 and 25. During the recordings, participants viewed various product images for a duration of 4 seconds, while their EEG signals were captured using a 14-channel Emotiv Epoc+ device. The stimuli consisted of 14 distinct product categories, such as glasses, bags, shirts, pens, and others, with each category comprising three different images. Every image is displayed for 4 seconds on the computer screen. This resulted in a total of 42 unique products. Subsequently, participants were required to express their preference or aversion towards each presented product. As a result, 1050 epochs of EEG signals were generated, of which 1045 epochs were made publicly available. The EEG signals were recorded at a sampling rate of 128 Hz.

The dataset was obtained by collecting features from 14 specific electrodes positioned at "AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4" locations. To identify preferences, four electrodes (F3, F4, AF3, and AF4) relevant to the task were selected. The choice of these electrodes was based on preference mapping using brain regions, as described in a previous study [8]. Each participant was asked to provide feedback in the form of "like" or "dislike" while viewing a product, and their EEG data was recorded simultaneously. The user's preferred choice was documented after each picture presentation.

B. *Preprocessing Stage*

The purpose of data pre-processing is to reduce or even reject certain artifacts labeled in the recordings and control undesired disturbances. EEG recordings are highly sensitive to various sources of interference, making it challenging to discern their characteristics by analyzing EEG signals. There are numerous existing methods available to effectively address these interferences [18,19].

- **Savitzky-Golay filter**

The Savitzky-Golay (SG) filters find extensive usage in various fields, including elastography, EEG, magnetocardiogram signal processing, and particularly absorption spectroscopy, for the purposes of smoothing and differentiation. The underlying method relies on the following principle: a set of 2m+1 equidistant points (-m, … 0, …, +m), obtained through sampling, is assumed to represent samples of a p-degree polynomial corrupted by additive zero-mean measurement noise. The SG method estimates the values of the least-squares polynomial or its derivatives at the point $i = 0$. There is no necessity to repeatedly fit new polynomials for subsequent points, as this can be achieved automatically by applying a convolution with constant coefficients. Consequently, a continuous least-squares polynomial fitting is performed automatically on the samples [9].

- **ICA**

ICA is a statistical technique used to separate mixed signals into independent stationary sources. Extensive research has demonstrated the utility of ICA decompositions in identifying EEG features such as alpha waves and steady-state responses, facilitating artifact removal, and contributing to EEG analysis in general. While ICA methods have been extensively discussed in the literature, special attention is required for the SLICA approach in EEG classification. Standard ICA algorithms fail to converge when the input matrix has a defective rank, as is the case with SL-transformed data. This issue arises during the sphering step, where a decorrelated data matrix is generated to enhance convergence. Sphering involves first zero-centering the rows of the input matrix, followed by a linear transformation using the inverse of the principal square root of the covariance matrix. The outcome is a decorrelated data ensemble with a diagonal covariance matrix [10].

- **Averaging**

To obtain the average of a set of numerical values, one can calculate the sum of the values and divide it by the number of terms in the set. This calculation yields another value known as the arithmetic mean [5]. When developing the arithmetic mean (AM), given n numbers, with each number denoted by a_k (where $k = 1, 2, ..., n$), the arithmetic mean is

obtained by dividing the sum of the numbers by n, expressed as Eq(1).

$$AM = \frac{1}{n}\sum_{k=1}^{n} a_k = \frac{1}{n}(a_{1+}a_{2+\dots+}a_n) \qquad Eq(1)$$

- **Bandpass Filter**

The ideal bandpass filter is typically used to isolate a specific frequency range within a time series. However, in practical applications, it is not feasible to have an infinite-length dataset. Therefore, approximations are necessary. By utilizing projections, we can derive optimal approximations for time series representations that exhibit a unit root or exhibit stationarity around a trend. While there is one approximation that is specifically optimal for a particular time series representation, it can still be effective for standard macroeconomic time series. To demonstrate the utility of this approximation, we apply it to analyze changes in the Phillips curve and the money-inflation relationship before and after the 1960s. Surprisingly, we find that there is minimal change in the Phillips curve, but significant changes in the relationship between money growth and inflation. [12].

To mitigate the impact of noise, an initial step involved applying a bandpass filter ranging from 0.5 Hz to 40 Hz to the raw EEG signals. Subsequently, the focus shifted to extracting the significant frequency bands, namely delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), and gamma (above 30 Hz). The next stage entailed calculating the spectral energy associated with each of these frequency bands [4].

C. Feature Extraction

Feature extraction aims to identify and extract important and relevant information from EEG signals. In previous studies conducted with this dataset, the following feature extraction methods have been used.

- **PSD**

Feature extraction plays a crucial role in the advancement of BCI (Brain-Computer Interface) applications. Within neuromarketing research, one commonly employed approach for feature extraction is the power-spectral density (PSD) method, which analyzes the frequency of EEG signals. The PSD method facilitates the translation of data between the frequency domain and time domain. This conversion relies on the Fast Fourier

Transform, which computes both the discrete Fourier transform and its inverse. By utilizing the Welch-PSD method, we acquired EEG frequencies and utilized them as preference indicators to gauge valences. The Welch method divides each EEG signal into four distinct frequency bands: gamma (30–40 Hz), beta (13–30 Hz), alpha (8–13 Hz), and theta (4–8 Hz) [3].

- **Discrete wavelet transform (DWT)**

The DWT (Discrete Wavelet Transform) is a method used for time- frequency domain analysis, which breaks down signals into various coefficients. It can be characterized as a multi-resolution or multi-scale analysis, where each coefficient serves as a distinct representation of the underlying mind signals. This decomposition allows for the examination of signal components at different scales and resolutions, enabling detailed analysis of the frequency content of the original signal.

The convolution operation is a two-function multiplication process. Therefore, the DWT can be expressed using the following Equation (2) [7,17].

$$W(j,k) = \sum_{N=0}^{M-1} f(n).\Psi_{j,k}^*(n) \qquad Eq(2)$$

- **Statistical Features**

In the previous study we examined, four statistical features, namely Mean (M), Standard Deviation (SD), Energy (EN), and Root-Mean-Square (RMS), have been extracted using the feature vector (Ft) [1]. The features have been computed using (3)–(6).

$$M = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad Eq(3)$$

where x_i is the ith sample of the data sequence and n denotes total number of data points

$$SD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \qquad Eq(4)$$

where x is the mean of the sample and n denotes the number of items in the sample.

$$EN = D \sum_{i=0}^{n-1} x_i^2 \qquad Eq(5)$$

Where D is duration of signal, and x_i is discrete samples of the signal at regular intervals (0 to n-1).

$$RMS = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} \qquad Eq(6)$$

Where x_i is the ith item of the sequence and n is the number of items in the sequence [1].

D. Feature Selection

The objectives of feature selection are manifold, with the most important ones being: (1) to avoid overfitting and improve model performance, i.e., enhancing prediction accuracy in supervised classification and better cluster detection in clustering, (2) to provide faster and more cost-effective models, and (3) to gain deeper insights into the underlying processes that generate the data. Feature selection methods can generally be divided into filter and wrapper methods. Wrapper methods select features based on their interaction with an underlying model (classifier), while filter methods are independent of the model. Filters have the advantage of requiring less computational power compared to wrappers, making them more suitable for large datasets [20,21]. Several notable feature selection algorithms include below.

- **PCA**

PCA is a widely used unsupervised feature selection method for reducing dimensionality. It utilizes statistical techniques to convert a set of correlated measurements into a set of linearly independent principal components. The significance of PCA lies in its ability to reduce dimensionality without losing information, while considering the complexity of signal extraction and classification. By analyzing the time series data of EEG signals, PCA extracts meaningful signals. It compresses the EEG signals into unrelated components for signal preprocessing and feature selection, effectively reducing noise during signal separation. Consequently, the EEG signals are reconstructed after noise removal. When PCA identifies patterns in a signal, it can be visualized as a rotation of coordinate axes combined with the arrangement of time points. The principal components are the components that exhibit the highest variance [13].

- **ReliefF**

This method, known as a univariate technique, is an expansion of the Relief algorithm. It employs a subset of the total instances to modify the weights assigned to individual features. The adjustments are made based on how effectively the features can distinguish between two distinct classes [14].

To the best of our understanding, the Relief algorithm and its variations are the sole evaluation filter algorithms that have the capability to identify feature dependencies. Unlike other algorithms, they don't explore feature combinations directly; instead, they rely on the concept of nearest neighbors to compute feature statistics that indirectly capture interactions [15].

- **mRMR**

The minimal-Redundancy-Maximal-Relevance (mRMR) is the widely recognized technique that utilizes mutual information to assess the appropriateness of a feature subset. It has gained immense popularity for its ability to effectively characterize the suitability of feature subsets [14].

Peng, Long, and Ding (2005) introduced the mRMR filtering approach, which utilizes relevance and redundancy metrics to select features. This technique initially entails picking features that have the highest correlation with the class label and the lowest correlation with other features. Mutual information I is employed to calculate the statistical correlation, thereby ensuring that the joint distribution optimizes the following two criteria simultaneously:

(1) maximum relevance/dependence D between features and the class label $I(x_a; y)$, and

(2) minimum redundancy R of features $I(x_a; x_b)$, which enhances classification accuracy [16].

- **RFE**

RFE is a popular wrapper method employed for feature selection. Its functionality involves iterative reduction of features and building a classification model using the remaining features. During each iteration, the algorithm reorganizes the ranking of potential feature subsets based on their classification accuracies. The classification accuracy serves as a criterion to determine the combination of features that provide the highest contribution to predicting class labels. In the RF-RFE method, the JRF classifier is utilized to assess the significance of features and construct a classifier with notable importance scores [5].

E. Classification

In our literature review conducted for the dataset created by Yadava et al. for neuro-marketing research, we found that support vector machine (SVM), Random Forest (RF), K- Nearest Neighbors (KNN), Deferred Acceptance (DA), Naïve Bayes (NB), Hidden Markov Model (HMM), Decision Tree (DT), and Latent Dirichlet Allocation (LDA) classifiers were utilized. Furthermore, machine learning and deep learning techniques (convolutional neural network (CNN)) were compared in terms of their accuracy rates. In the

study conducted by Kumar et al., the Artificial Bee colony algorithm was employed [1,2,4,5,6].

The signal processing stages of the articles using the dataset are summarized in the Table 1.

Table 1. Reviewed papers that used Yadava et. al. dataset

Paper	Preprocessing	Feature Extraction	Feature Selection	Classification
Yadava et al.	S-Golay	DWT, Statistical Features	-	HMM, SVM, RF, NN
Kumar et al.	S-Golay	DWT, Statistical Features	-	RF
Amin et al.	Averaging	DWT	-	KNN, DA, NB, DT, SVM, RF
Aldayel et al.	ICA, S-Golay, Bandpass, Averaging	PSD, Statistical Features	-	KNN, SVM, RF
Alimardani et al.	Bandpass	-	-	CNN
Ykhlef et al.	ICA, S-Golay, Averaging, Bandpass	PSD, DWT, Statistical Features	-	DNN, KNN, SVM, RF
Al-Nafjan	ICA, Averaging, Bandpass	PSD, DWT, Statistical Features	PCA, Relief F mRMR, RFE	LDA, SVM, RF, KNN, DNN

III. RESULTS

The publication years of the papers and the accuracy values obtained according to each classification method are given in Table 2. The results showed that, the dataset is highly innovative and has only been of interest to researchers for a few years. When the obtained accuracies are evaluated, it is seen that traditional methods lag behind deep learning methods. The effect of using the feature selection method on success is also shown in this study. Using less data is critical also timing cost.

Table 2. Accuracy for reviewed papers

Paper	Year	Accuracy (%)	
Yadava et al.	2017	HMM	70
Kumar et al.	2017	RF	71
Amin et al.	2020	KNN	77
		DA	60
		NB	76
		DT	93
		SVM	87
		RF	54
Aldayel et al.	2021	KNN	72
		SVM	71
		RF	83
Alimardani et al.	2021	CNN	52
Ykhlef et al.	2021	DNN	93
		KNN	78
		SVM	86
		RF	93
Al-Nafjan	2022	LDA	70
		SVM	86
		RF	93
		KNN	78
		DNN	93

Figure 5 shows the highest classification performance achieved in each study. Accordingly, the lowest and highest accuracy value obtained is 52% and 93% respectively.
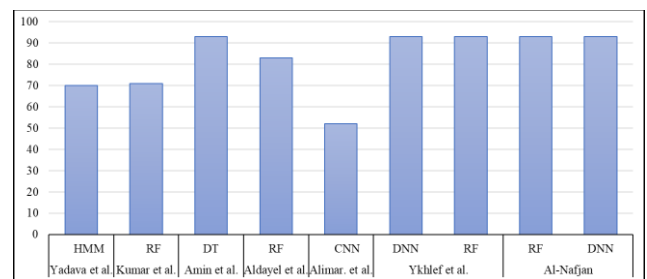

Fig. 5 Best accuracy for each reviewed paper

IV. DISCUSSION

According to the results obtained, the Yadava et. al. dataset is quite suitable for novel studies. By the articles compared, especially deep learning methods have been shown to be overwhelmingly superior to traditional methods. Another finding is important for the classification performance of the feature selection stage. No evaluation of timing cost was found in any of the studies using this dataset. This is a huge gap in the literature. In applications that are likely to be adapted to online systems such as neuromarketing, timing cost should be evaluated as well as performance.

V. CONCLUSION

In this study, articles using the Yadava et. al. dataset was evaluated and compared. results, advantages and disadvantages are presented in the discussion section. In particular, very guiding findings have been obtained for researchers who will use the Yadava et. al. data set. Within the scope of this study, WOS and Scopus databases were scanned. Although a large body of articles has been reached, a more comprehensive comparison can be made by searching other databases.

REFERENCES

[1] Yadava, M., Kumar, P., Saini, R., Roy, P. P., & Prosad Dogra, D. (2017). Analysis of EEG signals and its application to neuromarketing. *Multimedia Tools and Applications*, *76*, 19087-19111.

[2] Amin, C. R., Hasin, M. F., Leon, T. S., Aurko, A. B., Tamanna, T., Rahman, M. A., & Parvez, M. Z. (2020, December). Consumer behavior analysis using EEG signals for neuromarketing application. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 2061-2066). IEEE.

[3] Aldayel, M., Ykhlef, M., & Al-Nafjan, A. (2021). Consumers' preference recognition based on brain–computer interfaces: advances, trends, and applications. *Arabian Journal for Science and Engineering*, *46*(9), 8983-8997.

[4] Alimardani, M., & Kaba, M. (2021, May). Deep learning for neuromarketing; classification of user preference using EEG signals. In *12th Augmented Human International Conference* (pp. 1-7).

[5] Al-Nafjan, A. (2022). Feature selection of EEG signals in neuromarketing. *PeerJ Computer Science*, *8*, e944.

[6] Kumar, S., Yadava, M., & Roy, P. P. (2019). Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction. *Information Fusion*, *52*, 41-52.

[7] Aldayel, M., Ykhlef, M., & Al-Nafjan, A. (2021). Recognition of consumer preference by analysis and classification EEG signals. *Frontiers in Human Neuroscience*, *14*, 604639.

[8] Aldayel, M., Ykhlef, M., & Al-Nafjan, A. (2020). Deep learning for EEG-based preference classification in neuromarketing. *Applied Sciences*, *10*(4), 1525.

[9] Hargittai, S. (2005, September). Savitzky-Golay least-squares polynomial filters in ECG signal processing. In *Computers in Cardiology, 2005* (pp. 763-766). IEEE.

[10] Carvalhaes, C. G., Perreau-Guimaraes, M., Grosenick, L., & Suppes, P. (2009, June). EEG classification by ICA source selection of Laplacian-filtered data. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (pp. 1003-1006). IEEE.

[11] binti Mustaffa, I., & bin Zaid, Z. (2017, November). Significance of averaging method signal denoising. In *2017 International Conference on Robotics, Automation and Sciences (ICORAS)* (pp. 1-6). IEEE.

[12] Furceri, D., & Karras, G. (2008). Business-cycle synchronization in the EMU. *Applied economics*, *40*(12), 1491-1501.

[13] Xie, Y., & Oniga, S. (2020). A review of processing methods and classification algorithm for EEG signal. *Carpathian Journal of Electronic and Computer Engineering*, *13*(1), 23-29.

[14] Jenke, R., Peer, A., & Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Transactions on Affective computing*, *5*(3), 327-339.

[15] Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: Introduction and review. *Journal of biomedical informatics*, *85*, 189-203.

[16] Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, *27*(8), 1226-1238.

[17] Chen, L. L., Zhao, Y., Zhang, J., & Zou, J. Z. (2015). Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning. *Expert Systems with Applications*, *42*(21), 7344-7355.

[18] Mateo, J., Torres, A. M., García, M. A., & Santos, J. L. (2016). Noise removal in electroencephalogram signals using an artificial neural network based on the simultaneous perturbation method. *Neural Computing and Applications*, *27*, 1941-1957.

[19] Parvez, M. Z., & Paul, M. (2015). Epileptic seizure detection by exploiting temporal correlation of electroencephalogram signals. *IET Signal Processing*, *9*(6), 467-475.

[20] R. Jenke, A. Peer and M. Buss, "Feature Extraction and Selection for Emotion Recognition from EEG," in *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 327-339, 1 July-Sept. 2014, doi: 10.1109/TAFFC.2014.2339834.

[21] Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *bioinformatics*, *23*(19), 2507-2517.