Journal of Computational Science 68 (2023) 102000

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

Principal component analysis and manifold learning techniques for the design of brain-computer interfaces based on steady-state visually evoked potentials

Bartu Yesilkaya [a], Ebru Sayilgan [b], Yilmaz Kemal Yuce [c], Matjaž Perc [d,e,f,g,h], Yalcin Isler [a,*]

[a] Izmir Katip Celebi University, Department of Biomedical Engineering, Balatcik Campus, Cigli, 35620 Izmir, Turkey
[b] Izmir University of Economics, Department of Mechatronics Engineering, Balcova, 35330 Izmir, Turkey
[c] Department of Computer Engineering, Alanya Alaaddin Keykubat University, Alanya, 07425 Antalya, Turkey
[d] Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, 2000 Maribor, Slovenia
[e] Department of Medical Research, China Medical University Hospital, China Medical University, Taichung 404332, Taiwan
[f] Alma Mater Europaea, Slovenska ulica 17, 2000 Maribor, Slovenia
[g] Complexity Science Hub Vienna, Josefstädterstraße 39, 1080 Vienna, Austria
[h] Department of Physics, Kyung Hee University, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Manifold learning
Brain-computer interface
Steady-state visual evoked potential
Principal component analysis
Feature reduction

ABSTRACT

Steady-state visually evoked potentials (SSVEP) are stochastic and nonstationary bioelectric signals. Because of these properties, it is difficult to achieve high classification accuracy, especially when many considered features lead to a complex structure. We therefore propose a manifold learning framework to decrease the number of features and to classify SSVEP data by comparing lower dimensional matrices with well-known machine learning algorithms. We use the AVI-SSVEP Dataset, which includes stimuli at seven different frequencies and 15360 samples per person. The SSVEP features are extracted from relevant and distinctive frequency-domain, time-domain, and time–frequency domain properties, creating a total of 55 feature vectors. We then analyze and compare five divergent manifold learning methods with respect to their performance on nine different machine-learning algorithms. Among all considered manifold learning methods, we show that the Principal Component Analysis has the best classifier performance with an average of 22 components. Moreover, the Naive Bayes classifier with the Principal Component Analysis achieves the maximum accuracy of 50.0%–80.95% for a 7-class classification problem. Our research thus shows that the proposed analytical framework can significantly improve the decoding accuracy of 7-class SSVEP problems, and that it exhibits notable robustness and efficiency for small group datasets.

1. Introduction

Brain-computer interface (BCI) is a computer-assisted system that acquires brain signals via brain monitoring methods, and convert them into commands [1]. These commands are communicated to an output device to carry out desired actions. BCI systems, at first, must employ a method for data acquisition to capture brain activities [2]. For this purpose, Electroencephalography (EEG) is so popular among the numerous methods reflecting brain activity [3]. Since EEG have many advantages; such as ability to produce data promptly, ease of use, cost (cheaper equipment), and disposable equipment, EEG has become a common method to capture and measure brain activity for BCI applications [4].

Many BCI applications were developed and improved using EEG control signals such as event-related potentials (ERP), visually evoked potentials (VEPs), and many others [5,6]. Among these control signals, VEP-based BCI systems show a reliable and robust performance proved in many clinical tests from different laboratories throughout the world [7]. Along side the mentioned advantages of EEG, VEP's high information transfer rate (ITR), high signal-to-noise ratio (SNR), simplicity in configuration, and users' shorter training time have led VEP attract many researchers recently [8].

A variant to VEP is called Steady-State Visually-Evoked Potential (SSVEP). Generally, SSVEP is a transient answer of the occipital region that occurs after performing a short optical stimulus, which generates consistent and small-scale amplitude VEP signals [9]. In other words, SSVEPs are resonance phenomena that take place mainly in the visual cerebral cortex when people's visual attention focus on a light source

* Corresponding author.
E-mail addresses: bartu.yesilkaya@yandex.com (B. Yesilkaya), ebru.sayilgan@ieu.edu.tr (E. Sayilgan), yilmazkemalyuce@gmail.com (Y.K. Yuce), matjaz.perc@gmail.com (M. Perc), islerya@yahoo.com (Y. Isler).

https://doi.org/10.1016/j.jocs.2023.102000
Received 9 January 2023; Received in revised form 27 February 2023; Accepted 13 March 2023
Available online 21 March 2023

that flashes above 4 Hz frequency [10]. SSVEPs also include a few harmonic components, which are happened periodically [11,12].

Using the adopted approach in this study, to improve the efficiency of the SSVEP-based BCI systems and to produce meaningful commands from the acquired SSVEP data, it is crucial to use and analyze various kinds of feature extraction methods, feature reduction methods, and machine learning algorithms. Although many feature extraction methods, feature reduction methods, and machine learning algorithms have been offered for SSVEP-based BCI applications, it is challenging to decide and pick the most efficient combination due to a lack of comparison. It is also essential to obtain optimal features that help achieving higher classifier performances. Therefore, it is vital to use adjustable features and feature reduction techniques to facilitate the subsequent operation of the classifier(s) [13].

In the early part of this century, many distinct manifold learning methods were offered such as Multi-dimensional Scaling (MDS), Locally Linear Embedding (LLE) [14], Isometric Feature Mapping (Isomaps) [15,16], t-Distributed Stochastic Neighbor Embedding (t-SNE) [17,18]. The aforementioned feature reduction methods clarifies visualization when data samples depend on manifolds and are involved under the umbrella phrase of manifold-learning [4,19]. Manifold learning is a framework for dimensional reduction, which mentions to the difficulty of mapping high-dimensional data to low-dimensional data while preserving as much of the original data quality as possible [20].

In this study, we applied state-of-the-art methods like; LLE, t-SEN, ISOMAP for manifold learning, and conventional methods like; Principal Component Analysis (PCA) and MDS, to observe the success of dimension reduction methods comparatively. The proposed approach was tested by using the features obtained in the time, frequency, and time–frequency domains from the SSVEP data which were obtained by giving stimuli at seven different frequencies (6, 6.5, 7, 7.5, 8.2, 9.3, 10 Hz). PCA, MDS, LLE, t-SEN, and ISOMAP feature reduction methods were analyzed via the Orange Data Mining program. Then without feature selection, new and fewer features were composed and classified by the nine different machine learning algorithms (Stochastic Gradient Descent, AdaBoost, Random Forests, Naive Bayes, Decision Trees, Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Multi-Layer Perceptron). In the final step, by comparing, it was investigated which method(s) gives higher accuracy in classification results. Ultimately, it is crucial to have more interpretable SSVEP-based features that will lead to more efficient BCI design in terms of accurate recognition rates and to learn more about the mental processes in that BCI users control the system.

1.1. State of the art

In the study conducted by Aruna Tyagi and Vijay Nehra [21] in 2017, authors analyzed dimension reduction methods like Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), Factor Analysis (FA), Isometric Feature Mapping (ISOMAP) and Multi Dimensional Scaling (MDS) methods on the dataset "Motor Imagery of BCI competition IV" which is available and accessible on the Internet for free. This data set includes EEG data obtained during both hand and foot motor imagination. They used Artificial Neural Network (ANN) classifier for classification. Using dimension reduction methods, they obtained 0.2451, 0.1143, 0.2376, 0.2156, 0.2410 Mean square error (MSE) values, respectively.

Wanzhong Chen et al. [22] conducted a study in 2020, using BCI Competition II and III datasets. As dimension reduction method, the authors suggested MDS and they used LDA classifier in the classification stage. In their study, they also adopted and applied MDS, PCA, Kernel Principal Component Analysis (KPCA), Locally Linear Embedding (LLE), and Laplacian Eigenmaps (LE) methods for left-hand and right-hand motor imagination classification. They obtained the features with Flexible Analytic Wavelet Transform (FAWT). They achieved the highest accuracy 95% using MDS and FAWT.

In 2016, Ming-ai Li et al. [23] carried out a brain-computer interface study on a dataset that classified left and right hand movements provided by "BCI Competition 2003". In this study, the researchers propose the P. t-SNE (parametric t-Distributed Stochastic Neighbor Embedding) method. They obtained the attributes using DWT(Discrete Wavelet Transform). They compared the proposed method with the MDS, PCA, KPCA, LLE and LE methods. Using these methods and attributes, they achieved 94.1% success with the SVM (Support Vector Machine) classifier.

In another study conducted in 2016, Egor Krikov and Mikhail Belyaev [24] worked on the classification of left hand, right hand, legs, and tongue movements using EEG signals. They applied CSP (common spatial patterns), Isomap and PGA (principal geodesic analysis) methods on EEG data, which were acquired from 9 subjects. They used the LDA method for classification. They achieved an average of 61% success with the PGA method.

In 2019, Ming-ai Li et al. [25] compared many dimension reduction methods, especially ISOMAP, LLE, LE and MVU, with the Landmark version of Maximum Variance Unfolding (L-MVU). Using this method, they achieved 89.64% success on the 'BCI Competition 2003' dataset.

In the study of Muhammad Tariq Sadiq et al. [26] in 2021, dimension reduction methods were tested using 7 different classifiers. One of the dataset they used contains EEG signals containing the right hand-right foot activities of 5 subjects. In the other dataset, there are EEG signals that contain the left hand and right foot activities of a single subject. PCA, ICA (independent component analysis), LDA and NCA (neighborhood component analysis) methods were tested on data from different numbers (3,7,18) EEG channels. They found the most successful result using 7-channel EEG signals. The success rate achieved by applying EWT (empirical wavelet transform) and NCA methods together is 100%.

In 2017, Ming-ai Li et al. [27] recommends the use of wavelet packet decomposition (WPD) and supervised explicit isomap (SE-isomap) methods on 'BCI Competition 2008 Datasets 2b' in their study. There are EEG recordings of right and left hand movements taken from 9 people in the dataset. They extracted the feature using OWP (optimal wavelet Packets). They compared the SE-isomap method with PCA, MDS, LLE, which are dimension reduction methods. According to the classification result they performed using K-nearest neighbor (K-NN) (K=7), they achieved an average of 92.7% success with the method they suggested.

In 2020, Ping Tan et al. [28] propose dimensionality reduction mechanism (DimReM)- Evolutionary Algorithms (EAs) method in their study using EEG data from BCI Competition III dataset IVa and BCI Competition IV dataset IIb. In their study using two different datasets, there are EEG recordings of 5 people in the first dataset and 9 people in the second dataset. They used 3 different classifiers: K-NN, SVM and Discriminant Analysis (DA). They preferred Evolutionary Algorithms methods for feature selection. Binary Particle Swarm Optimization (BPSO), Novel modified binary differential evolution (NMBDE), Genetic algorithm (GA) are the Evolutionary Algorithms methods they recommend for comparison. Finally, they show the superiority of their proposed methods by comparing these methods with PCA, ICA, NCA, variable-length PSO (VLPSO).

In 2004, Felix Lee et al. [29] classified EEG signals of left hand, right hand, tongue and foot movements. EEG recordings were taken from 6 people with the experimental procedure they prepared. They tried PCA, LLE and Isomap dimension reduction methods. They obtained the evaluation results with the leave-one-out method. They concluded that ISOMAP was more successful than LLE and PCA.

In 2006, John Q. Gan [30] conducted a BCI study having 3 subjects with EEG recordings. The author converted 96 features obtained using power spectral density (PSD) into lower-dimensional feature matrices thanks to the dimension reduction methods PCA, LLE, Locality Preserving Projection (LPP) and Supervised PCA (SPCA). These matrices were classified by LDA using the 3-fold cross-validation method. According to the classification results of the three mental tasks, an average of 71.95% success was achieved with the LPP method. He stated that LPP gave better results than SPCA and other methods.

1.2. Contributions

The primary contributions of this research study are summarized as follows:

- To affirm a flexible manifold learning framework for recognition of **7-class SSVEP problem** from either sufficient or small training samples automatically,
- To present and perform the **PCA, LLE, MDS, ISOMAP, and t-SNE approaches** for reducing the SSVEP feature matrix obtained from the various and distinctive domains,
- To suggest correlation-based **components** for implicit selection of the PCA, LLE, MDS, t-SNE, and ISOMAP, so that they can be used as powerful biomarkers of SSVEP for 7-class problem findings,
- To investigate an appropriate and sustainable **machine learning model** for the proposed features to recognize the SSVEP tasks, and enhance performance success rate as compared with the existing methods,
- To design an effective subject-independent **expert 7-class SSVEP-based BCI system**.

In this study, we suggest a novel flexible framework and proved valid results for the SSVEP signal that is efficient for small training samples. Last but not least we proposed a solution to 7-class SSVEP problems, because in literature, generally, the high-performance solutions are only for 2-class problems. For the accurate evaluation of the suggested framework, we employed the Accuracy (ACC) metric.

1.3. Organization

The rest of the paper is organized as follows: Materials and methods used in this study are described in Section 2. In addition, the adopted assessment criteria for experiments are explained in Section 2, as well. Section 3 describes the observations and presents the results of the proposed approaches. Finally, it is summarized and concluded in Section 4.

2. Materials and methods

This study thoroughly consists of five steps as seen. These are Data Acquisition, Feature Extraction, Feature Reduction, Classification, and BCI Application, respectively. The operations performed in each step are explained in detail in the sub-headings.

2.1. Dataset description

In this study, open source data set called "AVI SSVEP Dataset" [31] is adopted, which contains SSVEP signals acquired by Adnan Vilic in a specific setting. The data are composed of SSVEP measurements from 4 healthy individuals staring at the flickering target, whose color changed rapidly from black to white, to trigger responses of signals at 7 different frequencies. Participants were composed of 3 men and 1 woman, for whom age ranges from 27 to 32. AVI SSVEP data were acquired using 3 electrodes (the signal electrode is Oz, the reference electrode is Fz, and the ground is Fpz). These electrodes were placed by following the international standard 10–20 system for electrode placement. In addition, the sampling frequency was 512 Hz.

During the experiment, participants have seated in front of an LCD computer monitor with a refresh rate of 120 Hz. In order to eliminate the interference frequency (50 Hz), an analog notch filter was applied to the data [31]. The stimulus is a flickering box at seven different frequencies (6.0, 6.5, 7.0, 7.5, 8.2, 9.3, and 10.0 Hz) shown on the screen. The data set contains four sessions for each participant. In one session each trial lasts 30 s, and attendees take a tiny break between the trials. Experiments for all attendees were repeated at least three times for each frequency.

2.2. Feature extraction

SSVEP signals are usually recorded with a series of electrodes ranging from 1 to 512 and a sampling frequency ranging from 4 Hz to 100 Hz [7,9,10]. Therefore, a huge amount of data is collected using such settings [11]. From data collected, extracting useful features suitable for the signal structure is a crucial step in BCI design [32]. To this purpose, a feature extraction method should be adopted. In this respect, Some studies [11–13,32,33] show that the selection of an appropriate preprocessing and feature extraction method has more impact on the final performance than the choice of a machine learning algorithm. Many feature extraction techniques have been researched and proposed for SSVEP-based BCI systems [34].

In this study, a total of 55 features were extracted from time, frequency, and time–frequency domains. Firstly, the SSVEP time-domain features are extracted from the original field of the EEG signal. These features are based on the amplitude (e.g. average amplitude change value, root mean square, interquartile ranges, etc.) and statistical changes of the EEG signal (e.g., mean, variance, skewness, and kurtosis, etc.) Secondly, SSVEP frequency-domain features were extracted from the frequency domain representation of the SSVEP signal using a Fourier Transform [35]. The relevant and distinctive SSVEP frequency characteristics we detected are based on the spectral information such as energy, variance and spectral entropy for each EEG rhythm. These features explain how power, variance, and irregularity (entropy) change in certain related frequency bands [36–38]. Last but not least, using Db8 DWT function, SSVEP signals are subdivided into frequency bands (delta, theta, alpha, beta, gamma), hence, the energy, entropy and variance were calculated for each band. Every DWT frequency band is associated with one or two EEG rhythms. Thus, a number of features represented in the frequency bands were obtained [39].

2.3. Data reduction using principal component analysis and manifold learning methods

Following data extraction, data reduction is the next step to realize. Data reduction's objective is to explore a set of features with the least number of elements that shall guide to optimal classification. To realize that a data reduction method should be adopted among many. Data reduction allows invalid data that is no good to classification to be eliminated, thereby, enables the system to get the optimal classifier performance from a designated system. While trying to get the optimal classifier performance from the system, it is necessary to work with fewer data describing some relevant features of SSVEP signals. For this purpose, one of the most famous methods called Principal Component Analysis and the manifold learning methods that forms up the originality of this study are briefly summarized in the following subsections.

2.3.1. Principal Component Analysis

Principal Component Analysis (PCA) is one of the linear dimension reduction methods based on the covariance matrix of variables. The main purpose of PCA is to keep the data set with the highest variance in high-dimensional data, but to provide dimension reduction at the same time. By finding the general features in the over-dimensional data, it reduces the number of dimensions and compresses the data. It is certain that some features will be lost with dimension reduction but the intent is that these disappearing traits contain little information about the population. Basically, PCA combines highly correlated variables to create a smaller set of artificial variables, called "principal components", that make up the most variation in the data [40,41].

First of all, in the PCA method, the average of the data set is found, as seen in Eq. (1). Assuming we call our m*n-sized dataset matrix **X**, m is number of feature, n is number of sample and i shows us the row index in $x_i (i \epsilon [1, 2, \ldots, n])$.

$$\overline{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{1}$$

As a second step, the mean value is subtracted from each element in the data set, and the covariance matrix(C) is calculated as seen in Eq. (2).

$$\mathbf{C} = \frac{1}{1-n}\sum_{i=1}^{n}(X-\overline{X})(X-\overline{X})^T \tag{2}$$

Covariance is a measure of how much two variables change together. The covariance matrix is the expression of the covariances between the elements of a vector in a matrix. As a third step, eigenvalues (λ) and eigenvectors(V_k) are calculated using the covariance matrix as seen in Eq. (3).

$$det(\lambda I - C) = 0, \tag{3}$$

$$(C - \lambda_k I)V_k = 0 \tag{4}$$

After finding eigenvectors from the covariance matrix, the vectors are ordered according to their eigenvalues. The vector with the highest eigenvalue in the dataset is called the fundamental component of this dataset. Even though discarding vectors with small eigenvalues causes information loss, the dataset size is indicated as being smaller than the first. If the PCA method is applied on an n-dimensional dataset, n eigenvectors and n eigenvalues are obtained. The n-dimensional dataset is expressed with a k-dimensional dataset by choosing the k eigenvectors with the highest eigenvalues.

2.3.2. t-Distributed Stochastic Neighbor Embedding (t-SNE)

T-Distributed Stochastic (Random) Neighbor Embedding (t-SNE) is an unsupervised, non-linear technique used primarily for data exploration and visualization of high-dimensional data. It calculates the probability that pairs of data points in high-dimensional space are related and then chooses a low-dimensional embedding that produces a similar distribution. In simpler terms, t-SNE gives you a sense or intuition of how data is arranged in a high-dimensional space. It was developed by Laurens van der Maaten and Geoffrey Hinton in 2008 [17].

The T-SNE algorithm calculates the similarity measure between pairs of samples in high- and low-dimensional space. It then tries to optimize these two similarity measures using a cost function. This is done in 3 basic steps: In Step 1, similarities between points in high-dimensional space are measured. Consider a bunch of data points scattered in 2D space. For each data point (x_i), a Gaussian distribution is centered on that point. Then the intensity of all points (x_j) in this Gaussian distribution is measured. All points are then normalized. This gives us a set of probabilities (p_{ij}) for all points. These probabilities are proportional to the similarities. This means that if x_1 and x_2 data points have equal values under this Gaussian circle, then their ratios and similarities are the same, and hence there are local similarities in the structure of this higher-dimensional space [17]. Step 2 is also similar to Step 1, but instead of using a Gaussian distribution, a Student's t-distribution with one degree of freedom, also known as the Cauchy distribution, is used. This gives us a second set of possibilities (q_{ij}) in low-dimensional space. The Student t-distribution has heavier tails than the normal distribution. Heavy tails allow better modeling of long distances. In the last step, this set of possibilities is asked to reflect as well as possible from the low-dimensional space (q_{ij}) to the high-dimensional space (p_{ij}). We expect the two map structures to be similar. The difference between the probability distributions of two-dimensional spaces is measured using the Kullback-Liebler deviation (KL). KL is an asymmetric approach that effectively compares p_{ij} and q_{ij} values.

2.3.3. Multidimensional scaling (MDS)

Classical multidimensional scaling analysis, which is the metric multidimensional scaling approach, was proposed by Torgerson k.k. 1952 [42]. In classical multidimensional scaling analysis, which is also called Principal Coordinates Analysis (PCoA) due to its similarity to PCA, the distances between the units in the **X** data matrix are determined by the Euclidean distance [43].

When working with quantitative variables, the most widely used distance function for determining the distances between units is the Euclidean distance function. The Euclidean distance between two units is calculated as in the formula.

$$d(x_i, x_j) = \{\sum_{l=1}^{p}(x_{il} - x_{jl})^2\}^{1/2} \tag{5}$$

Classical multidimensional scaling consists of the square distance matrix $D^{(2)}$ and the centralization matrix **H** which is formed by squaring each distance in the distance matrix **D** formed by the Euclidean distance.

$$B = -\frac{1}{2}HD^2H \tag{6}$$

It is an approach based on the eigenvalues and eigenvectors of the defined **B** matrix. The centralization matrix **H** is obtained by the equation

$$H = I - n^{-1}LL^T \tag{7}$$

and with a unit matrix **I** and a vector **L**. The reduced coordinate system is created by Eq. (8), in classical multidimensional scaling analysis.

$$Y = A_d L_d^{1/2} \tag{8}$$

2.3.4. Isometric Feature Mapping (ISOMAP)

The Isometric Mapping (ISOMAP) method is a graph-based, nonlinear dimension reduction method that aims to represent high-dimensional data in a lower-dimensional coordinate system. With principal component analysis and multidimensional scaling analysis, effective results may not be obtained in data sets where data points have nonlinear relationships. In order to solve this problem, it is proposed the isometric mapping method [15]. This method is an extended version of the classical multidimensional scaling method. According to this method, Geodetic distance is used instead of Euclidean distance. Geodetic distance is defined as the shortest curvilinear length between two points along a manifold path on a surface.

Normally, Euclidean distance is used to calculate the similarity between two units. However, if the Euclidean length is used, the internal geometry of the manifold becomes unprotected. Two points that are similar in terms of Euclidean distance, i.e. close, actually may be far apart, since their true distance may be the length of the path between these points along the manifold [15].

In the Isometric matching method, which is one of the graph-based dimension reduction methods, the points that are close after the dimension reduction process remain close to each other. Long-distance points keep their places at a distance, too.

The local linearity principle is used in the isometric mapping method, and neighboring points are assumed to lie on a linear patch of the manifold. Therefore, Euclidean distances for nearby points are considered to accurately estimate geodetic distances. For remote points, geodetic distances are estimated by adding adjacent distances on the manifold. The algorithm of the isometric matching method is as follows:

- Step 1 Determine the nearest neighbors for all data points.
- Step 2 A weighted graph is created connecting each point to its nearest neighbors, with nodes showing data points and links showing distances between points.
- Step 3 In the generated neighborhood graph, their distances are redefined, hence, the length of the shortest path between two points.
- Step 4 Classical multidimensional scaling is applied to the new distance matrix defined in step 3.

Adjacent points are considered to have edges between them. The weights of the edges, called $d_x(i,j)$ are assumed to be the Euclidean distance between neighbors i and j. Under these two assumptions, all data points are represented as nodes in a weighted G-graph. The shortest paths named $d_G(i,j)$ in the weighted graph resembling the G letter are calculated using the Floyd algorithm between all pairs of data points. The calculated lengths are used as an estimate of the geodetic distance called $d_M(i,j)$ on the M manifold. Finally, dimension reduction is performed by classical multidimensional scaling based on the nxn-dimensional symmetric geodetic distance matrix $D_G = d_G(i,j)$ between all pairs of data points on the M manifold.

The isometric mapping method aims to reduce the size in a way that preserves the geodesic distance matrix called D_G between the data point pairs. The error function is trying to be minimized in the isometric matching method with the reduced-dimensional distance matrix called D_Y.

Isometric mapping finds the true dimensionality of nonlinear distances as long as sufficient data is provided. The success of the method depends on the number of k and the radius ϵ, which determine the neighborhood. Since the isometric mapping method does not define a mathematical function between the input and the output, it is one of the disadvantages of the method that the whole process is repeated from scratch when a new data point is added [44].

2.3.5. Locally Linear Embedding (LLE)

The LLE algorithm characterizes the low-dimensional local geometry of the data points by finding linear coefficients that reconstruct each data point using only its nearest neighbors according to the Euclidean distance. The error in re-weighting is also measured by Eq. (9).

$$E(W) = \sum_i \left| x_i - \sum_j w_{ij} x_j \right|^2 \tag{9}$$

The j index in the equation shows the data points in the nearest k neighborhood of the x_i data point. Optimal weights in the error function are obtained using the least squares method under the constraint

$$\sum_j w_{ij} x_j = 1 \tag{10}$$

By fixing the optimal weights w_{ij} distances from the original dimension are represented in reduced dimension. The reduced-dimensional representation of the original distances is accomplished by minimizing under the constraints of the quadratic objective function defined by Eq. (11).

$$\Phi(S) = \sum_i \left| s_i - \sum_j w_{ij} s_j \right|^2 \tag{11}$$

$$\frac{1}{n} \sum_{i=1}^{n} s_i^T s_i = I \tag{12}$$

$$\sum_{i=1}^{n} s_i = 0 \tag{13}$$

The sparse eigenvalue–eigenvector approach can be used to solve the minimization problem. Symmetrical and semi-positive nxn dimensional sparse matrix for eigenvalue decomposition will be performed is obtained by Eq. (14).

$$M = (I - W)(I - W)^T \tag{14}$$

The eigenvectors corresponding to the smallest nonzero eigenvalue M of the d matrix provide independent coordinates centered at the origin. The algorithm for the local linear mapping method is as follows;

- Step 1 Determine k for the neighborhood and the number of dimensions d in the reduced coordinate system.
- Step 2 For each x_i point, the nearest k neighbor is determined.
- Step 3 The weights of each point x_i to its nearest neighbors Equation is reconstructed with x to calculate linear weights w_{ij}.
- Step 4 Equation xx and s_i points are created in the reduced d-dimensional space so that the weights determined in Step 3 remain the same.

Table 1
All classifier performances were tested in this study using the Principal Component Analysis method. The maximum classifier accuracies are shown in boldface for each subject. Classifier performances are given as accuracies where the perfect accuracy is 1.00. Naive Bayes classifier gave the highest accuracies for all subjects where subjects' features had been utilized using the principal component analysis method.

Classifier	Subject 1	Subject 2	Subject 3	Subject 4	Average
k-Nearest Neighbors	0.17	0.15	0.24	0.14	0.18
Decision Tree	0.21	0.27	0.43	0.19	0.27
Support Vector Machines	0.38	0.42	0.38	0.33	0.38
Stochastic Gradient Descent	0.38	0.23	0.33	0.38	0.33
Random Forest	0.25	0.38	0.29	0.29	0.30
Multi-Layer Perceptron	0.25	0.38	0.33	0.24	0.30
Naive Bayes	**0.58**	**0.50**	**0.81**	**0.71**	**0.65**
Logistic Regression	0.17	0.31	0.14	0.19	0.20
AdaBoost	0.25	0.23	0.48	0.33	0.32

Table 2
All classifier performances were tested in this study using the t-Distributed Stochastic Neighbor Embedding method. The maximum classifier accuracies are shown in boldface for each subject. Classifier performances are given as accuracies where the perfect accuracy is 1.00. Naive Bayes classifier gave the highest accuracies for almost all subjects where subjects' features had been utilized using the t-Distributed Stochastic Neighbor Embedding method.

Model	Subject 1	Subject 2	Subject 3	Subject 4	Average
k-Nearest Neighbors	0.21	0.19	0.24	0.19	0.21
Decision Tree	**0.33**	0.31	0.19	0.43	0.32
Support Vector Machines	0.17	0.27	0.05	0.24	0.18
Stochastic Gradient Descent	0.17	0.27	0.14	0.19	0.19
Random Forest	0.17	0.35	0.14	0.24	0.22
Multi-Layer Perceptron	0.17	0.38	0.10	0.33	0.24
Naive Bayes	0.25	**0.42**	**0.48**	**0.48**	**0.41**
Logistic Regression	0.21	0.23	0.10	0.19	0.18
AdaBoost	0.17	0.42	0.33	0.38	0.33

Table 3
All classifier performances were tested in this study using the Multidimensional Scaling method. The maximum classifier accuracies are shown in boldface for each subject. Classifier performances are given as accuracies where the perfect accuracy is 1.00. Naive Bayes classifier gave the highest accuracies for almost all subjects where subjects' features had been utilized using the Multidimensional Scaling method.

Model	Subject 1	Subject 2	Subject 3	Subject 4	Average
k-Nearest Neighbors	0.25	0.35	0.14	0.29	0.26
Decision Tree	0.21	0.15	0.29	0.14	0.20
Support Vector Machines	0.25	**0.46**	0.29	0.14	0.29
Stochastic Gradient Descent	0.25	0.31	0.33	0.29	0.29
Random Forest	0.25	0.27	0.24	0.29	0.26
Multi-Layer Perceptron	0.13	0.42	0.24	0.19	0.24
Naive Bayes	**0.54**	**0.46**	**0.67**	**0.48**	**0.54**
Logistic Regression	0.25	0.27	0.14	0.14	0.20
AdaBoost	0.29	0.31	0.33	0.29	0.30

When these performances are examined in detail, it is seen that the most successful classifier is Naive Bayes algorithm. Therefore, the highest performance metrics and performing number of components obtained for all feature reduction methods using Naive Bayes classifier are given in Table 6. The average classifier performances are 0.65, 0.57, 0.52, 0.35, and 0.40 with 22, 15, 17, 6, and 9.5 components for PCA, LLE, MDS, ISOMAP, and T-SNE, respectively.

2.4. Classification

In pattern recognition problems, it is so vital to do the identification process accurately. For SSVEP-based BCI studies, this process generally

Table 4
All classifier performances were tested in this study using the Isometric Feature Mapping method. The maximum classifier accuracies are shown in boldface for each subject. Classifier performances are given as accuracies where the perfect accuracy is 1.00. Naive Bayes classifier gave the highest accuracies for almost all subjects where subjects' features had been utilized using the Isometric Feature Mapping method.

Model	Subject 1	Subject 2	Subject 3	Subject 4	Average
k-Nearest Neighbors	0.25	0.31	0.29	0.24	0.27
Decision Tree	0.13	0.27	0.14	0.24	0.19
Support Vector Machines	0.17	0.31	0.29	0.19	0.24
Stochastic Gradient Descent	0.13	0.19	0.29	0.29	0.22
Random Forest	0.17	0.27	0.24	0.29	0.24
Multi-Layer Perceptron	0.13	0.23	0.29	0.24	0.22
Naive Bayes	**0.29**	**0.35**	0.33	**0.43**	**0.35**
Logistic Regression	0.21	0.23	0.19	0.14	0.19
AdaBoost	0.17	**0.35**	**0.38**	0.29	0.29

Table 5
All classifier performances were tested in this study using the Locally Linear Embedding method. The maximum classifier accuracies are shown in boldface for each subject. Classifier performances are given as accuracies where the perfect accuracy is 1.00. Naive Bayes classifier gave the highest accuracies for all subjects where subjects' features had been utilized using the Locally Linear Embedding method.

Model	Subject 1	Subject 2	Subject 3	Subject 4	Average
k-Nearest Neighbors	0.21	0.27	0.24	0.24	0.24
Decision Tree	0.21	0.35	0.38	0.14	0.27
Support Vector Machines	0.25	0.42	0.29	0.24	0.30
Stochastic Gradient Descent	0.21	0.31	0.29	0.19	0.25
Random Forest	0.17	0.35	0.24	0.19	0.24
Multi-Layer Perceptron	0.17	0.46	0.33	0.29	0.31
Naive Bayes	**0.50**	**0.54**	**0.71**	**0.52**	**0.57**
Logistic Regression	0.25	0.35	0.10	0.19	0.22
AdaBoost	0.17	0.35	0.43	0.24	0.29

Table 6
The comparison of classifier accuracies were given where the classifier algorithm was Naive Bayes. The maximum classifier accuracies were boldface and the number of features (or components) are given in parentheses to achieve this performance for each subject. Abbreviations of algorithms are PCA (Principal Component Analysis), LLE (Locally Linear Embedding), MDS (Multi-Dimensional Scaling), ISOMAP (Isometric Feature Mapping), and T-SNE (t-Distributed Stochastic Neighbor Embedding).

Subjects	PCA	Manifold learning methods			
		LLE	MDS	ISOMAP	T-SNE
Subject 1	**0.58** (24)	0.50 (21)	0.46 (15)	0.29 (9)	0.21 (2)
Subject 2	0.50 (24)	**0.54** (21)	0.46 (16)	0.35 (3)	0.42 (7)
Subject 3	**0.81** (20)	0.71 (11)	0.67 (16)	0.33 (3)	0.48 (13)
Subject 4	**0.71** (20)	0.52 (7)	0.48 (21)	0.43 (9)	0.48 (16)
Average	**0.65** (22)	0.57 (15)	0.52 (17)	0.35 (6)	0.40 (9.5)

relies on choosing appropriate machine learning (classification) algorithms. These algorithms aim automatically estimate the class of the data as represented by feature vectors [34,45]. In another word, the classification performance of the SSVEP-based BCI is directly affected by the machine learning algorithms. For that reason, in this study, SSVEP records have been evaluated with nine well-known machine learning algorithms [13,46–50]. These algorithms are explained below briefly:

- Naive Bayes: The Naive Bayes method is based on the Bayesian probability model [51]. The Naive Bayes classifier has a powerful assumption of independence between predictors [52]. According to this assumption, the probability of one feature does not affect the probability of the other. A small amount of training data is sufficient to estimate the required parameters. There are four subtypes of the Naive Bayes algorithm. These are Gaussian, Bernoulli, multinomial, and complement Naive Bayes algorithms.
- Multi-Layer Perceptron (MLP): The Multilayer Perceptron Algorithm (MLP) is a supervised learning algorithm that can learn a function by training on a dataset. It consists of input, output, and hidden layers [13]. The input layer receives the data, and the output layer makes decisions or predictions about the input. There is an arbitrary number of hidden layers between these two layers, which is the computational engine of MLP. It learns by modeling the correlation between inputs and outputs. While the model is being trained, the weights of the parameters are adjusted to minimize errors.
- K-Nearest Neighbors (K-NN): The k-nearest neighbor algorithm is a non-parametric method used in classification and regression [13]. The purpose of the KNN is to determine the class of an object whose class is unknown in the multidimensional feature space. A parameter "k" is user-defined in the algorithm. The class of the undefined object is determined by taking the most common class information from the "k" closest objects. When calculating the distance of objects from each other, the "Euclidean" distance metric is generally used.
- AdaBoost: The AdaBoost algorithm is an ensemble learning method that was proposed by Freund and Schapire [47]. The model can improve the classification accuracy of weak classifiers by changing the distribution of sample weights. The AdaBoost algorithm can transform a weak classifier that makes erroneous predictions into a new classifier with high classification accuracy.
- Logistic Regression: The first study in the field of Logistic Regression (LR) was made by Berkson [53] in 1944. Contrary to its name, the LR algorithm is generally more suitable for classification problems. The LR method is preferred when the outcome variable is a two or multi-level categorical variable [54].
- Decision Trees: Decision trees are one of the common algorithms in supervised learning. It has a predetermined target variable. It offers a top-down strategy by design. It breaks a data set into smaller pieces according to some rules. Decision trees can process both categorical and numerical data. The decision tree algorithm may differ according to the target variable type. The entropy and Gini algorithms are the most widely used [13].
- Support Vector Machines: Support Vector Machines (SVM) were first introduced by Vapnik [55] and it is one of the most used machine learning algorithms today. It is used in data classification and regression analysis. While generating a model, it places the data in a high-dimensional space and configures the hyperplane that will best separate this data in space [50].
- Stochastic Gradient Descent (SGD): The Gradient descent (GD) is used in learning widely. Mathematically, they can be defined as partial derivatives of an array of parameters with respect to their inputs. In the stochastic gradient descent (SGD) algorithm, several samples are randomly selected for each iteration, rather than all of the adjusted data [46].
- The Random Forests (RF): The Random Forests (RF) algorithm was developed by Leo Breiman for classification issues. Breiman [56] added the "randomness" layer in addition to the "bagging" layer he had previously developed in the Random Forest method. In standard trees, each node is separated using the best separator among all variables. On the other hand, in RF trees, the discrimination is made by the best classifier from a randomly selected sub-classifier set at that node. Thus, RF is a method that can produce better results compared to other classification methods and is more resistant to the over-fitting problem [49].

2.5. Performance verification

In literature, commonly used leave one out and confusion matrix evaluation criteria were used to assess the performance of the machine learning algorithms. The performance metric is the Accuracy (Acc) [13]. According to the metrics, the AVI-SSVEP dataset is split into two groups train and test data. The machine learning algorithms' parameters are decided with the train data where the algorithms' performances are predicted over the test data. In this scope of the study, we preferred to divide the dataset into 70% for training and 30% for

Table 7
Comparison of classifier performances with the similar studies from the literature.

Study	Classifier	Optimization	Number of Classes	Accuracy (%)
[21]	Multi-Layer Perceptron	Principal Component Analysis	2	53.3
[25]	Multi-Layer Perceptron	Landmark version of Maximum Variance Unfolding	2	89.6
[27]	k-Nearest Neighbors	Supervised Explicit Isometric Feature Mapping	2	92.7
[23]	Support Vector Machines	t-Distributed Stochastic Neighbor Embedding	2	94.1
[22]	Linear Discriminant Analysis	Multidimensional Scaling	2	95.0
[28]	Support Vector Machines	Dimensionality Reduction Mechanism	2	100
[26]	Multi-Layer Perceptron	Neighborhood Component Analysis	2	100
[30]	Linear Discriminant Analysis	Locality Preserving Projection	3	71.9
[24]	Linear Discriminant Analysis	Principal Geodesic Analysis	4	61.0
[29]	Linear Discriminant Analysis	Locally Linear Embedding	4	74.2
This study	**Naive Bayes**	**Principal Component Analysis**	**7**	**65.2**

testing. Besides, we iterated each machine learning algorithm ten times and averaged their success of them to reduce the effect of randomness in the performances.

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{15}$$

In this study, mentioned criteria above calculated with (15). In the formula, TP means true positive indicators of the correctly estimated samples true class; on the other hand, TN means true negative represents the correctly predicted samples, not true classes; in addition, FP presents false positives representing the number of samples estimated as positive where it belongs to negative class and finally, FN predicted as negative samples but belongs to the positive class.

3. Results and discussion

In this study, the multi-class classification problem of SSVEP signals was analyzed with 9 different machine learning algorithms via Orange Data Mining Programming and their performances were evaluated. The metrics and values selected for the parameters of feature reduction and machine learning algorithms are as follows:

- In the ISOMAP algorithm, the number of neighbors is chosen as 5,
- In the MDS algorithm, the max. iteration was determined as 300 and the initialization was determined as PCA (torgenson),
- In the t-SNE algorithm, distance metric was selected as Euclidean, Perplexity equal 30, early exaggeration 200, and max iteration equal 1000. In addition, PCA method were chosen as Initialization,
- The number of neighbors in the LLE algorithm was set to 10 and the max. iteration was set to 100,
- RBF kernel function is applied to the Support Vector Machine algorithm, and the iteration limit was set to 100,
- In the K-NN algorithm, the “k” value was chosen as 5. Euclidean was chosen as the distance metric,
- In the MLP algorithm, the Relu activation function was set to Adam solver and the maximum iteration was set to 200,
- Number of estimators 50 and learning rate 1 were preferred in Adaboost algorithm.

While composing feature matrices with feature reduction methods, the number of components in each dimension reduction method has been increased one by one. Thus, it has been proved that the maximum success rate is achieved with how many components. The highest performance metrics obtained for all feature reduction methods are given in Tables 1, 2, 3, 4 and 5 for Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (T-SNE), Multidimensional Scaling (MDS), Isometric Feature Mapping (ISOMAP), and Locally Linear Embedding (LLE) methods, respectively.

4. Conclusion

In this study, we developed an automated signal processing process to design an expert 7-class SSVEP-based BCI system. First, 55 different features were extracted using temporal, spatial, and spatio-temporal methods. Then, traditional and recently hot topic manifold learning methods were analyzed and compared to successfully classify 7 different frequencies with the most meaningful and least data. Finally, classification was made using 9 different machine learning methods. According to the studies in the literature, the result is quite successful for a 7-class study (Table 7). According to the classification results, PCA was found to be the most successful method for all 4 subjects. The accuracy rates were found between 0.50 and 0.81 using PCA. The accuracy obtained using manifold methods was compared with PCA, which is considered the golden ratio. However, there is not a method as successful as PCA. The LLE method is also seen as the most successful method after PCA. Although MDS gives better results than ISOMAP and t-SNE methods, it does not seem to be more successful than LLE for all 4 subjects. In this study, LLE was found to be the Manifold Learning Method with the best results. It has been concluded that the LLE method for SSVEP-based BCI studies is more successful than other manifold learning methods. In addition, in this study, all methods except the ISOMAP method had the best results for Subject 3. Among the classification methods, the Naive Bayes method was seen to be the most successful machine learning method among all feature reduction methods tested in this study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

Matjaž Perc was supported by the Slovenian Research Agency (Javna agencija za raziskovalno dejavnost Republike Slovenije) (Grant Nos. P1-0403 and J1-2457).

References

[1] D.J. McFarland, J.R. Wolpaw, Brain-computer interfaces for communication and control, Commun. ACM 54 (5) (2011) 60–66, http://dx.doi.org/10.1145/1941487.1941506.
[2] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, J.R. Wolpaw, BCI2000: a general-purpose brain-computer interface (BCI) system, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1034–1043, http://dx.doi.org/10.1109/TBME.2004.827072.
[3] J.J. Vidal, Real-time detection of brain events in EEG, Proc. IEEE 65 (1977) 633–641, http://dx.doi.org/10.1109/PROC.1977.10542.

[4] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T. Jung, S. Gao, High-speed spelling with a noninvasive brain-computer interface, Proc. Natl. Acad. Sci. USA 112 (44) (2015) E6058–67, http://dx.doi.org/10.1073/pnas.1508080112.

[5] Y.J. Wang, R.P. Wang, X.R. Gao, B. Hong, S. Gao, A practical VEP-based brain-computer interface, IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2006) 234–239, http://dx.doi.org/10.1109/TNSRE.2006.875576.

[6] D. Regan, Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine, Elsevier, New York, 1989.

[7] Y. Wang, X. Gao, B. Hong, C. Jia, S. Gao, Brain-computer interfaces based on visual evoked potentials, IEEE Eng. Med. Biol. 27 (5) (2008) 64–71, http://dx.doi.org/10.1109/MEMB.2008.923958.

[8] S. Gao, Y. Wang, X. Gao, B. Hong, Visual and auditory brain-computer interfaces, IEEE Trans. Biomed. Eng. 61 (5) (2014) 1436–1447, http://dx.doi.org/10.1109/TBME.2014.2300164.

[9] D. Regan, Comparison of transient and steady-state methods, Ann. NY Acad. Sci. 388 (1982) 45–71, http://dx.doi.org/10.1111/j.1749-6632.1982.tb50784.x.

[10] E. Basar, EEG-brain dynamics: Relation between EEG and brain evoked potentials, in: Brain Lang 1980, Elsevier, Amsterdam.

[11] E. Sayilgan, Y.K. Yuce, Y. Isler, Evaluation of wavelet features selected via statistical evidence from steady-state visually-evoked potentials to predict the stimulating frequency, J. Fac. Eng. Archit. Gazi Univ. 36 (2) (2021) 593–605, http://dx.doi.org/10.17341/gazimmfd.664583.

[12] E. Sayilgan, Y.K. Yuce, Y. Isler, Evaluation of mother wavelets on steady-state visually-evoked potentials for triple-command brain-computer interfaces, Turk. J. Electr. Eng. Comput. Sci. 29 (5) (2021) 2263–2279, http://dx.doi.org/10.3906/elk-2010-26.

[13] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, second ed., Wiley Interscience, 2000.

[14] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326, http://dx.doi.org/10.1126/science.290.5500.2323.

[15] J.B. Tenenbaum, V.D. Silvaand, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323, http://dx.doi.org/10.1126/science.290.5500.2319.

[16] R.R. Coifman, S. Lafon, A.B. Lee, M. Maggioni, B. Nadler, F. Warner, S.W. Zucker, Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps, Appl. Math. 102 (21) (2005) 7426–7431, http://dx.doi.org/10.1073/pnas.050033410.

[17] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) (2008) 2579–2605, https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf?fbcl.

[18] G. Hinton, S.T. Roweis, Stochastic neighbor embedding, in: Advances in Neural Information Processing Systems, Vol. 15, 2003, pp. 857–864, https://proceedings.neurips.cc/paper/2002/file/6150ccc6069bea6b5716254057a194ef-Paper.pdf.

[19] K. Li, H. Xu, X. Liu, Analysis and visualization of accidents severity based on LightGBM-TPE, Chaos Solitons Fractals 157 (2022) 111987, http://dx.doi.org/10.1016/j.chaos.2022.111987.

[20] B. Yesilkaya, M. Perc, Y. Isler, Manifold learning methods for the diagnosis of ovarian cancer, J. Comput. Sci. 63 (2022) 101775, http://dx.doi.org/10.1016/j.jocs.2022.101775.

[21] A. Tyagi, V. Nehra, A comparison of feature extraction and dimensionality reduction techniques for EEG-based BCI system, IUP J. Comput. Sci. 11 (1) (2017) 51.

[22] Y. You, W. Chen, T. Zhang, Motor imagery EEG classification based on flexible analytic wavelet transform, Biomed. Signal Process. Control 62 (2020) 102069, http://dx.doi.org/10.1016/j.bspc.2020.102069.

[23] M.A. Li, X.Y. Luo, J.F. Yang, Extracting the nonlinear features of motor imagery EEG using parametric t-SNE, Neurocomputing 218 (2016) 371–381, http://dx.doi.org/10.1016/j.neucom.2016.08.083.

[24] E. Krivov, M. Belyaev, Dimensionality reduction with isomap algorithm for EEG covariance matrices, in: 2016 4th International Winter Conference on Brain-Computer Interface, BCI, IEEE, pp. 1–4, http://dx.doi.org/10.1109/IWW-BCI.2016.7457448.

[25] M.A. Li, H.W. Xi, Y.J. Sun, Feature extraction and visualization of MI-EEG with L-MVU algorithm, in: World Congress on Medical Physics and Biomedical Engineering, Springer, Singapore, 2019, pp. 835–839, http://dx.doi.org/10.1007/978-981-10-9023-3_151.

[26] M.T. Sadiq, X. Yu, Z. Yuan, Exploiting dimensionality reduction and neural network techniques for the development of expert brain–computer interfaces, Expert Syst. Appl. 164 (2021) 114031, http://dx.doi.org/10.1016/j.eswa.2020.114031.

[27] M.A. Li, W. Zhu, H.N. Liu, J.F. Yang, Adaptive feature extraction of motor imagery EEG with optimal wavelet packets and SE-Isomap, Appl. Sci. 7 (4) (2017) 390, http://dx.doi.org/10.3390/app7040390.

[28] P. Tan, X. Wang, Y. Wang, Dimensionality reduction in evolutionary algorithms-based feature selection for motor imagery brain-computer interface, Swarm Evol. Comput. 52 (2020) 100597, http://dx.doi.org/10.1016/j.swevo.2019.100597.

[29] H. Bischof, F.Y.-T. Lee, R. Scherer, R. Leeb, A. Schlogl, G. Pfurtscheller, Feature mapping using pca. Locally linear embedding and isometric feature mapping for EEG-based brain computer interface, in: Digital Imaging in Media and Education, Proceedings of the 28th AAPR Workshop, Osterreichische Computer Gesellschaft, 2004, pp. 189–196.

[30] J.Q. Gan, Feature dimensionality reduction by manifold learning in brain-computer interface design, 2006.

[31] A. Vilic, AVI steady-state visual evoked potential (SSVEP) signals dataset, 2013, https://www.setzner.com/avi-ssvep-dataset/.

[32] S.N. Carvalho, T.B.S. Costa, L.F.S. Uribe, D.C. Soriano, G.F. Yared, L.C. Coradine, R. Attux, Comparative analysis of strategies for feature extraction and classification in SSVEP BCIs, Biomed. Signal Process. Control 21 (2015) 34–42, http://dx.doi.org/10.1016/j.bspc.2015.05.008.

[33] R. Zerafa, T. Camilleri, O. Falzon, K.P. Camilleri, To train or not to train? A survey on training of feature extraction methods for SSVEP-based BCIs, J. Neural Eng. 15 (05) (2018) 1001, http://dx.doi.org/10.1088/1741-2552/aaca6-e.

[34] E. Sayilgan, Y.K. Yuce, Y. Isler, Investigating the effect of flickering frequency pair and mother wavelet selection in steady-state visually-evoked potentials on two-command brain-computer interfaces, IRBM 43 (6) (2022) 594–603, http://dx.doi.org/10.1016/j.irbm.2022.04.006.

[35] E. Sayilgan, Y.K. Yuce, Y. Isler, Duragan Durum Gorsel Uyaran Potansiyellerinden Fourier Donusumu ile Uc Farklı Frekansın Kestirimi, Duzce Univ. Bilim Teknoloji Derg. 8 (4) (2020) 2337–2343, http://dx.doi.org/10.29130/dubited.716386.

[36] E. Sayilgan, Y.K. Yuce, Y. Isler, Frequency recognition from temporal and frequency depth of the brain-computer interface based on steady-state visual evoked potentials, J. Intell. Syst. Appl. 4 (1) (2021) 68–73, http://dx.doi.org/10.54856/jiswa.202105160.

[37] M. Degirmenci, E. Sayilgan, Y. Isler, Evaluation of Wigner–Ville distribution features to estimate steady-state visual evoked potentials' stimulation frequency, J. Intell. Syst. Appl. 4 (2) (2021) 133–136, http://dx.doi.org/10.54856/jiswa.202112178.

[38] M.B. Avci, R. Hamurcu, O.A. Bozbas, E. Gurman, A.E. Cetin, E. Sayilgan, Design of steady-state visually-evoked potential based brain-computer interface system, J. Intell. Syst. Appl. 5 (2) (2022) 86–89, http://dx.doi.org/10.54856/jiswa.202212214.

[39] E. Sayilgan, Y.K. Yuce, Y. Isler, Evaluating steady-state visually evoked potentials-based brain-computer interface system using wavelet features and various machine learning methods, in: Brain-Computer Interface, IntechOpen, London, United Kingdom, 2021, [Online]. Available: https://doi.org/10.5772/intechopen.98335.

[40] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometr. Intell. Lab. Syst. 2 (1–3) (1987) 37–52, http://dx.doi.org/10.1016/0169-7439(87)80084-9.

[41] M.B. Selek, B. Yesilkaya, S.S. Egeli, Y. Isler, The effect of principal component analysis in the diagnosis of congestive heart failure via heart rate variability analysis, Proc. Inst. Mech. Eng. H 235 (12) (2021) 1479–1488, http://dx.doi.org/10.1177/09544119211036806.

[42] W.S. Torgerson, Multidimensional scaling: I. Theory and method, Psychometrika 17 (1952) 401–419, http://dx.doi.org/10.1007/BF02288916.

[43] M.C. Hout, M.H. Papesh, S.D. Goldinger, Multidimensional scaling, WIREs Cogn. Sci. 4 (2013) 93–103, http://dx.doi.org/10.1002/wcs.1203.

[44] B. Turhan, Nonlinear Dimensionality Reduction Methods for Pattern Recognition (diss. Master's thesis), 2004.

[45] D.D. Lopes, B.R. da Cunha, A.F. Martins, S. Gocalves, E.K. Lenzi, S. Hanley, M. Perc, H.V. Ribeiro, Machine learning partners in criminal networks, Sci. Rep. 12 (2022) 15746, http://dx.doi.org/10.1038/s41598-022-20025-w.

[46] L. Bottou, Stochastic gradient learning in neural networks, Proc. Neuro-Nimes 91 (8) (1991) 12, https://leon.bottou.org/publications/pdf/nimes-1991.pdf.

[47] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. System Sci. 55 (1997) 119–139, http://dx.doi.org/10.1006/jcss.1997.1504.

[48] M. Barstugan, R. Ceylan, The effect of dictionary learning on weight update of AdaBoost and ECG classification, J. King Saud Univ. Comput. Inf. Sci. 32 (2020) 1149–1157, http://dx.doi.org/10.1016/j.jksuci.2018.11.007.

[49] G. Biau, E. Scornet, A random forest guided tour, TEST 25 (2016) 197–227, http://dx.doi.org/10.1007/s11749-016-0481-7.

[50] C.M. Bishop, Pattern recognition, Mach. Learn. 128 (9) (2006) 2006, https://www.springer.com/gp/book/9780387310732.

[51] H. Zhang, The optimality of naive Bayes proc FLAIRS, 2004, https://www.cs.unb.ca/hzhang/publications/FLAIRS04ZhangH.pdf.

[52] S. Mukherjee, N. Sharma, Intrusion detection using naive Bayes classifier with feature reduction, Proc. Technol. 4 (2012) 119–128, http://dx.doi.org/10.1016/j.protcy.2012.05.017.

[53] J. Berkson, Application of the logistic function to bio-assay, J. Amer. Statist. Assoc. 39 (227) (1944) 357–365, http://dx.doi.org/10.1080/01621459.1944.10500699.
[54] A.C. Gok, A. Ozdemir, Forecasting the sector portions of banks by logistic regression analysis, Dokuz Eylul Univ. Isletme Fak. Derg. 12 (1) (2011) 43–51, https://dergipark.org.tr/en/download/article-file/234828.
[55] V. Vapnik, The Nature of Statistical Learning, Springer-Verlag, New York, 1995, https://books.google.com.tr/books?id=EqgACAAAQBAJ.
[56] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, http://dx.doi.org/10.1023/A:1010933404324.

Bartu Yesilkaya graduated from the Dokuz Eylül University with B.Sc. Electrical and Electronics Engineering degree in 2016. In 2020, he achieved M.Sc. degree from department of Biomedical Technologies in Izmir Katip Celebi University. Since 2018, he has been working in Biomedical Engineering as a research assistant at the same university. He is currently studying to achieve Phd degree at the department of biomedical technologies and his research interests are biomedical signals and machine learning algorithms.

Ebru Sayilgan is an assistant professor at the Department of Mechatronics Engineering, Izmir University of Economics, Turkey. She received the Ph.D. degree in the Department of Biomedical Technologies from Izmir Katip Celebi University in 2020. Her current research interests include biomedical signal processing, pattern recognition, machine/deep learning, medical robotics, and medical device design.

Yilmaz Kemal Yuce is assistant professor of Computer Engineering in Alanya Alaaddin Keykubat University. His main research interests include e-health, human–computer interaction, usability, process communication patterns and social computing. In addition to academic background, he is a single dad who loves spending time with his daughter. He is an IEEE and TurkMIA member.

Matjaž Perc is professor of physics at the University of Maribor. He is a member of Academia Europaea and the European Academy of Sciences and Arts, and among top 1 most cited physicists according to 2020 and 2021 Clarivate Analytics data. He is also the 2015 recipient of the Young Scientist Award for Socio and Econophysics from the German Physical Society, and the 2017 USERN Laureate. In 2018 he received the Zois Award, which is the highest national research award in Slovenia. In 2019 he became Fellow of the American Physical Society. Since 2021 he is Vice Dean of Natural Sciences at the European Academy of Sciences and Arts.

Yalcin Isler is associate professor of Biomedical Engineering at Izmir Katip Celebi University. His main research interests cover biomedical signal processing, pattern recognition, medical device design, programming, and embedded systems. In addition to academics, he founded Islerya Medical and Information Technologies Company, Izmir, Turkey, in 2017. He was an IEEE Student Member (2009–2011) and IEEE Member (2011–2014) and has been IEEE Senior Member since 2014.