Sādhanā (2022) 47:281
https://doi.org/10.1007/s12046-022-02015-w

Optimal classification of N-back task EEG data by performing effective feature reduction

RAJESH PATEL*, K GIREESAN, R BASKARAN and N V CHANDRA SHEKAR

MEG Lab, SQUIDs Application Section, SDTD, Materials Science Group, Indira Gandhi Centre for Atomic Research, A CI of Homi Bhabha National Institute, Kalpakkam, Tamil Nadu 603102, India
e-mail: prajesh@igcar.gov.in, prajeshec@yahoo.com

MS received 18 June 2022; revised 14 September 2022; accepted 23 September 2022

Abstract. Many studies have been carried out related to the analysis of cognitive workload assessment using the N-back task. However, fixed analytic functions like time-frequency spectrum and wavelet-based approaches have been primarily used to analyze non-stationary EEG signals. Moreover, these approaches removed redundant information present in the input features by implementing the feature reduction approaches like correlation analysis and Principal Component Analysis, which are primarily based on the assumption of linearity in the input features. In the present work, we have analyzed multichannel EEG data for the N-back (0, 1, 2-back) task using a data-driven technique called multivariate empirical mode decomposition (MEMD). MEMD breaks down multichannel data into a fixed number of intrinsic mode functions (IMFs). Various features have been extracted from each IMF based on statistical parameters (variance, skewness, and kurtosis), spectral power (related to brain waves: delta, theta, alpha, beta, and gamma), and parameters based on time-series data (relative MEMD energy and zero-crossing rate). The effective feature reduction is obtained by kernel principal component analysis (kPCA). These new reduced transformed features are taken as input for training and testing different machine learning (ML) models viz K-nearest neighbor (KNN), Support Vector Machine (SVM), Multilayer Perceptron (MLP), and random forest. The best average classification accuracy of 97.34% could be achieved using KNN with kPCA transformed features (based on third-order polynomial kernel function). The proposed approach performs better in classifying the N-back EEG data than earlier techniques.

Keywords. Electroencephalogram; multivariate empirical mode decomposition; classification; features reduction; N-back task; kernel principal component analysis.

1. Introduction

A person's ability to process the received information and make use of it for correct decisions is critical in modern industries [1–3]. It is reported that one of the major causes of accidents (60–80%) in the plants is due to human errors [4–7]. Earlier studies also show the relation of human errors concerning the cognitive aspects of humans [8, 9], and hence it is essential to assess the cognitive workload of the plant's operator [10]. Electroencephalography (EEG) technique is regarded as one of the best techniques to study and analyze the dynamics of brain signals as it offers excellent temporal resolution and excellent spatial localization accuracy [11]. Many EEG studies were carried out to investigate cognitive workload by analyzing N-back tasks [12–19]. In most of the earlier studies, short-term Fourier transforms (STFT) and wavelet-based approaches have been used widely to analyze the EEG data. However, these approaches required some fixed basis functions such as sinusoids or wavelets for decomposition of the data[20–22]. The selection of a suitable basis function requires prior knowledge about the signal's morphology, which necessitates checking many such basis functions to find the optimum basis function [23, 24]; this can be avoided by employing data-driven approaches such as empirical mode decomposition (EMD) and its variant ensemble empirical mode decomposition (EEMD) [25]. It may, however, be noted that despite its success [26], EMD may not be suitable for analysis of multi-channel data, as EMD based decomposition of the data generates a variable number of intrinsic mode functions (IMFs) from one channel to another, and IMFs also may suffer from mode mixing issues [27].

In the present work, multi-channel EEG data were decomposed using multivariate empirical mode decomposition (MEMD), which results in a fixed number of IMFs for every channel [28, 29]. MEMD reduces the issues of mode mixing and mode misalignment within IMFs by processing all the input signals altogether in multi-dimensional space [30, 31]. Many features were extracted

*For correspondence
Published online: 19 December 2022

from each IMF based on spectral power, time series, and statistical parameters. In the machine learning field, a large number of features may result in model over-fitting and poor generalization [32, 33]. Hence, feature reduction is commonly performed to resolve these issues [34, 35]. Principal component analysis (PCA) is widely used for feature reduction, data analysis, and denoising the data [36–38]. In PCA, nearly the same information present in the raw data is expressed using reduced features by calculating correlation (linear relation) among the features.

The operation of PCA is based on the assumption of linearity [39], and hence linear PCA is unable to perform effective feature reduction . The kernel PCA (kPCA) is proposed against linear PCA to resolve the above issue. In kPCA, the covariance of nonlinear features is taken as the kernel function in place of the linear covariance; this allows kPCA to decode nonlinearity present in the features and thereby achieve a better feature reduction compared to linear PCA. A few studies have shown that kPCA is better than linear PCA by effectively extracting the breathing signal from the Electrocardiogram (ECG) data [39] and removing the eye-blink artifact from the EEG data [40].

Figure 1 shows the steps followed in the present study.

2. MEMD algorithm and kPCA: a brief introduction and notation

MEMD: MEMD is an extension of EMD and allows the decomposition of multi-channel data. The multivariate signal is decomposed into intrinsic mode functions (IMFs), containing multiple oscillatory modes. In MEMD, the sifting process is used to obtain successive IMFs, which have to satisfy the following two conditions:

(1) The mean of local envelopes of maxima and minima must always be zero.
(2) The number of zero crossings and extrema points should either be equal or at most differ by one.

In MEMD, multi-dimensional signals are projected along with different directions in n-dimensional space, and then many envelopes are obtained from those projections. These envelopes are averaged to calculate the corresponding local means. The resultant local means are subtracted from the original signals to obtain oscillatory modes, which are further subjected to an identical procedure until IMFs are derived. Consider a sequence of m-dimensional vectors $s(t) = s_1(t), s_2(t), \ldots, s_m(t))$, representing a multivariate signal with m channels and $x^v = x_1^v, x_2^v, ..., x_m^v$ denoting a set $v = 1, 2, ..., V$ direction vectors along the directions given by angles $\theta_v = \theta_{v1}, \theta_{v2}, \ldots, \theta_{vm}$ in m space. The details of the MEMD algorithms are as follows:

(i) Select an appropriate point set for sampling a (m-1) sphere. Calculate the corresponding projections, denoted by $w_{\theta_v(t)}$ of input signal $s(t)$ along the direction vectors x^v for $v = 1, 2, ..., V$
(ii) Find the time instants $t_{\theta v}^i$ corresponding to the maxima of the set of projected signals for $v = 1, 2, ..., V$.
(iii) Interpolate $[t_{\theta v}^i, s(t_{\theta v}^i)]$ to obtain the multivariate envelope curves $e_{\theta v}(t)$ for $v = 1, 2, ..., V$.
(iv) Obtain the mean of the V multidimensional envelopes

$$m(t) = 1/V \sum_{v=1}^{V} e_{\theta v}(t) \tag{1}$$

(v) Substract the mean from signal to get detail, $d(t) = s(t) - m(t)$. If $d(t)$ satisfies the conditions of the stopping criteria for a multivariate IMF, use the above procedure to $s(t) - d(t)$, else repeat to obtain $d(t)$.

Finally, MEMD decompose a m-variate signal s (t) as

$$s(t) = \sum_{p=1}^{l} c_p(t) + r(t) \tag{2}$$

where the m-variate IMFs, $\{c_m\}_{p=1}^{l}$ contain scale-aligned intrinsic joint rotational modes, and $r(t)$ denotes the residual.

Kernel PCA: Linear PCA is one of the popular approaches for feature reduction and data exploration. However, PCA is based on an assumption of linearity among the input features. Kernel PCA overcomes this issue and reveals nonlinearity present in the input features by projecting the input features into the higher dimension and then performing PCA to obtain the new reduced transformed features. For the input data $\{f_i\}_{i=1}^{n}$, the covariance matrix for kPCA can be written as:

$$Cov_{mat} = \frac{1}{n} \sum_{j=1}^{n} \phi(f_j)\phi(f_j)^T \tag{3}$$

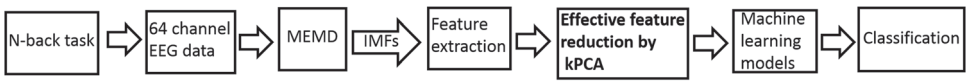

Figure 1. The proposed approach for the classification of the N-back task.

where n is the number of data-points and ϕ is a nonlinear function. The corresponding eigen values can be obtained as

$$\lambda\vartheta = Cov_{mat}\vartheta \tag{4}$$

where λ represents eigen values and ϑ refers to the eigen vector in feature space In the present study, two kernel functions have been used for implementing kPCA.

(i) Polynomial kernel function can be expressed as:

$$Ker_{Poly.}(f_i, f_j) = (f_i^T f_j + t)^d \tag{5}$$

where t refers to the intercept and d refers to the degree of the polynomial.

(ii) Radial basis function (RBF) can be described as:

$$Ker_{RBF}(f_i, f_j) = exp\left(\frac{||f_i - f_j)||^2}{2\sigma^2}\right) \tag{6}$$

where σ^2 refers to the variance of the Gaussian kernel.

3. Materials and methods

3.1 *Experimental design and protocol*

A total of 20 (8 female) subjects participated in N-back EEG experiments. All the subjects were informed about the procedure and motive of the present study. Written consent for participating in this study was obtained from each participant, and they had no history of any issue related to cognitive ability.

The age of the subjects was 28±6 years. In earlier N-back studies, letters/numbers were presented as objects to perform cognitive workload tasks [16, 17]. The present study uses shapes instead of letters/numbers to generalize the stimulus for performing the N-back task. In this work, six different sequences of visual stimuli were presented to the subjects. The sequences consisted of two distinct objects (square and circle) located at the corner of the computer screen. During the stimulus presentation, the spatial location of objects varied randomly on the screen. The subject was instructed to respond by pressing a specific key on the keypad if the spatial locations of objects presented on the current screen were identical to those in the stimulus presented "n" step immediately preceding the current stimulus (n = 1,2). Figure 2 shows an illustration of a typical sequence of stimuli presented for one back experiment, where the exact match between the spatial locations of objects in the second and third stimuli of the sequence is required to be noticed by the subject, and the subject is required to give a response (a click on the response keypad). In the case of 0-back experiments, a particular pattern was given as a fixed target, and subjects were instructed to respond to it in the sequences of stimuli presented. The STIM2 system was used to program and present these sequences of stimuli, which also sends a trigger code to the EEG system corresponding to the different tasks (0, 1, 2-back). The inter-stimulus interval was selected as 2 seconds. N-back experiments were performed on all the subjects using such randomly generated sequences of stimuli. Each N-back experiment was conducted for a total duration of 10 minutes, presenting about 300 visual stimuli to the subject.

Before starting the experiment, the subject was informed about the nature of the task to be performed. After each experiment, a 10-minute break was given to the subject. A 64 channel EEG system (Neuroscan SynAmps 2 from Compumedics, Australia) was used with a 1 kHz sampling rate to acquire the brain signals, and the bandwidth was selected from DC to 200 Hz. EEG data were epoched based on the trigger onset of stimulus as the starting point over a duration of 1800 ms. A small number of data segments of the recorded EEG were found to be contaminated by severe artifacts. Hence, it was decided to remove the data segments whenever an absolute deviation exceeded 150 micro-volts. Eye-blink and cardiac artifacts were also suppressed before analyzing the EEG data using the reference artifact electrodes (ocular:EOG, cardiac:EKG)[41].

3.2 *Features extraction*

MEMED is performed on the cleaned EEG data to obtain IMFs. Different features were extracted from each IMF, including spectral power in selected frequency bands and statistical parameters like variance, skewness, and kurtosis. Along with these features, parameters based on time series (zero-crossing rate and relative MEMD energy) were also calculated. These features can be easily extracted, unlike features based on the evoked related potentials (ERPs). ERPs extraction require the presentation of the stimulus many times to the subject for obtaining a large number of trials ($>$ 100) to perform trigger locked averaging [42, 43], which cancels all uncorrelated signal and add only stimulus related response [44]. However this necessitates to extend the EEG recording duration and this can increase the chance of inducing the effect of fatigue in the EEG studies [45].

4. Data analysis

4.1 *Classification algorithms*

In the present study, we implemented different machine learning (ML) models for the classification task. Different machine learning models will have various data classification strengths that will affect the classification performance [46]. Three non-parametric classifiers (Random Forest, k-Nearest Neighbor, and Support Vector Machine)

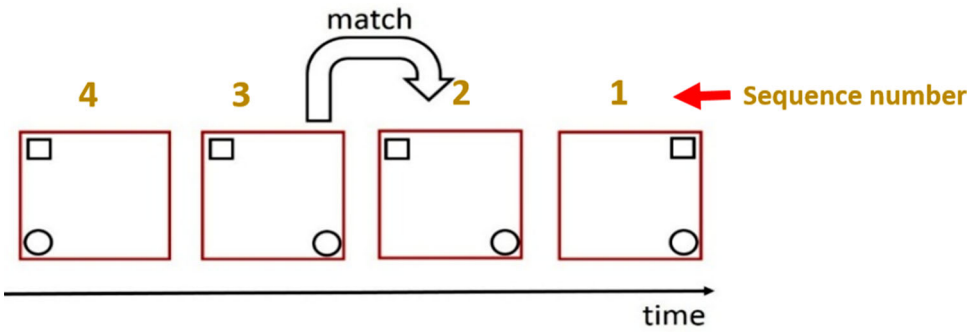

Figure 2. Sequences of stimuli presented during one-back EEG experiment. Subject is expected to respond with a click on the keypad when the third stimulus is presented since it happened to be identical to that presented in the second stimulus (one step earlier).

were reported in the earlier classification studies as the foremost classifiers capable of achieving high accuracies [47]. In the present work, these three non-parametric classifiers and one additional model based on multilayer perceptron were investigated to obtain an optimal classification. After the feature extraction process, features were normalized by removing the mean and dividing the resultant by standard deviation. All the algorithms were implemented using the scikit-learn package in Python [48]. Following different ML models were used :

(i) K-Nearest Neighbour (KNN): This supervised learning algorithm classifies the new data set according to most k-nearest classes training points. In the present study, KNN is implemented based on Minkowski distance, and the value of k was set to 3,5,7, 9,11, and the optimum value of k was found to be 5. [49].

(ii) Support Vector Machine (SVM): The SVM classifier transforms the input data into higher-dimensional space using kernel functions. The SVM finds an optimal hyper-plane that separates the data classes with maximal margin. In this work, SVM has been implemented using a radial basis function [49].

(iii) Multilayer Perceptron (MLP): MLP consists of three sequential layers: input, layer, hidden layer, and output layer. It uses a non-linear activation function and back-propagation for its operation. In our study, we implemented the MLP model with 50 neurons in the hidden layers [50].

(iv) Random forest: This ensemble approach trains several decision trees using the classification and regression trees algorithm. All the decision trees are trained parallelly, and classification decision is made based on the highest priority vote [51]. In the present study, the optimum value for the number of trees was found to be 95.

4.2 *Training and evaluation*

(a) K-fold cross-validation: A given data set is split into a k-number of sections. (k−1) sections are used for training, and one remaining section is taken for testing. In this, each section is used for testing at a particular point. The most significant advantage of this approach is that all the instances of a dataset are employed for validation. In our study, k is taken as 10 to train and test extracted features for all machine learning models [52].

(b) Classifier performance evaluation: The performance of the classifier was measured based on conventional parameters [52]

(i) Accuracy = Total no. of correctly classified cases/Total number of cases.

(ii) Recall = True Positive/(True Positive + False Negative)

(iii) Precision = True Positive/(True Positive + False Positive)

The above classification metrics represent the results, which are the average of the outcomes taken over all the classes. The MEMD technique was used to decompose multichannel EEG data into multivariate intrinsic mode functions (IMFs). Decomposed data (IMFs) of the FP1 channel is shown in figure 3 to represent a part of multichannel EEG data decomposition.

After obtaining the IMFs, a total of 10 features were extracted from each IMFs. The definitions of all the parameters are summarized below.

(i) Spectral features: Spectral analysis is performed using Welch's method [53] in different frequency bands (Delta: 0-4 Hz, Theta: 4-7 Hz, Alpha: 7-13 Hz, Beta: 13-30 Hz and Gamma 30-80 Hz) [54, 55].

(ii) Statistical parameters: To characterize the distribution of the signal amplitude, we calculated the variance, absolute skewness, and kurtosis [56, 57].

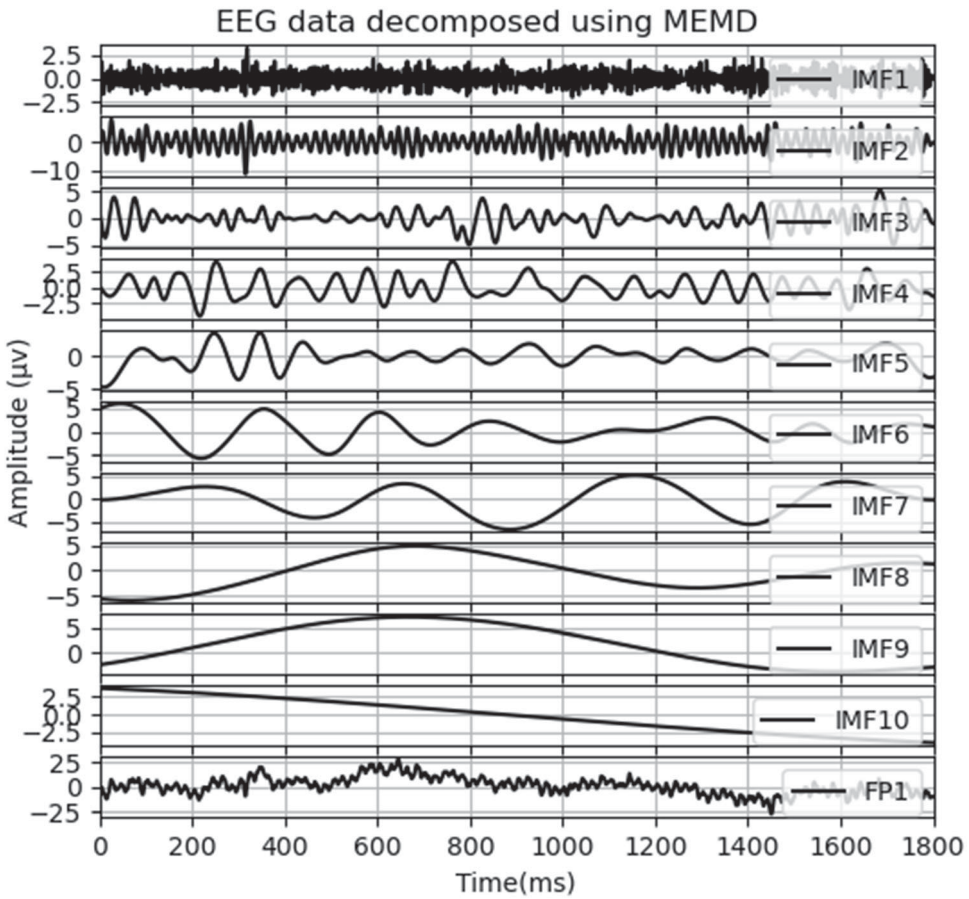

Figure 3. MEMD decomposed data corresponding to FP1.

(a) Variance: It is a second-order moment of central tendency measure and represents the spread of the data around the mean.

$$Variance = E[(x-\mu)^2] \tag{7}$$

where E is the expectation operator and μ is the mean.

(b) Absolute Skewness (abs_skew): Asymmetry of distribution about the mean is measured using skewness. An absolute value of skewness describes lateral shift compared to a distribution that is symmetric about the peak.

$$Skewness = E\left[\left(\frac{x-\mu}{\sigma}\right)^3\right] \tag{8}$$

where σ is the standard deviation.

(c) Kurtosis: It indicates the degree of peakedness of a distribution.

$$Kurtosis = \left(\frac{\mu_4}{\sigma^4}\right) \tag{9}$$

(iii) Time-series based parameters:

(a) Zero crossing rate (zer_cr): This parameter represents the rate of sign-changes in a signal and is commonly used in speech recognition systems [58].

$$zer\_cr = \frac{1}{2N}\sum_{n=0}^{N-1}|sgn[x_i(n)] - sgn[x_i(n-1)]| \tag{10}$$

where

$$sgn[x_i(n)] = \begin{cases} 1, & x_i(n) \geq 0 \\ -1, & x_i(n) < 0 \end{cases} \tag{11}$$

(b) Relative MEMD energy (RME): It gives information about the relative energy associated with the different IMFs.

$$RME = Energy_j / Energy_t \tag{12}$$

where $Energy_j$ is the energy at jth IMF while $Energy_t$ is total energy corresponding to all the IMFs.

Few studies reported that the changes in the brain waves are correlated with the cognitive load [59]. A decrease in alpha and increase in theta has been linked to an increase in mental workload during visual tasks [59, 60]. These studies also reported the association of cognitive tasks with a further smaller section of alpha and beta bands [61]. High-alpha power (10–13 Hz) is inversely related to cognitive tasks [62], and higher beta-band power (20–32 Hz) is associated with working memory [61]. Based on these studies, brain wave power is calculated after the decomposition of EEG data into various frequency bands (IMFs). In our experiments, 64-channels EEG data were decomposed by MEMD technique into IMFs. Table 1 shows the extracted features from IMFs (corresponding to the FP1 channel) after implementing the MEMD technique on EEG data.

The total number of features extracted from each epoch was 6400= 10 (no. of different features) x 10 (no. of IMFs) x 64 (no. of EEG channel). Every feature has been standardized by subtracting the mean and dividing by its standard deviation. These raw features are extracted from each subject and given as input for training and testing different machine learning models; the best classification achieved is only 87.13%. The high-dimensionality of the raw input features raises a concern about achieving good classification accuracy, and hence feature reduction is implemented to obtain optimal accuracy [63]. Many different techniques have been commonly used for feature reduction based on criteria set by principal component analysis, correlation [64–66]. Linear PCA has been performed on all the features for every class. The number of principal components retained was progressively increased to hold the variance around 90% for all the classes, as shown in figure 4. Important transformed features (principal components) have been selected based on the scree plot

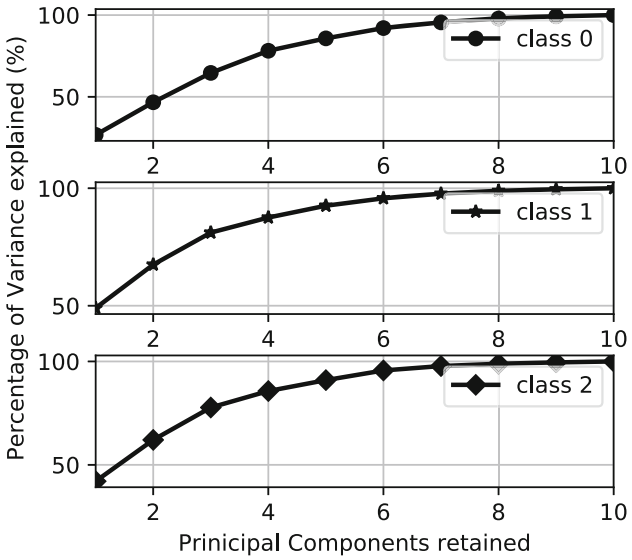

Figure 4. Selection of significant principal components.

[67] for each class. It is found that 5 Principal Components (PCs) are sufficient to retain the variance of around (90%) for each class.

A section of new reduced transformed feature (reduction from 10 features to 5 features) corresponding to IMFs of FP1 is given in table 2. These new reduced features were utilized for training and testing different machine learning models. Table 3 shows the comparison of the classification accuracy based on the raw features and reduced transformed features using linear PCA. The classification accuracy improved after performing feature reduction using linear PCA. Hence taking into account these results, kPCA was used to perform effective feature reduction.

The kPCA is carried on the raw features using different kernel functions. In the present study, second and third degrees of the polynomial with intercept one has been used to obtain the polynomial-based kernel functions. Further

Table 1. A section of extracted features from 10 IMFs after decomposition of EEG data corresponding to FP1 channel.

IMF	Delta	Theta	Alpha	Beta	Gamma	Variance	Abs_skew	kurtosis	Zer_cr	RME
1	0.0001	0.0001	0.0001	0.0001	0.0010	0.6604	0.0872	−0.3500	527	0.0096
2	0.0022	0.0019	0.0043	0.0111	0.1148	8.3786	0.0258	−0.6122	180	0.1226
3	0.0103	0.0238	0.0494	0.0891	0.0031	2.7122	0.0928	0.6637	87	0.0398
4	0.0104	0.1228	0.2745	0.0731	1.64E−05	2.7094	0.1655	−0.5500	45	0.0396
5	0.0149	0.1507	0.4784	0.0031	3.79E−07	1.8987	0.6172	1.4142	26	0.0278
6	0.8789	0.2698	0.1984	1.21E−05	5.60E−09	6.7479	0.0122	−0.2818	12	0.0998
7	1.8053	0.0069	0.1508	6.91E−07	9.02E−10	11.008	0.3838	−0.9692	8	0.1613
8	0.2942	0.0010	0.0090	1.31E−06	4.05E−09	10.082	0.0058	−1.0335	3	0.1501
9	0.1869	0.0008	0.0004	1.29E−06	4.11E−09	15.527	0.0037	−1.4992	2	0.2669
10	0.0179	9.07E−05	2.44E−06	1.69E−07	5.50E−10	5.5908	0.0009	−1.3311	1	0.0820

Table 2. A section of reduced transformed features (10 features to 5 features) based on linear PCA corresponding to the table 1.

IMF	PC1	PC2	PC3	PC4	PC5
1	0.7452	2.2040	−0.236	−1.1353	−1.073
2	−0.307	2.5624	0.5033	1.18831	0.8142
3	3.3277	−0.892	1.2224	−0.6375	1.2914
4	1.9461	−0.839	−1.017	1.38032	−1.255
5	−1.040	−1.267	−1.549	2.01868	0.9772
6	−1.605	−0.446	−1.370	−1.3926	0.6039
7	−0.438	0.1609	−0.593	−1.3845	−0.728
8	−1.291	−0.085	−0.334	−0.3981	0.0361
9	−1.335	−1.395	3.3766	0.36087	−0.665
10	0.4809	3.6355	0.3428	0.48106	0.4774

Table 3. Classifications performance achieved based on raw features and transformed features using linear PCA

ML models	Avg. accuracy based on raw features (%)	Avg. accuracy based on linear PCA transformed features (%)
KNN	86.13	92.28
SVM	82.70	86.63
MLP	83.60	90.18
Random forest	87.13	91.47

higher degree polynomials have not been considered due to increased complexity. In case of RBF-based kernel function, various values of the variance of the Gaussian have been used as the tuning parameter, and the best optimum value was considered to be 0.3 using grid search. The kPCA performs kernel transformation of the raw data in a different feature space. Thus, no direct relation can be derived between variance and kernel Principal Components (kPCs).

Table 4. A section of reduced transformed features (10 features to 5 features) based on kernel PCA (poly, degree = 3) corresponding to the table 1.

IMF	kPC1	kPC2	kPC3	kPC4	kPC5
1	−0.2061	1.0251	0.1699	−0.5887	−1.6958
2	−0.8676	1.6762	1.0927	0.24953	0.54861
3	3.36460	−0.296	1.1914	−0.4949	0.59151
4	0.75055	0.0880	−0.838	2.04788	−0.8309
5	−0.3936	−0.112	−1.445	0.53220	1.90968
6	−0.4442	−0.051	−1.088	−1.2022	−0.0143
7	−0.2179	0.0727	−0.559	−0.5809	−0.5343
8	−0.4113	−0.070	−0.530	−0.5139	−0.0253
9	−1.4191	−2.936	1.8552	0.43180	−0.2659
10	−0.9329	2.4873	1.5961	0.08493	−0.1761

Hence, the optimum number of kPCs to be retained was approximated based on linear PCA; this also allows the proposed approach to operate in an automated way. Table 4 shows the new reduced transformed features using kPCA (ploy., degree=3) after retaining five kPCs.

5. Results

The performance of various machine learning models based on different transformed features using kPCA is given in table 5. The best average classification accuracy of 97.34% (95% confidence interval [CI]: 96.23%, 98.45%) , 93.48% (95% CI: 90.31%, 96.65%) , 95.57% (95% CI: 93.77%, 97.37%), and 95.34% (95% CI: 92.91%, 97.31%) could be achieved based on kPCA (Ploy., degree=3) transformed feature using the KNN, SVM, MLP, and random forest classifier, respectively. Among different machine learning models investigated, the best classification performance could be achieved using KNN. Table 6 indicates the classification metric corresponding to the best average classification. It can be inferred from table 6 that the proposed approach achieves better classification accuracy and also performs reasonably well in precision and recall.

Table 5. Classifications performance achieved using various transformed features.

ML models	Avg. accuracy based on kPCA (RBF, variance = 0.3) transformed features (%)	Avg. accuracy based on kPCA (Ploy., degree = 2) transformed features (%)	Avg. accuracy based on kPCA (Ploy., degree = 3) transformed features (%)
KNN	94.02	95.73	**97.34**
SVM	88.24	91.32	**93.48**
MLP	91.34	94.85	**95.57**
Random forest	92.78	93.58	**95.11**

Table 6. Classification metric using transformed feature by kPCA (Poly., degree = 3).

ML	Accuracy (%)	Precision (%)	Recall (%)
KNN	97.34	97.51	96.15
SVM	93.48	96.87	91.54
MLP	95.57	96.12	94.02
Random forest	95.11	95.73	94.84

Table 7. Comparison of classification accuracy achieved in recent studies.

Authors	Features	Techniques	Accuracy (%)
Shouyi Wang et al. [16]	Wavelet entropy + statistical features + morphological features	Wavelet approach	89.42
Rwan Mahmoud et al. [17]	Statistical features	Discrete Wavelet Transform + stepwise regression	92.86
Małgorzata Plechawska et al. [18]	Spectral analysis	Fast Fourier Transform + logarithm transformation	93.53
Gupta et al. [22]	Statistical + LBP + morphological	Analytic wavelet transform	90.86
Proposed approach	Spectral + statistical + time-series based parameters	MEMD + kPCA (Poly., degree = 3)	97.34

6. Discussion

The present study aims to find an appropriate approach for effectively classifying the N-back Task EEG data. In most of the earlier studies, time-frequency spectrum, wavelet transform, and spatial filtering were used for the classification task. These approaches were based on predefined basis functions, and feature transformation/reduction was achieved based on the linearity assumption. Hence, unable to capture critical information effectively from non-stationary signals like EEG. The present investigation decomposes the multi-channel EEG data using the MEMD approach, which derives the basis functions using the input data itself. Hence, it better suits for analyzing non-stationary signals like EEG. In the current study, all the features are derived from the EEG data recorded from 20 subjects. However, it may be noted that the K-fold cross-validation is performed to improve the situation of limited data, as suggested by other researchers [68, 69]. PCA and kPCA can achieve a feature reduction of about 50% (based on tables 2 and 4), but reduced transformed features based on kPCA can significantly improve classification accuracy by decoding the hidden patterns present in the raw input features using non-linear kernel functions. Table 7 compares the proposed approach with the other techniques used in the earlier cognitive studies based on the data recorded in the present study. An ANOVA test was performed to evaluate the classification accuracy using various techniques [70, 71]. It shows that the classification accuracy among different approaches is statistically significant ($F = 14.9, p < 0.000008$) at alpha level = 0.05; moreover, post-hoc tests using Tukey's HSD test revealed a significant difference ($p < 0.05$) between the proposed approach and other approaches.

7. Conclusion

A new approach is proposed by combing the data-driven MEMD technique with kPCA to obtain effective feature reduction, which also resulted in improved classifying accuracy of the EEG data corresponding to the cognitive N-back study. The best result is obtained using the third-order polynomial function as the kernel function in kPCA, and the best classification accuracy obtained using KNN is 97.34%. The proposed approach performed better than the recent approaches with an added advantage of not depending upon the predefined basis functions for analyzing the EEG data. Moreover, in the present method, extracted features do not depend upon ERPs; hence, EEG recording time is less, resulting in reduced fatigue issues. In the future, this work can be extended with more subjects where other advanced approaches like deep learning can be probed to obtain an optimal classification of N-back task EEG data. The proposed approach has another important implication even beyond the application in the biomedical field; it can be employed in different fields where there is a requirement for feature reduction from the huge raw feature dataset.

Acknowledgements

The authors would like to thank Director MSG, IGCAR, for their valuable support throughout this research work.

Funding No funding to declare.

Data availability The data that support the findings of this study can be available only with the permission of the participants.

Declarations

Conflict of interest All authors have no conflict of interest to report this study.

Consent to participate All subjects gave written informed consent for participation.

Consent for publication The participants agreed to publish the findings based on their data with the following conditions: (1) All information taken from the study will be coded to protect subject name. (2) No names or other identifying information will be used when discussing or reporting data.

Ethical approval This study was carried out on volunteers, with informed consent obtained from them prior to commencing the collaborative work between the Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER), Pondicherry and the Indira Gandhi Centre for Atomic Research (IGCAR), Kalpakkam.

References

[1] M Peruzzini, M Tonietti and C Iani 2019 Transdisciplinary design approach based on driver's workload monitoring, *Journal of Industrial Information Integration* 15 91–102. https://doi.org/10.1016/j.jii.2019.04.001. https://www.sciencedirect.com/science/article/pii/S2452414X1830147X

[2] M U Iqbal, B Srinivasan and R Srinivasan 2020 Dynamic assessment of control room operator's cognitive workload using electroencephalography (eeg), *Computers and Chemical Engineering* 141, 106726

[3] M U Iqbal, M A Shahab, M Choudhary, B Srinivasan and R Srinivasan 2021 Electroencephalography (eeg) based cognitive measures for evaluating the effectiveness of operator training, *Process Safety and Environmental Protection* 150 51–67. https://doi.org/10.1016/j.psep.2021.03.050. https://www.sciencedirect.com/science/article/pii/S0957582021001737

[4] B P 2014 Improving human performance: tackling the challenges to develop effective safety culture, *Oil and Gas Facilities* 03: 18–23

[5] S Yang, L Yang and C He 2001 Improve safety of industrial processes using dynamic operator training simulators, *Process Safety and Environmental Protection* 79 (6): 329–338. https://doi.org/10.1205/095758201753373096. https://www.sciencedirect.com/science/article/pii/S0957582001709834

[6] M Abu-Khader 2004 Impact of human behaviour on process safety management in developing countries, *Process Safety and Environmental Protection* 82 (6): 431–437, risk Management. https://doi.org/10.1205/psep.82.6.431.53206. https://www.sciencedirect.com/science/article/pii/S0957582004711994

[7] J T Selvik and L J Bellamy 2020 Addressing human error when collecting failure cause information in the oil and gas industry: A review of iso 14224:2016, *Reliability Engineering and System Safety* 194 : 1064–18. https://doi.org/10.1016/j.ress.2019.03.025. https://www.sciencedirect.com/science/article/pii/S0951832018311062

[8] D Manca, S Brambilla and S Colombo 2013 Bridging between virtual reality and accident simulation for training of process industry operators, *Adv. Eng. Softw*. 55: 1–9. https://doi.org/10.1016/j.advengsoft.2012.09.002

[9] L Das, M U Iqbal, P Bhavsar, B Srinivasan and R Srinivasan 2018 Toward preventing accidents in process industries by inferring the cognitive state of control room operators through eye tracking, *ACS Sustainable Chemistry and Engineering* 6 (2): 2517–2528. https://doi.org/10.1021/acssuschemeng.7b03971

[10] P Bhavsar, B Srinivasan and R Srinivasan 2016 Pupillometry based real time monitoring of operators cognitive workload to prevent human error during abnormal situations, *Industrial and Engineering Chemistry Research* 55 (12): 3372–3382. https://doi.org/10.1021/acs.iecr.5b03685

[11] C Mulert 2013 Simultaneous EEG and fMRI: towards the characterization of structure and dynamics of brain networks, *Dialogues in clinical neuroscience* 15 (3) 381–386

[12] S M Jaeggi, M Buschkuehl, W J Perrig and B Meier 2010 The concurrent validity of the n-back task as a working memory measure, *Memory* 18 (4): 394–412. https://doi.org/10.1080/09658211003702171

[13] H U Amin, W Mumtaz, A R Subhani, M N M Saad and A S Malik 2017 Classification of *eeg* signals based on pattern recognition approach, *Frontiers in Computational Neuroscience* 11: 103. https://doi.org/10.3389/fncom.2017.00103

[14] P D Gajewski, E Hanisch, M Falkenstein, S Thönes and E Wascher 2018 What does the n-back task measure as we get older? *r*elations between working-memory measures and other cognitive functions across the lifespan, *Frontiers in Psychology* 9: 2208. https://doi.org/10.3389/fpsyg.2018.02208

[15] A M Brouwer, M A Hogervorst, J B F van Erp, T Heffelaar, P H Zimmerman and R Oostenveld 2012 Estimating workload using *eeg* spectral power and ERPs in the n-back task, *Journal of Neural Engineering* 9 (4): 045008. https://doi.org/10.1088/1741-2560/9/4/045008

[16] S Wang, J Gwizdk and W A Chaovalitwongse 2016 Using wireless EEG signals to assess memory workload in the *n*-back task, *IEEE Transactions on Human-Machine Systems* 46 (3): 424–435. https://doi.org/10.1109/THMS.2015.2476818

[17] R Mahmoud, T Shanableh, I P Bodala, N V Thakor and H Al-Nashash 2017 Novel classification system for classifying cognitive workload levels under vague visual stimulation, *IEEE Sensors Journal* 17 (21):7019–7028. https://doi.org/10.1109/JSEN.2017.2727539

[18] M Plechawska-Wójcik, M Tokovarov, M Kaczorowska and D Zapała 2019 A three-class classification of cognitive workload based on eeg spectral data, *Applied Sciences* 9 (24). https://doi.org/10.3390/app9245340. https://www.mdpi.com/2076-3417/9/24/5340

[19] Y Zhou, S Huang, Z Xu, P Wang, X Wu and D Zhang 2021 Cognitive workload recognition using eeg signals and machine learning: A review, *IEEE Transactions on Cognitive and Developmental Systems* 1–1 https://doi.org/10.1109/TCDS.2021.3090217

[20] R N Roy, S Charbonnier, A Campagne and S Bonnet 2016 Efficient mental workload estimation using task-independent *eeg* features, *Journal of Neural Engineering* 13 (2): 026019. https://doi.org/10.1088/1741-2560/13/2/026019

[21] C Scharinger, A Soutschek, T Schubert and P Gerjets 2017 Comparison of the working memory load in n-back and working memory span tasks by means of eeg frequency band power and p300 amplitude, *Frontiers in human neuroscience* 11: 6–6. https://doi.org/10.3389/fnhum.2017.00006

[22] S S Gupta and R R Manthalkar, Classification of visual cognitive workload using analytic wavelet transform, *Biomedical Signal Processing and Control* 61: 101961. https://doi.org/10.1016/j.bspc.2020.101961. http://www.sciencedirect.com/science/article/pii/S1746809420301178

[23] B N Singh and A K Tiwari 2006 Optimal selection of wavelet basis function applied to ecg signal denoising, *Digital Signal Processing* 16 (3): 275–287. https://doi.org/

10.1016/j.dsp.2005.12.003. https://www.sciencedirect.com/science/article/pii/S1051200405001703

[24] N K Al-Qazzaz, S H Bin Mohd Ali, S A Ahmad, M S Islam and J Escudero 2015 Selection of mother wavelet functions for Multi-Channel EEG signal analysis during a working memory task, *Sensors* (Basel, Switzerland) 15 (11): 29015–29035

[25] N E Huang, Z Shen, S R Long, M C Wu, H H Shih, Q Zheng, N-C Yen, C C Tung and H H Liu 1998 The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 454 (1971): 903–995 https://doi.org/10.1098/rspa.1998.0193

[26] Z WU and N E HUANG 2009 Ensemble empirical mode decomposition: A noise-assisted data analysis method, *Advances in Adaptive Data Analysis* 01 (01): 1–41 https://doi.org/10.1142/S1793536909000047

[27] D Labate, F L Foresta, G Occhiuto, F C Morabito, A Lay-Ekuakille and P Vergallo 2013 Empirical mode decomposition vs. wavelet decomposition for the extraction of respiratory signal from single-channel ecg: A comparison, *IEEE Sensors Journal* 13 (7): 2666–2674. https://doi.org/10.1109/JSEN.2013.2257742

[28] Q Wei, Q Liu, S-Z Fan, C-W Lu, T-Y Lin, M F Abbod and J-S Shieh 2013 Analysis of eeg via multivariate empirical mode decomposition for depth of anesthesia based on sample entropy, *Entropy* 15: 3458–3470. https://doi.org/10.1109/JBHI.2014.2333010

[29] N Rehman and D P Mandic 2010 Multivariate empirical mode decomposition, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 466 (2117): 1291–1302. arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.2009.0502. https://doi.org/10.1098/rspa.2009.0502. https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2009.0502

[30] SMU Abdullah, Nu Rehman, MM Khan and DP Mandic 2015 A multivariate empirical mode decompositionbased approach to pansharpening, *IEEE Transactions on Geoscience and Remote Sensing* 53 (7): 3974–3984. https://doi.org/10.1109/TGRS.2015.2388497

[31] S Murawwat, H M Asif, S Ijaz, M Imran Malik and K Raahemifar 2022 Denoising and classification of arrhythmia using memd and ann, *Alexandria Engineering Journal* 61 (4): 2807–2823. https://doi.org/10.1016/j.aej.2021.08.014. https://www.sciencedirect.com/science/article/pii/S1110016821005299

[32] I H Sarker 2021 Machine learning: Algorithms, real-world applications and research directions, *SN Computer Science* 2 (3):160. https://doi.org/10.1007/s42979-021-00592-x

[33] P Mehta, M Bukov, C-H Wang, A G Day, C Richardson, C K Fisher and D J Schwab 2019 A high-bias, low-variance introduction to machine learning for physicists, *Physics Reports* 810 (2019) 1–124, a high-bias, low-variance introduction to Machine Learning for physicists. https://doi.org/10.1016/j.physrep.2019.03.001. https://www.sciencedirect.com/science/article/pii/S0370157319300766

[34] B Mwangi, T S Tian and J C Soares 2014 A review of feature reduction techniques in neuroimaging, *Neuroinformatics* 12 (2): 229–244 https://doi.org/10.1007/s12021-013-9204-3

[35] R-C Chen, C Dewi, S-W Huang and R E Caraka 2020 Selecting critical features for data classification based on machine learning methods, *Journal of Big Data* 7 (1) 52 https://doi.org/10.1186/s40537-020-00327-4

[36] M Z F Nasution, O S Sitompul and M Ramli 2018 PCA based feature reduction to improve the accuracy of decision tree c4.5 classification 978 012058. https://doi.org/10.1088/1742-6596/978/1/012058

[37] J Lever, M Krzywinski and N Altman 2017 Principal component analysis, *Nature Methods* 14 (7): 641–642. https://doi.org/10.1038/nmeth.4346

[38] I T Jolliffe and J Cadima 2016 Principal component analysis: a review and recent developments, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374

[39] D Widjaja, C Varon, A Dorado, J A K Suykens and S Van Huffel 2012 Application of kernel principal component analysis for single-lead-ecg-derived respiration, *IEEE Transactions on Biomedical Engineering* 59 (4): 1169–1176. https://doi.org/10.1109/TBME.2012.2186448

[40] R Patel, K Gireesan and S Sengottuvel 2020 Decoding nonlinearity for effective extraction of the eye-blink artifact pattern from eeg recordings, *Pattern Recognition Letters* 139: 42–49. https://doi.org/10.1016/j.patrec.2018.01.022. https://www.sciencedirect.com/science/article/pii/S0167865518300291

[41] R Patel, K Gireesan, S Sengottuvel, M P Janawadkar and T S Radhakrishnan 2017 Common methodology for cardiac and ocular artifact suppression from eeg recordings by combining ensemble empirical mode decomposition with regression approach, *Journal of Medical and Biological Engineering* 37 (2): 201–208

[42] C Davila and M Mobin 1992 Weighted averaging of evoked potentials, *IEEE Transactions on Biomedical Engineering* 39 (4): 338–345. https://doi.org/10.1109/10.126606

[43] R Patel, M P Janawadkar, S Sengottuvel, K Gireesan and T S Radhakrishnan 2017 Effective extraction of visual event-related pattern by combining template matching with ensemble empirical mode decomposition, *IEEE Sensors Journal* 17 (7): 2146–2153. https://doi.org/10.1109/JSEN.2017.2661993

[44] V Pergher, B Wittevrongel, J Tournoy, B Schoenmakers and M M Van Hulle 2019 Mental workload of young and older adults gauged with erps and spectral power during n-back task performance, *Biological Psychology* 146: 107726. https://doi.org/10.1016/j.biopsycho.2019.107726. https://www.sciencedirect.com/science/article/pii/S0301051118300139

[45] Z Guo, R Chen, X Liu, G Zhao, Y Zheng, M Gong and J Zhang 2018 The impairing effects of mental fatigue on response inhibition: An erp study, *PloS one* 13 (6): e0198206–e0198206. https://doi.org/10.1371/journal.pone.0198206. https://pubmed.ncbi.nlm.nih.gov/29856827

[46] R-C Chen, C Dewi, S-W Huang and R E Caraka 2020 Selecting critical features for data classification based on machine learning methods, *Journal of Big Data* 7 (1): 52. https://doi.org/10.1186/s40537-020-00327-4

[47] P Thanh Noi and M Kappas 2017 Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery, *Sensors* 18 (2) https://doi.org/10.3390/s18010018

[48] F Pedregosa, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, P Prettenhofer, R Weiss, V Dubourg, J Vanderplas, A Passos, D Cournapeau, M Brucher, M Perrot and E Duchesnay 2011 Scikit-learn: Machine learning in python, *J. Mach. Learn. Res.* 12 2825–2830

[49] M P Hosseini, A Hosseini and K Ahi 2010 A review on machine learning for eeg signal processing in bioengineering, *IEEE Reviews in Biomedical Engineering* 1–1 https://doi.org/10.1109/RBME.2020.2969915

[50] C Ieracitano, N Mammone, A Hussain and F C Morabito 2020 A novel multi-modal machine learning based approach for automatic classification of eeg recordings in dementia, *Neural Networks* 123: 176–190. https://doi.org/10.1016/j.neunet.2019.12.006

[51] T M Oshiro, P S Perez and J A Baranauskas 2012 How many trees in a random forest?, in: P. Perner (Ed.), *Machine Learning and Data Mining in Pattern Recognition*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 154–168

[52] S Santaji and V Desai 2020 Analysis of eeg signal to classify sleep stages using machine learning, *Sleep and Vigilance*https://doi.org/10.1007/s41782-020-00101-9

[53] O Faust, R Acharya, A Allen and C Lin 2008 Analysis of eeg signals during epileptic and alcoholic states using ar modeling techniques, *IRBM* 29 (1): 44–52. https://doi.org/10.1016/j.rbmret.2007.11.003. https://www.sciencedirect.com/science/article/pii/S1297956207001209

[54] M Roohi-Azizi, L Azimi, S Heysieattalab and M Aamidfar 2017 Changes of the brain's bioelectrical activity in cognition, consciousness, and some mental disorders, *Medical journal of the Islamic Republic of Iran* 31: 53–53

[55] A Ameera, A Saidatul and Z Ibrahim 2019 Analysis of EEG spectrum bands using power spectral density for pleasure and displeasure state 557 012030. https://doi.org/10.1088/1757-899x/557/1/012030

[56] E Alickovic, J Kevric and A Subasi 2018 Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction, *Biomedical Signal Processing and Control* 39 94 – 102. https://doi.org/10.1016/j.bspc.2017.07.022

[57] I Stancin, M Cifrek and A Jovic 2021 A review of eeg signal features and their application in driver drowsiness detection systems, *Sensors* (Basel, Switzerland) 21 (11): 3786. https://doi.org/10.3390/s21113786. https://pubmed.ncbi.nlm.nih.gov/34070732

[58] T Staudinger and R Polikar 2011 Analysis of complexity based eeg features for the diagnosis of alzheimer's disease, in: *Annu Int Conf IEEE Eng Med Biol Soc,* 2011, pp. 2033–2036 https://doi.org/10.1109/IEMBS.2011.6090374

[59] E Halgren, C Boujon, J Clarke, C Wang and P Chauvel 2002 Rapid distributed fronto-parieto-occipital processing stages during working memory in humans, *Cereb Cortex* 12 (7): 710–728

[60] A Fink, R Grabner, C Neuper and A Neubauer 2005 Eeg alpha band dissociation with increasing task demands, *Cognitive Brain Research* 24 (2): 252–259. https://doi.org/10.1016/j.cogbrainres.2005.02.002. https://www.sciencedirect.com/science/article/pii/S0926641005000418

[61] J Daume, S Graetz, T Gruber, A K Engel and U Friese 2017 Cognitive control during audiovisual working memory engages frontotemporal theta-band interactions, *Scientific Reports* 7 (1): 12585. https://doi.org/10.1038/s41598-017-12511-3

[62] T H Budzynski, H K Budzynski, J R Evans and A Abarbanel 2009 Introduction, in: T H Budzynski, H K Budzynski, J R Evans, A Abarbanel (Eds.), *Introduction to Quantitative EEG and Neurofeedback* (Second Edition), second edition Edition, Academic Press, San Diego, 2009, pp. xxi–xxii. https://doi.org/10.1016/B978-0-12-374534-7.00020-4. https://www.sciencedirect.com/science/article/pii/B9780123745347000204

[63] C Chu, A-L Hsu, K-H Chou, P Bandettini and C Lin 2012 Does feature selection improve classification accuracy? impact of sample size and feature selection on classification using anatomical magnetic resonance images, *NeuroImage* 60 (1) 59–70. https://doi.org/10.1016/j.neuroimage.2011.11.066. https://www.sciencedirect.com/science/article/pii/S1053811911013486

[64] N Salem and S Hussein 2019 Data dimensional reduction and principal components analysis, *Procedia Computer Science* 163 (2019) 292–299, 16th Learning and Technology Conference *Artificial Intelligence and Machine Learning: Embedding the Intelligence*. https://doi.org/10.1016/j.procs.2019.12.111. https://www.sciencedirect.com/science/article/pii/S1877050919321507

[65] M Asadur Rahman, M Foisal Hossain, M Hossain and R Ahmmed 2020 Employing pca and t-statistical approach for feature extraction and classification of emotion from multichannel eeg signal, *Egyptian Informatics Journal* 21 (1): 23–35. https://doi.org/10.1016/j.eij.2019.10.002. https://www.sciencedirect.com/science/article/pii/S1110866519301720

[66] J Jin, Y Miao, I Daly, C Zuo, D Hu and A Cichocki 2019 Correlation-based channel selection and regularized feature optimization for mi-based bci, *Neural Networks* 118: 262–270. https://doi.org/10.1016/j.neunet.2019.07.008. https://www.sciencedirect.com/science/article/pii/S0893608019301960

[67] O Rodionova, S Kucheryavskiy and A Pomerantsev 2021 Efficient tools for principal component analysis of complex data–a tutorial, *Chemometrics and Intelligent Laboratory Systems* 213: 104304. https://doi.org/10.1016/j.chemolab.2021.104304. https://www.sciencedirect.com/science/article/pii/S0169743921000721

[68] A Vabalas, E Gowen, E Poliakoff and A J Casson 2019 Machine learning algorithm validation with a limited sample size, *PLOS ONE* 14 (11): 1–20. https://doi.org/10.1371/journal.pone.0224365

[69] M A Little, G Varoquaux, S Saeb, L Lonini, A Jayaraman, D C Mohr and K P Kording 2017 Using and understanding cross-validation strategies. perspectives on saeb *et al*, *Gigascience* 6 (5): 1–6

[70] N B Toosi, A R Soffianian, S Fakheran, S Pourmanafi, C Ginzler and L T Waser 2019 Comparing different classification algorithms for monitoring mangrove cover changes in southern iran, *Global Ecology and Conservation* 19: e00662. https://doi.org/10.1016/j.gecco.2019.e00662. https://www.sciencedirect.com/science/article/pii/S2351989419300617

[71] V Pergher, B Wittevrongel, J Tournoy, B Schoenmakers and M M Van Hulle, Mental workload of young and older adults gauged with erps and spectral power during n-back task performance, *Biological Psychology* 146: 107726. https://doi.org/10.1016/j.biopsycho.2019.107726. https://www.sciencedirect.com/science/article/pii/S0301051118300139