Heliyon 9 (2023) e15258

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Electroencephalography (EEG) eye state classification using learning vector quantization and bagged trees

Mehrbakhsh Nilashi [a,b,*], Rabab Ali Abumalloh [c], Hossein Ahmadi [d], Sarminah Samad [e], Abdullah Alghamdi [f], Mesfer Alrizq [f], Sultan Alyami [g], Fatima Khan Nayer [h]

[a] UCSI Graduate Business School, UCSI University, No. 1 Jalan Menara Gading, UCSI Heights, 56000, Cheras, Kuala Lumpur, Malaysia
[b] Centre for Global Sustainability Studies (CGSS), Universiti Sains Malaysia, 11800 Penang, Malaysia
[c] Department of Computer Science and Engineering, Qatar University, Doha 2713, Qatar
[d] Centre for Health Technology, Faculty of Health, University of Plymouth, Plymouth PL4 8AA, UK
[e] Department of Business Administration, College of Business and Administration, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia
[f] Information Systems Department, College of Computer Science and Information Systems, Najran University, Najran, Saudi Arabia
[g] Computer Science Department, College of Computer Science and Information Systems, Najran University, Najran, Saudi Arabia
[h] Artificial Intelligence and Data Analytics (AIDA) Research Lab, College of Computer and Information Sciences, Prince Sultan University, Saudi Arabia

ARTICLE INFO

Keywords:
Electroencephalography
Human cognition state
Bagged trees
Learning vector quantization
Eye state classification

ABSTRACT

The analysis of Electroencephalography (EEG) signals has been an effective way of eye state identification. Its significance is highlighted by studies that examined the classification of eye states using machine learning techniques. In previous studies, supervised learning techniques have been widely used in EEG signals analysis for eye state classification. Their main goal has been the improvement of classification accuracy through the use of novel algorithms. The trade-off between classification accuracy and computation complexity is an important task in EEG signals analysis. In this paper, a hybrid method that can handle multivariate signals and non-linear is proposed with supervised and un-supervised learning to achieve a fast EEG eye state classification with high prediction accuracy to provide real-time decision-making applicability. We use the Learning Vector Quantization (LVQ) technique and bagged tree techniques. The method was evaluated on a real-world EEG dataset which included 14976 instances after the removal of outlier instances. Using LVQ, 8 clusters were generated from the data. The bagged tree was applied on 8 clusters and compared with other classifiers. Our experiments revealed that LVQ combined with the bagged tree provides the best results (Accuracy = 0.9431) compared with the bagged tree, CART (Classification And Regression Tree) (Accuracy = 0.8200), LDA (Linear Discriminant Analysis) (Accuracy = 0.7931), Random Trees (Accuracy = 0.8311), Naïve Bayes (Accuracy = 0.8331) and Multilayer Perceptron (Accuracy = 0.7718), which demonstrates the effectiveness of incorporating ensemble learning and clustering approaches in the analysis of EEG signals. We also provided the time complexity of the methods for prediction speed (Observation/ Second). The result showed that LVQ + Bagged Tree provides the best result for prediction speed (58942 Obs/Sec) in relation to Bagged Tree (28453 Obs/Sec), CART (27784 Obs/Sec), LDA

* Corresponding author. UCSI Graduate Business School, UCSI University, No. 1 Jalan Menara Gading, UCSI Heights, 56000, Cheras, Kuala Lumpur, Malaysia.
E-mail address: mehrbakhsh@ucsiuniversity.edu.my (M. Nilashi).

https://doi.org/10.1016/j.heliyon.2023.e15258
Received 24 January 2023; Received in revised form 30 March 2023; Accepted 31 March 2023
Available online 5 April 2023
2405-8440/

(26435 Obs/Sec), Random Trees (27921), Naïve Bayes (27217) and Multilayer Perceptron (24163).

1. Introduction

The ability to decode human intentions using neural activity has a lot of practical applications [1–3]. Various neuroimaging modalities (e.g., magnetoencephalography, electroencephalography, functional near-infrared spectroscopy, and functional magnetic resonance imaging) have been used in the development of decoding applications based on neural activity. Electroencephalography or EEG has been widely used in various identification systems [4,5]. One of the applications of EEG has been in eye state identification [6]. Distinguishing between open and closed states of the eyes in real-life situations using EEG signals is a difficult research goal that is important for medical care and daily life tasks. Eye state identification [3,7,8] is a type of common time-series classification problem that has recently received a lot of attention in the research community [8,9]. EEG [10] is a technique commonly used in eye state classification to determine a human's cognition state [9]. There have been several successful applications of EEG eye state classification in the literature which are in areas of, for example, driving drowsiness detection [11], infant sleep-waking state identification [12], emotional arousal detection [13], personal authentication [14] and driver alertness monitoring [15].

Previous studies on EEG analysis have shown that machine learning is effective in eye state classification [3,8,16]. The prior research on EEG analysis for eye state classification widely have applied supervised learning techniques such as support vector machines [17,18], deep learning [19–22], neural network [23], and so on. These studies have mainly attempted to improve the accuracy of eye state classification. The trade-off between classification accuracy and computation complexity is an important task in EEG signals analysis which is rarely explored in previous studies. In this study, a hybrid method that can handle multivariate signals and non-linear is proposed with supervised and un-supervised learning to achieve a fast EEG eye state classification with high prediction accuracy which provides real-time decision-making applicability. We use the Learning Vector Quantization (LVQ) technique and bagged tree techniques. The method aims to improve both classification accuracy and time complexity of EEG signals (prediction speed) for eye state identification. To the best of our knowledge, there is no study to combine ensemble learning with learning vector quantization for improving the efficiency of eye identification systems using the analysis of EEG signals.

The remainder of our work is as follows. A literature review is presented in Section 2. In Section 3, we present the techniques used in this work. We introduce the LVQ and bagged tree techniques in this section. In Section 4, method evaluation is performed. In this section, the dataset is introduced and analyzed and the method evaluation is performed and compared with the other classifiers. In Section 5, we present the conclusion. The overall results are provided along with the shortcomings and future works.

2. Literature review

Research on EEG classification has been carried out in diverse directions based on the application domain of the study. The literature has deployed several approaches in the feature extraction and classification processes and utilized diverse datasets. Varli and Yilmaz [24] deployed a hybrid deep learning approach that entails Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) methods for EEG signals classification and the diagnosis of epileptic seizure activity. For epileptic seizure detection also, Amin, Yusoff and Ahmad [25] deployed Machine Learning (ML) approaches and used Computer Aided Diagnostic (CAD) approach to distinguish between normal signals and epileptic seizures. Liu, Shi, Hui, Xu, Wang, Na, Sun, Ding, Zheng and Chen [26] adopted a temporal and channel attention CNN approach for MI-EEG classification. The results of the study indicated an improvement in the performance of the proposed method over other states of art approaches in terms of efficiency and accuracy. In the context of sentence classification, Keles, Yildiz, Barua, Dogan, Baygin, Tuncer, Demir, Ciaccio and Acharya [27] deployed a new graph-based signal classification model by utilizing the testosterone chemical shape. The proposed approach indicated the superiority of the performance in comparison with other approaches in terms of classification accuracy. In the research by Al-Salman, Li, Oudah and Almaged [28], the authors deployed a new approach based on *k*-means clustering and the Least Square Support Vector Machine (LS-SVM) classification approach to classify the sleep stages. The results of the study indicated a satisfactory accuracy of the classification approach with a ratio of 97.4%. Another study that focused on the classification of the sleep stage is presented by Mousavi, Rezaii, Sheykhivand, Farzamnia and Razavi [29] by deploying a Deep CNN and utilizing the Sleep-EDF dataset. The study deployed a new single-channel approach to address the high computational issue for discriminative feature extraction and selection. In terms of classification accuracy, the results presented an enhanced performance when compared to other approaches. Another research by Michielli, Acharya and Molinari [30] focused on sleep stage classification and deployed LSTM and RNN approaches. The research also utilized a single-channel approach to the Sleep-EDF dataset. In terms of the neurocognitive performance evaluation, the research presented satisfactory results. Amin, Alsulaiman, Muhammad, Mekhtiche and Hossain [31] proposed new multi-layer CNN for EEG motor imagery classification. When compared with other ML and Deep Learning (DL) approaches, the proposed approach presented a better performance in terms of classification accuracy, with ratios of 75.7% and 95.4% on the deployed datasets. Focusing on phase synchronization, Li, Fan, Wang and Wang [32] introduced an enhanced approach for EEG signals classification using the Common Spatial Patterns (CSP) approach. The deployment of the proposed method on three datasets proved the efficiency of the approach. San-Segundo, Gil-Martín, D'Haro-Enríquez and Pardo [33] deployed a Deep Neural Network (DNN) approach for the classification of epileptic EEG recordings, with an enhancement in the outcomes over the results obtained from other approaches. In the context of stress classification, the study by Asif, Majid and Anwar [34] deployed four classification algorithms; namely Multilayer Perceptron

Table 1
EEG classification in previous literature.

Ref.	Context	Data sets	DL	ANN	DNN	CNN	LSTM	Graph-based	LS-SVM	Clustering	SDG	SMO	MP	LR	CSP	SVM	RF	KNN	MLP	RNN	ML	WT
[24]	Epileptic seizure diagnosis	-Bonn dataset -Bern-Barcelona dataset -CHB-MIT dataset	✓			✓	✓															
[26]	MI-EEG	-BCIC IV 2a dataset -HGD dataset				✓																
[27]	EEG sentence classification	-EEG sentence datasets						✓														
[28]	Sleep stage classification	Two EEG datasets							✓	✓												
[29]	Sleep stage classification	Sleep-EDF dataset	✓			✓																
[31]	EEG motor imagery classification	-BCI Competition IV-2a dataset -High Gamma Dataset				✓																
[16]	-Eye State classification	-UCI Machine Learning Repository		✓										✓								
[32]	Phase synchronization	-BCI Competition IIa dataset -BCI Competition I dataset -BCI Competition III dataset													✓							
[33]	Epileptic EEG recordings	-Bern-Barcelona EEG dataset -Epileptic Seizure Recognition dataset			✓	✓																
[34]	Stress classification	Developed by the authors									✓	✓	✓	✓								
[35]	Stress classification	Developed by the authors												✓		✓	✓	✓	✓			
[30]	Sleep stage classification	Sleep-EDF database					✓													✓		
[25]	Epileptic seizure diagnosis	-Bonn dataset																			✓	
[37]	Brain electrical activity	-EEG data																				✓

Deep Learning: DL, Artificial Neural Networks: ANN, Convolutional Neural Networks: CNN, Long-Short Term Memory: LSTM, Motor Imagery Electroencephalogram: MI-EEG, Least Square Support Vector Machine: LS-SVM, Common Spatial Patterns: CSP, Deep Neural Network: DNN, Logistic Regression: LR, Multilayer Perceptron: MP, Sequential Minimal Optimization: SMO, Stochastic Decent Gradient: SDG, Support Vector Machine: SVM, Random Forest: RF, K-Nearest Neighbors: KNN, Multi-Layer Perceptron: MLP, Recurrent Neural Network: RNN, Machine Learning: ML, Wavelet Transform: WT.

(MP), Logistic Regression (LR), Stochastic Decent Gradient (SDG), and Sequential Minimal Optimization (SMO). The study focused on the impact of music tracks on the level of stress using EEG signals. The study revealed that LR has better performance in terms of classification accuracy. The study also examined the influences of different factors on stress; gender, music genres, and music tracks. In an experimental study, the author in Ref. [35] focused also on stress classification through the lenses of Power Spectral Density (PSD). For the classification process, the researchers deployed several techniques including LR, SVM, Random Forest (RF), K-Nearest Neighbors (KNN), and Multi-Layer Perceptron (MLP). The results of the study indicated better performance with respect to stress detection by SVM, RF, and MLP approaches. Although both works by Asif, Majid and Anwar [34] and Perez-Valero, Lopez-Gordo and Vaquero-Blasco [35] deployed several techniques for stress classification following an experimental design, their results were different. Contrary to the work by Ref. [34], the results by Perez-Valero, Lopez-Gordo and Vaquero-Blasco [35] indicated the worst performance for the LR approach. Much effort has been made to deploy effective discriminating statistical quantifiers focusing on analyzing the time series [36]. To analyze EEG time series, a new approach for EEG evaluation based on the wavelet transform was deployed by Ref. [37] aiming to address the qualitative variation of the signal in terms of frequency and time. In the context of eye state classification, the author in Ref. [16] deployed LR and ANN with an overall accuracy of 88.2%. Table 1 presents the summary of previous literature on EEG classification.

3. Method

This study employed LVQ and bagged tree techniques for eye state identification through the use of EEG signals. In this section, we present the techniques used in this work. We introduce the bagged tree and LVQ techniques in this section.

3.1. Bagged trees

A machine learning technique called ensemble combines different learning algorithms to produce a model with improved predictive performance over the model's individual components [38,39]. AdaBoost (adaptive boosting) [40], Boosting [41], Gradient boosting machines [42], Stacked generalization (blending or stacking) [43], Random Forest [44] and Bootstrapped aggregation (Bagging) [45,46] are some examples. What types of weak learners to combine and how to aggregate their outputs are two crucial design decisions. Decision Trees are supervised learning algorithms that are commonly used to create an explainable machine learning algorithm. An algorithm that has a tree-like structure and is constructed using sets of mutually exclusive if-then rules is known as a decision tree. The most crucial rules are usually kept close to the tree's root, and these if-then statements are drawn from the malware and benign training datasets fed to the model.

The high variance that decision trees experience is a drawback. This means that the outcomes we get could be very different if we divide the training data into two parts at random and fit a decision tree to each half. A Bootstrap aggregation technique called bagging is used to lower the variance of a statistical learning method. Bagged Trees is a machine-learning approach that aims to improve the accuracy of various types of models [46]. In this case, it is attempted to boost the accuracy of the constructed Decision Tree models. Bagging is the process of aggregating classifiers over bootstrap datasets, in which a dataset D is divided into multiple smaller datasets S_i, each with n values. These n sample values could be replaced in one of the other S_i sets. After this step, the ensemble classifier generates a number of fine decision trees. The model is then fed the test set, and one sample is run through each of the decision trees it generates. The model tallies the classification outcomes of each decision tree. The ensemble algorithm can then use a plurality rule to determine the classification of a given sample once it has reached this stage. A bagged tree has the potential to be more accurate than a simple decision tree because it has the ability to completely delete or erase the decision tree that produced an unstable result, improving the accuracy of the subsequent sample and the model as a whole.

We define $\widehat{f}(x)$ as a decision tree built from a dataset with predictors in x. Thus, $\widehat{f}^{1}(x), \widehat{f}^{2}(x), \ldots, \widehat{f}^{Q}(x)$ can be calculated using Q separate training sets and the average of them to minimize variance. In general, access to multiple training sets is limited, so bagging relies on Bootstrapping approach. By generating Q different bootstrapped training data sets and then training the data sets to obtain $\widehat{f}^{*q}(x)$. Finally, by taking the average of all predictions, we can define bagging as in Eq. (1).

$$\widehat{f}_{bag}(x) = \sum_{q=1}^{Q} \widehat{\mathrm{f}}^{*q}(\mathrm{x}) \tag{1}$$

Because it deals with the problem with high variance, bagging is frequently used in conjunction with decision trees. It is possible to make an overall prediction by building Q decision trees based on Q bootstrapped training sets; the mean predicted value for an observation corresponds to the predicted probability. The model classifier uses a selected threshold to classify the observation. It is common to use a threshold of 0.5, which causes the model to assign an observation to the Q predictors' most frequently predicted class; this is referred to as majority voting.

3.2. LVQ

In this study, the data clustering is performed using the LVQ technique. Clustering has been effective in handling large datasets for different applications [47–53]. LVQ algorithms are regarded as interpretable machine learning techniques. The initial LVQ models developed by Refs. [54,55] are heuristic methods driven by Bayes decision theory and vector quantization. Heuristic implies that we do

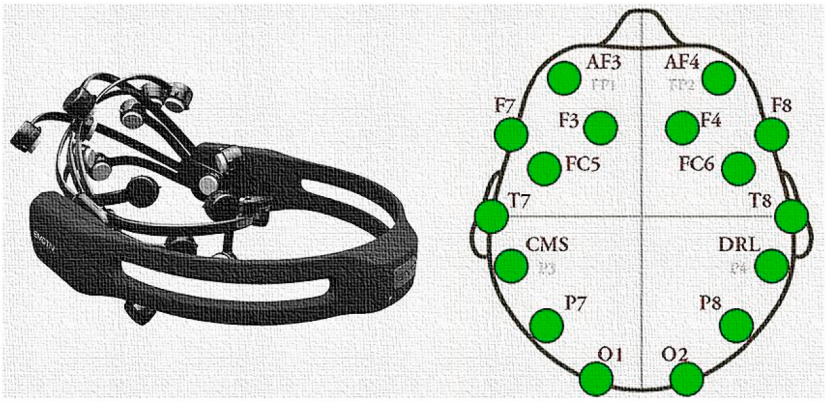

Fig. 1. Location of 14 electrodes of Emotiv EEG device.

not directly optimize a loss function. The simplest LVQ scheme is method LVQ1, which has the pseudocode in Algorithm 1. In order to ensure that each class in the dataset is represented by at least one prototype, we first define the set of prototypes W. A suitable scheme is then used to initialize the prototypes. As random vectors, for instance, or as randomly selected data samples from the pertinent class, or as k-means over the data samples from the pertinent class, and so forth. The next two steps are repeated:

i. We first randomly chose a training sample $(x, c(x))$ from the training dataset D and choose the most similar prototype w^* through Eq. (2).

$$w^*(x) = argmin_{w_k = W}\ d(x, w_k) \tag{2}$$

ii. We push the prototype w^* slightly in the direction of x (attraction) and pull it slightly in the opposite direction (repulsion) if the input's class label $c(x)$ and the closest prototype's class label $c(w^*)$ are equal.

A learning rate $\eta > 0$ controls the applied shift's magnitude. We apply the Best Matching Prototype Principle (BMPP) to categorize random data points after the model has been trained. The Euclidean distance d_E is typically used by LVQ1 in accordance with Eq. (3).

$$d_E(x, y) = \sqrt{(x - y)^T (x - y)} \tag{3}$$

With respect to a prototype, the gradient of d_E is calculated by Eq. (4):

$$\nabla_w d_E(x, w) = \frac{1}{d_E(x, w)}(x - w) \text{ if } d_E(x, w) \neq 0 \tag{4}$$

Algorithm 1: LVQ1 Procedure
1. Inputs
2. $(T,$ dissimilarity measure $d,$ prototypes $W,$ learning rate $\eta,$ number of steps $N)$
3. $\mathscr{W} \leftarrow$ initialize the set of prototypes
4. $i \leftarrow 0$
5. While $i < N$ **Do**
6. $i \leftarrow i + 1$
7. $\mathrm{x},\ \mathrm{c(x)} \leftarrow$ randomly select a training sample from D
8. $\mathrm{w}^* \leftarrow \mathrm{w}^*(x)$
9. If $c(\mathrm{w}^*) = c(x)$ **Then**
10. $s \leftarrow -1$
11. Else
12. $s \leftarrow 1$
13. $\triangle \mathrm{w}^* \leftarrow s(\mathrm{x} - \mathrm{w}^*)$
14. $\mathrm{w}^* \leftarrow \mathrm{w}^* — \eta \triangle \mathrm{w} *$
15. Return $\mathscr{W}$ (return the trained prototypes)

4. Method evaluation

In this section, method evaluation is performed. We present the dataset. The method evaluation is performed and the results are compared with the other classifiers.

4.1. Dataset

The data corpus used in this study was gathered and compiled by Roesler and made available for public use on the UCI data repository [56]. This dataset is widely used for eye state identification [8,23]. The dataset consists of unprocessed electromagnetic

Table 2
Dataset information.

Descriptive Statistics (Eye-Open)						
	N	Minimum	Maximum	Mean		Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
AF3	8254	4197.95	4504.10	4298.0686	.44746	40.65200
F7	8254	3924.10	4156.92	4012.7937	.34821	31.63500
F3	8254	4197.44	4386.15	4262.8903	.24437	22.20171
FC5	8254	4073.33	4250.26	4123.3197	.22636	20.56495
T7	8254	4304.62	4463.59	4341.5613	.17701	16.08134
P7	8254	4566.15	4756.92	4621.1798	.21219	19.27816
O1	8254	4027.18	4178.46	4071.9733	.19597	17.80437
O2	8254	4567.2	4731.8	4614.925	.2005	18.2165
P8	8254	4152.3	4320.0	4200.308	.1890	17.1726
T8	8254	4152.82	4362.56	4229.4507	.21929	19.92239
FC6	8254	4100.00	4332.31	4200.0248	.27390	24.88416
F4	8254	4201.03	4397.95	4276.9300	.23122	21.00663
F8	8254	4443.08	4833.85	4602.1036	.36759	33.39582
AF4	8254	4205.64	4573.33	4356.6255	.44705	40.61546
Descriptive Statistics Eye-Closed)						
	N	**Minimum**	**Maximum**	**Mean**		**Std. Deviation**
	Statistic	**Statistic**	**Statistic**	**Statistic**	**Std. Error**	**Statistic**
AF3	6722	4198.97	4445.13	4305.4429	.40806	33.45631
F7	6722	3905.64	4138.97	4005.4726	.33592	27.54161
F3	6722	4212.31	4367.18	4265.5488	.24547	20.12595
FC5	6722	4058.46	4214.36	4121.2210	.25996	21.31324
T7	6722	4309.74	4435.38	4341.5613	.22053	18.08085
P7	6722	4574.87	4708.72	4618.6865	.21273	17.44129
O1	6722	4026.15	4167.18	4073.8663	.29449	24.14483
O2	6722	4567.7	4695.9	4616.849	.2249	18.4396
P8	6722	4147.7	4287.7	4202.611	.2263	18.5517
T8	6722	4174.36	4323.08	4233.3547	.23610	19.35768
FC6	6722	4130.77	4319.49	4204.7597	.28914	23.70626
F4	6722	4225.64	4368.72	4281.7420	.22655	18.57421
F8	6722	4510.26	4811.28	4610.8060	.39992	32.78862
AF4	6722	4246.15	4552.82	4367.0500	.42470	34.81985

Table 3
Correlations among the features.

	F4	F3	FC6	T8	P8	F8	T7	AF4	AF3	O2	P7	FC5	O1	F7
F4	1.000	0.833	0.836	0.764	0.669	0.823	0.533	0.838	0.797	0.592	0.508	0.565	0.583	0.377
F3	0.833	1.000	0.667	0.640	0.563	0.624	0.644	0.708	0.755	0.546	0.595	0.765	0.491	0.565
FC6	0.836	0.667	1.000	0.801	0.678	0.843	0.533	0.741	0.650	0.594	0.495	0.407	0.513	0.250
T8	0.764	0.640	0.801	1.000	0.828	0.714	0.629	0.589	0.501	0.724	0.653	0.401	0.574	0.177
P8	0.669	0.563	0.678	0.828	1.000	0.537	0.640	0.389	0.319	0.867	0.708	0.368	0.667	0.143
F8	0.823	0.624	0.843	0.714	0.537	1.000	0.361	0.864	0.758	0.415	0.310	0.361	0.385	0.248
T7	0.533	0.644	0.533	0.629	0.640	0.361	1.000	0.299	0.349	0.655	0.835	0.680	0.662	0.491
AF4	0.838	0.708	0.741	0.589	0.389	0.864	0.299	1.000	0.943	0.285	0.214	0.473	0.334	0.414
AF3	0.797	0.755	0.650	0.501	0.319	0.758	0.349	0.943	1.000	0.225	0.231	0.606	0.318	0.585
O2	0.592	0.546	0.594	0.724	0.867	0.415	0.655	0.285	0.225	1.000	0.721	0.357	0.645	0.106
P7	0.508	0.595	0.495	0.653	0.708	0.310	0.835	0.214	0.231	0.721	1.000	0.544	0.657	0.329
FC5	0.565	0.765	0.407	0.401	0.368	0.361	0.680	0.473	0.606	0.357	0.544	1.000	0.396	0.750
O1	0.583	0.491	0.513	0.574	0.667	0.385	0.662	0.334	0.318	0.645	0.657	0.396	1.000	0.259
F7	0.377	0.565	0.250	0.177	0.143	0.248	0.491	0.414	0.585	0.106	0.329	0.750	0.259	1.000

recordings made from a participant's scalp and data on the participant's eye state (eyes open or closed) during the same time period. With the Emotiv EEG neuro headset, which is depicted in Fig. 1, a single continuous EEG measurement was used to generate all of the data. The dataset contains 14980 patterns and 14 features (see Table 2), where the 14 features are data collected by the 14 sensors depicted in Fig. 1. The correlations among the features are shown in Table 3. In Fig. 2, the features are visualized using two principal components (PC1 and PC2) [57–59] (see Table 4). The measurement lasted for 117 s. During the EEG measurement, the eye state was picked up using a camera, and after reviewing the video frames, it was manually added to the file.

Three instances of the numbers 899, 10387, 11510, and 13180 in this eye state corpus have obvious errors, making them outliers that should be removed before the experiments. As a result, only 14976 instances were used in the experiments performed by Roesler. In the corpus output, "0", for the eye-open state, and "1" stands for the eye-closed state. There are 8254 legal instances of eyes open and 6722 legal instances of eyes closed.

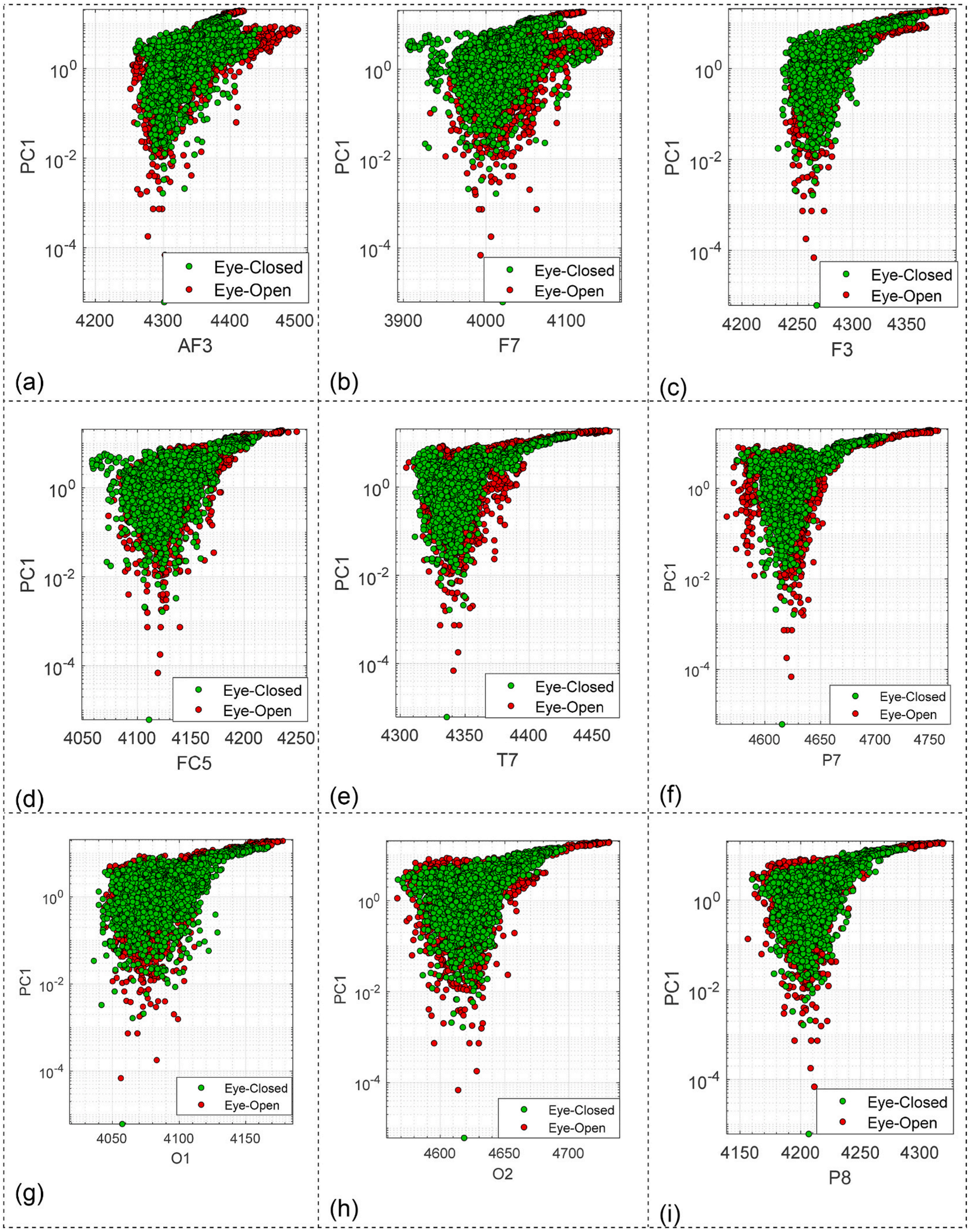

Fig. 2. Visualization of features for (a) F7-PC1, (b) F3-PC1, (c) FC5-PC1, (d) T7-PC1, (e) P7-PC1, (f) O1-PC1, (g) O2-PC1, (h) P8-PC, (i) T8-PC1, (j) FC6-PC1, (k) F4-PC1, (l) F8-PC1, (m) AF4-PC1, and (n) PC2-PC1.

4.2. LVQ and bagged trees results

We evaluated the proposed method on the EEG eye state dataset. The method was run on Microsoft Windows 10 Pro and a laptop with Processor Intel(R) Core(TM) i7-6700HQ CPU @ 2.60 GHz, 2592 Mhz, 4 Core(s), and 8 Logical Processor(s). Cross Validation is one of the most popular techniques for assessing model performance on in-sample estimates [52,60–65]. The entire training sample is used in the 5-fold cross-validation, which divides the sample into 5 roughly equal portions. The model trained on the remaining 4 folds will be tested on each fold in the fifth fold. We demonstrate a five-fold cross-validation example in Fig. 3. The percentage of samples

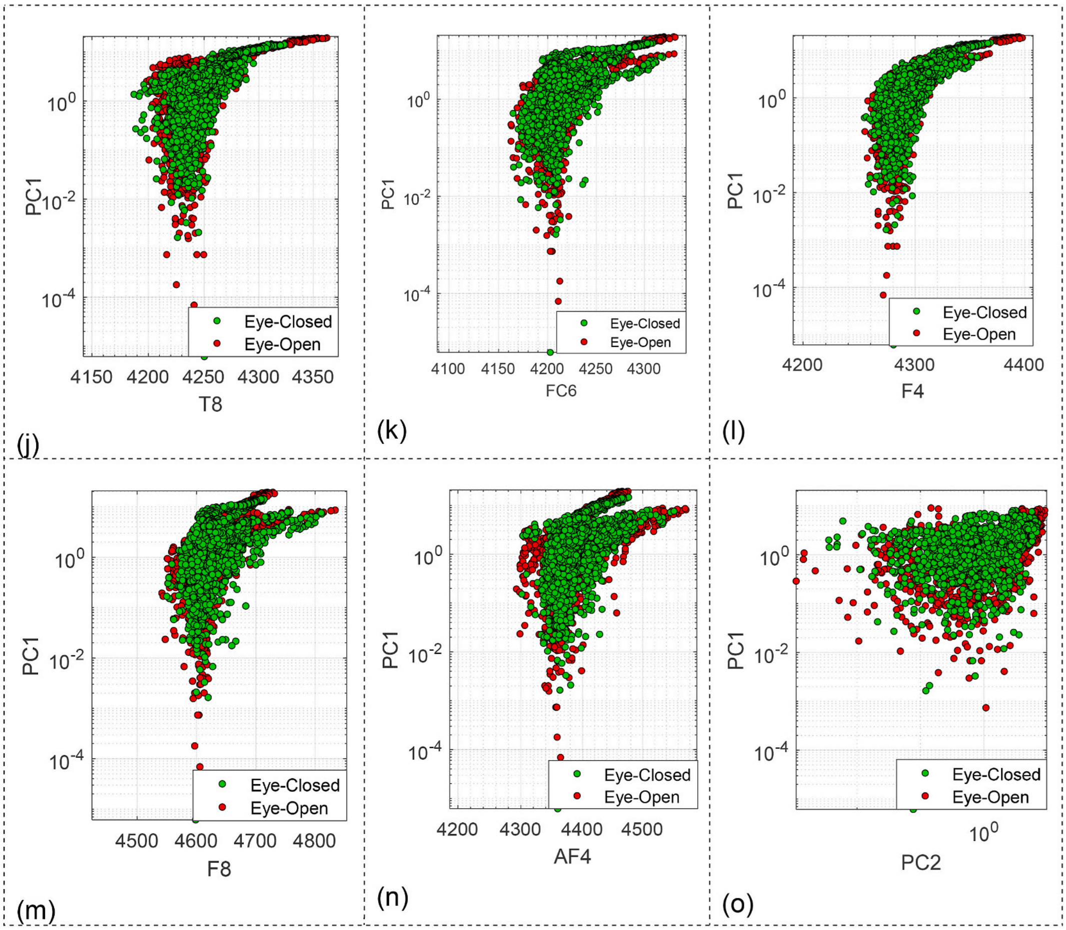

Fig. 2. (*continued*).

Table 4
PCA results.

Axis	Eigen value	Difference	Proportion (%)	Cumulative (%)
1	8.304187	6.177272	59.32%	59.32%
2	2.126915	0.528787	15.19%	74.51%
3	1.598128	1.146541	11.42%	85.92%
4	0.451587	0.119461	3.23%	89.15%
5	0.332127	0.055693	2.37%	91.52%
6	0.276434	0.093429	1.97%	93.50%
7	0.183005	0.013030	1.31%	94.80%
8	0.169975	0.026441	1.21%	96.02%
9	0.143534	0.016425	1.03%	97.04%
10	0.127109	0.027207	0.91%	97.95%
11	0.099902	0.010210	0.71%	98.66%
12	0.089692	0.018652	0.64%	99.30%
13	0.071039	0.044673	0.51%	99.81%
14	0.026366	–	0.19%	100.00%
Tot.	14.000000	–	–	–

from each class is preserved when using the stratified method, and a random fold will choose observations at random from the entire set. A dataset that is used as a training set and another set that is used to validate the trained model will be produced by the effect.

LVQ was applied to segment the data for eye state identification in bagged trees. The number of clusters per the dataset classes was set to 4. In addition, the learning rate in LVQ was set to 0.005. Distance normalization was performed by the variance. According to the pre-defined number of clusters, totally 8 clusters were generated from the data. The weights for global nodes and the nodes for clusters (eye-open) and clusters (eye-closed) are shown in Fig. 4. The clusters are visualized in Fig. 5 using three principal components (PC1,

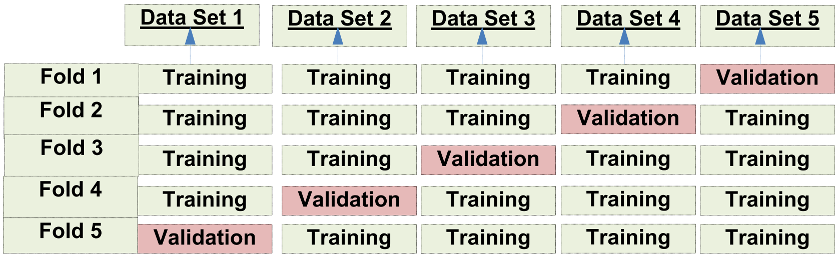

Fig. 3. 5-Fold cross-validation.

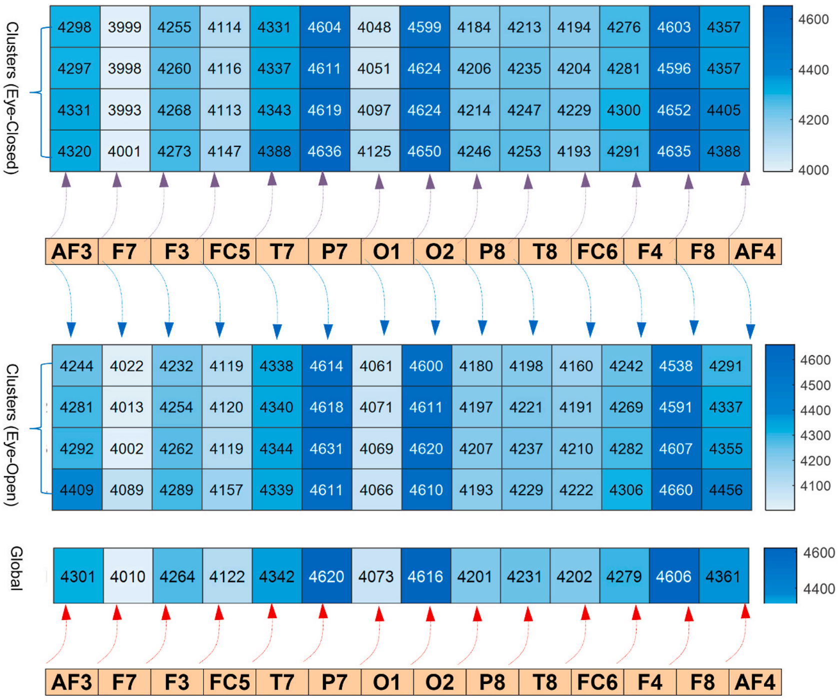

Fig. 4. The weights in LVQ nodes.

PC2, and PC3). Cluster centroids are presented in Table 5.

After data clustering, the ensemble models were developed by bagged trees for each cluster of LVQ. The bagged tree was performed on EEG signals for binary classification (eye-open and eye-closed). Using a 5-fold cross-validation, the models were verified and evaluated on the test data. We did not perform any feature selection, accordingly, in each cluster, we used all features (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 and AF4) of EEG signals in bagged trees for binary classification.

In this study, we used ROC (Receiver Operating Characteristics)-AUC (Area under the Curve) as a performance metric for classification methods that assessed class separability. For displaying the two types of errors simultaneously for all potential thresholds, the ROC curve was found to be a suitable diagram. The false positive rate (FPR) and true positive rate (TPR) were used in the ROC curve. The number of observations that are correctly classified in each class is indicated as true negatives or TN and true positives or TP. False positives or FP and false negatives or FN, on the other hand, represented observations that are misclassified. The AUC is a measure of a classifier's overall performance over all feasible thresholds. The model's ability to accurately predict classes of Eye-Closed and Eye-Open increases with increasing AUC, indicating that AUC falls within the range [0.5, 1], where AUC = 1 denotes perfect class separation and 0.5 denotes a classifier that performs no better than chance. It was also possible to determine the models' accuracy using the confusion matrix depicted in Fig. 6. We used accuracy in this study as a well-known performance indicator that calculated the model's overall classification performance as shown in Fig. 7 for the first segment of LVQ.

We also compare the results of the proposed bagged tree and LVQ with other classifiers, CART (Classification And Regression Tree) [66], bagged tree, LDA (Linear Discriminant Analysis) [67], Random Trees [68], Naïve Bayes [69] and Multilayer Perceptron [70]. The results of comparisons are presented in Table 6 for minimum, maximum, and mean accuracies.

The results reveal that the method which combines LVQ with a bagged tree (AUC = 0.9628; Accuracy = 0.9431) provides the best results in terms of AUC and accuracy in the test and train set. It is found that when LVQ is combined with the bagged tree, the results are

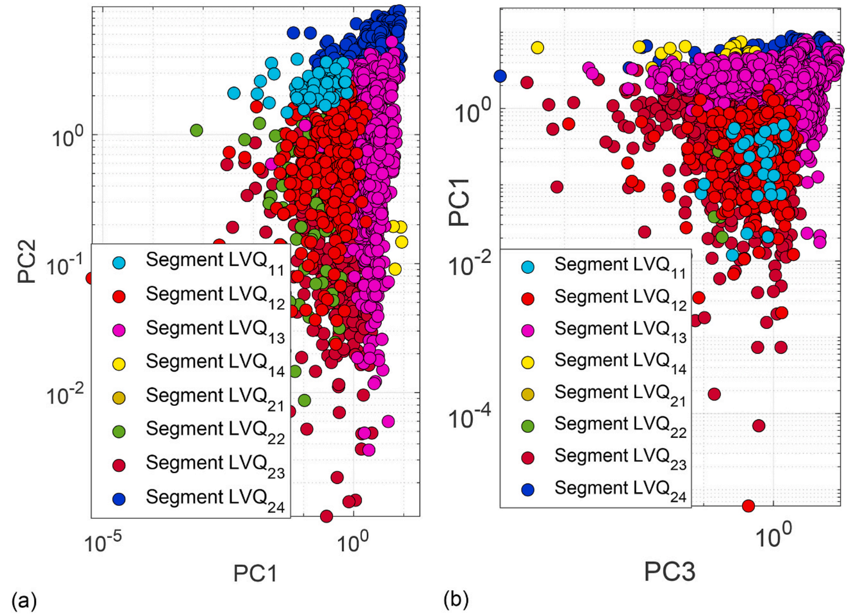

Fig. 5. Visualization of 8 clusters on (a) PC1-PC2 and (b) PC3-PC1.

Table 5
Cluster centroids.

Attribute	Cluster n1	Cluster n2	Cluster n3	Cluster n4	Cluster n5	Cluster n6	Cluster n7	Cluster n8
AF3	4291.893662	4293.461639	4329.504462	4353.221116	4241.191887	4284.156608	4299.611888	4415.060661
F7	3997.340619	3995.309873	4001.448576	4054.478789	4013.459247	4007.650741	4007.555985	4079.334841
F3	4256.021207	4260.733769	4273.032329	4322.626944	4234.078356	4255.021220	4265.802234	4294.333263
FC5	4114.175121	4114.025418	4118.180823	4175.210324	4116.286697	4118.183803	4122.791195	4155.497569
T7	4330.232405	4333.653021	4345.003914	4397.098012	4337.331089	4339.235980	4344.451099	4339.704990
P7	4606.189704	4612.409360	4622.634513	4678.426152	4614.453736	4617.768736	4626.066184	4609.346182
O1	4053.757717	4058.720797	4091.822022	4125.538864	4062.942708	4072.854426	4078.208139	4066.439378
O2	4600.219991	4619.660006	4626.070885	4669.181163	4600.573535	4610.022766	4620.813079	4601.678932
P8	4185.286600	4204.171351	4214.777764	4249.724782	4180.068302	4195.501709	4207.469540	4191.649751
T8	4218.055842	4233.330421	4248.880938	4280.194562	4199.413494	4222.012172	4237.873057	4231.173780
FC6	4189.974865	4200.576312	4229.030067	4251.166484	4158.501006	4189.856699	4207.860045	4226.092918
F4	4269.501285	4277.136310	4298.915453	4325.825390	4244.001129	4268.370172	4282.185059	4308.344747
F8	4596.870373	4600.532459	4645.177594	4657.171332	4536.340393	4590.661410	4609.382098	4660.494015
AF4	4351.722554	4356.681279	4399.484851	4409.851310	4287.285769	4342.120770	4361.486080	4463.490712

better than the method which solely uses the bagged tree (AUC = 0.8783; Accuracy = 0.8618). In addition, the comparison results show that overall the bagged tree method works better in relation to the CART (AUC = 0.8352; Accuracy = 0.8200), LDA (AUC = 0.8091; Accuracy = 0.7931), Random Trees (AUC = 0.8574; Accuracy = 0.8311), Naïve Bayes (AUC = 0.8702; Accuracy = 0.8331) and Multilayer Perceptron (AUC = 0.7641; Accuracy = 0.7718) on the EEG signals for eye state classification. Overall, our findings reveal that clustering of EEG signals with the aid of ensemble learning can improve the accuracy of the classification. This indicates that the use of ensemble learning combined with clustering techniques may be an efficient way in handling large EEG datasets and classification tasks.

In this research, we conducted a two-step statistical test to compare the developed machine learning method with other methods, Bagged Tree, CART, LDA, Random Trees, Naïve Bayes, and Multilayer Perceptron. We utilized the Friedman test [71] and post hoc tests to evaluate the methods (prediction models) over the dataset. The Holm method [72] was used for pairwise comparisons of the control model, which was set as in the proposed method. We found that LVQ + Bagged Tree had the lowest rank and a p-value of 1.45×10^{-11}, which was lower than the threshold of 0.05 (0.05 in this study) (see Table 7). This indicated a significant difference between LVQ + Bagged Tree and Bagged Tree, CART, LDA, Random Trees, Naïve Bayes, and Multilayer Perceptron. We then proceeded with the post hoc test using the Holm method for the estimation of the performance differences between LVQ + Bagged Tree and other methods. The outcomes of the post hoc test showed that the LVQ + Bagged Tree had significant differences in performance when compared to other methods (p-value < 0.05) (see Table 8).

We also provide the time complexity of the methods for prediction speed (Observation/Second). The result is provided in Fig. 8. The result shows that LVQ + Bagged Tree provides the best result for prediction speed (58942 Obs/Sec) in relation to Bagged Tree

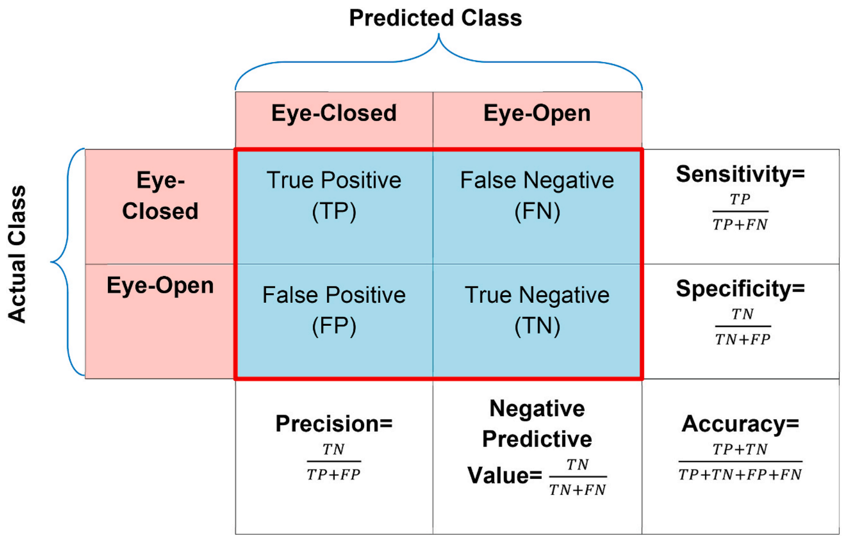

Fig. 6. Confusion matrix with classification metrics.

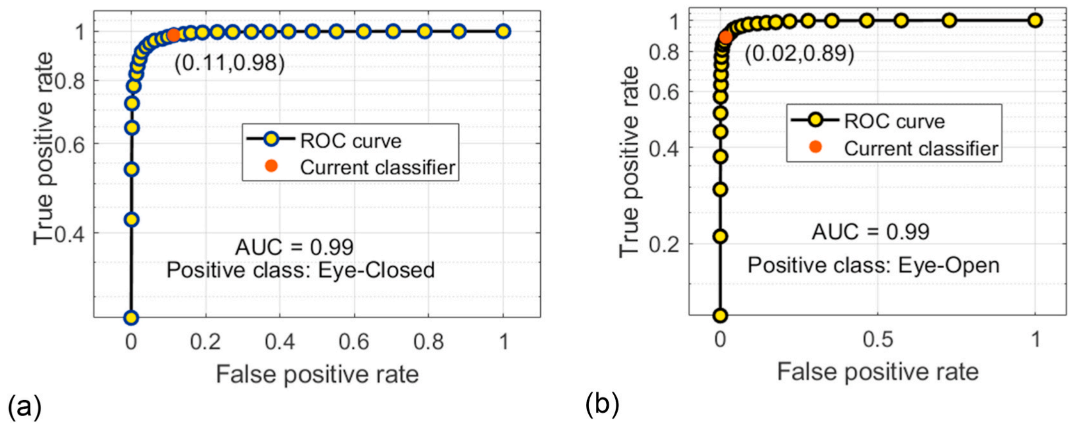

Fig. 7. ROC-AUC results in first segment for (a) Eye-Closed and (b) Eye-Open.

(28453 Obs/Sec), CART (27784 Obs/Sec), LDA (26435 Obs/Sec), Random Trees (27921), Naïve Bayes (27217) and Multilayer Perceptron (24163).

5. Conclusion

The analysis of EEG signals using machine learning has proven to be an effective method for determining eye states. New methods based on machine learning techniques have aided the identification systems in their accuracy improvements and real-time decision-making. Although the previous methods have been effective in EEG data analysis, the use of solely supervised learning techniques may not be an efficient way in the analysis of large datasets. In fact, handling large datasets can be a critical issue in the analysis of large datasets with many features in terms of time complexity and accuracy of classification. Accordingly, this study attempted to develop a new method using supervise and un-supervised learning techniques. We relied on ensemble and clustering machine learning approaches. A bagged tree along with LVQ was used in this study to increase the accuracy of EEG classification meanwhile improve the time complexity. The method was evaluated on a real-world EEG dataset which included 14976 instances after the removal of outlier instances. Using LVQ, 8 clusters were generated from the data. A bagged tree was applied on 8 clusters and compared with other classifiers. Our experiments revealed that LVQ combined with the bagged tree provides the best results (AUC = 0.9628; Accuracy = 0.9431) compared with the other classifiers, which demonstrates the effectiveness of incorporating ensemble learning and clustering approaches in the analysis of EEG signals. We also provided the time complexity of the methods for prediction speed (Observation/Second). The result showed that LVQ + Bagged Tree provides the best result for prediction speed (58942 Obs/Sec) in relation to Bagged Tree (28453 Obs/Sec), CART (27784 Obs/Sec), LDA (26435 Obs/Sec), Random Trees (27921), Naïve Bayes (27217) and Multilayer Perceptron (24163). Although the proposed method provides better results compared with the other classifiers, this method

Table 6
The comparisons of the methods.

Performance index		LVQ + Bagged Tree		Bagged Tree		CART		LDA		Random Trees		Naïve Bayes		Multilayer Perceptron	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Accuracy	Max	0.9434	0.9534	0.8696	0.8895	0.8212	0.8423	0.8055	0.8098	0.8331	0.8499	0.8427	0.8681	0.7987	0.8143
	Min	0.9427	0.9521	0.8541	0.8658	0.8187	0.8370	0.7807	0.7933	0.8291	0.8415	0.8236	0.8483	0.7591	0.7811
	Mean	0.9431	0.9528	0.8618	0.8777	0.8200	0.8397	0.7931	0.8015	0.8311	0.8457	0.8331	0.8582	0.7789	0.7977
AUC	Max	0.9644	0.9733	0.8845	0.9057	0.8372	0.8600	0.8218	0.8288	0.8666	0.8753	0.8744	0.8831	0.7718	0.7918
	Min	0.9611	0.9714	0.8721	0.8821	0.8331	0.8523	0.7964	0.8092	0.8483	0.8681	0.8661	0.873	0.7565	0.7743
	Mean	0.9628	0.9724	0.8783	0.8939	0.8352	0.8561	0.8091	0.8190	0.8574	0.8717	0.8702	0.8780	0.7641	0.7830

Table 7
Friedman test results.

Method	Friedman Rank	p-Value
LVQ + Bagged Tree	4.06	1.45×10^{-11}
Bagged Tree	8.21	
CART	7.78	
LDA	10.25	
Random Trees	8.03	
Naïve Bayes	9.57	
Multilayer Perceptron	11.43	

Table 8
The results of post hoc test.

Method Comparisons	Post Hoc p-Value	Result
LVQ + Bagged Tree vs. Bagged Tree	0.000843	Significant
LVQ + Bagged Tree vs. CART	0.000658	Significant
LVQ + Bagged Tree vs. LDA	0.001092	Significant
LVQ + Bagged Tree vs. Random Trees	0.001096	Significant
LVQ + Bagged Tree vs. Naïve Bayes	0.000884	Significant
LVQ + Bagged Tree vs. Multilayer Perceptron	0.001269	Significant

Fig. 8. The comparisons of the methods for prediction speed.

can be further developed using optimization and incremental learning techniques. In addition, other versions of LVQ with ensemble learning techniques could be applied and evaluated on the EEG datasets for their time complexity and accuracy.

Author contribution statement

Mehrbakhsh Nilashi: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Rabab Ali Abumalloh: Conceived and designed the experiments; contributed reagents, materials, analysis tools or data; Analyzed and interpreted the data; Wrote the paper.

Hossein Ahmadi, Sarminah Samad, Abdullah Alghamdi, Mesfer Alrizq, Sultan Alyami, Fatima Khan Nayer: Conceived and designed the experiments; contributed reagents, materials, analysis tools or data; analyzed and interpreted the data; Wrote the paper.

Data availability statement

Data associated with this study has been deposited at https://archive.ics.uci.edu/ml/datasets/EEG+Eye+State#:~:text=The%20eye%20state%20was%20detected,the%20top%20of%20the%20data; https://archive.ics.uci.edu/ml/machine-learning-databases/00264/.

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.

Acknowledgements

This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R4), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work, under the Research Groups Funding program grant code NU/RG/SERC/12/12.

References

[1] A. Vuckovic, V.J.F. Gallardo, M. Jarjees, M. Fraser, M. Purcell, Prediction of central neuropathic pain in spinal cord injury based on EEG classifier, Clin. Neurophysiol. 129 (8) (2018) 1605–1617.
[2] T.N. Alotaiby, S.A. Alshebeili, T. Alshawi, I. Ahmad, A. El-Samie, E. Fathi, EEG seizure detection and prediction algorithms: a survey, EURASIP J. Appl. Signal Process. 1 (2014) 1–21.
[3] N. Ahmadi, M. Nilashi, B. Minaei-Bidgoli, M. Farooque, S. Samad, N.O. Aljehane, W.A. Zogaan, H. Ahmadi, Eye state identification utilizing EEG signals: a combined method using self-organizing map and deep belief network, Sci. Program. 2022 (2022).
[4] Y. Sun, F.P.-W. Lo, B. Lo, EEG-based user identification system using 1D-convolutional long short-term memory neural networks, Expert Syst. Appl. 125 (2019) 259–267.
[5] L.A. Moctezuma, M. Molinas, Multi-objective optimization for EEG channel selection and accurate intruder detection in an EEG-based subject identification system, Sci. Rep. 10 (1) (2020) 1–12.
[6] L.C.D. Nkengfack, D. Tchiotsop, R. Atangana, V. Louis-Door, D. Wolf, Classification of EEG signals for epileptic seizures detection and eye states identification using Jacobi polynomial transforms-based measures of complexity and least-square support vector machine, Inform. Med. Unlocked 23 (2021), 100536.
[7] F. Laport, P.M. Castro, A. Dapena, F.J. Vazquez-Araujo, O. Fresnedo, Eye state identification based on discrete wavelet transforms, Appl. Sci. 11 (11) (2021) 5051.
[8] T. Wang, S.-U. Guan, K.L. Man, T. Ting, EEG eye state identification using incremental attribute learning with time-series classification, Math. Probl Eng. 2014 (2014).
[9] S. Ketu, P.K. Mishra, Hybrid classification model for eye state detection using electroencephalogram signals, Cogn. Neurodyn. 16 (1) (2022) 73–90.
[10] M. Sahu, N. Nagwani, S. Verma, S. Shirke, An incremental feature reordering (IFR) algorithm to classify eye state identification using EEG, in: Information Systems Design and Intelligent Applications, Springer, 2015, pp. 803–811.
[11] Z. Mardi, S.N.M. Ashtiani, M. Mikaili, EEG-based drowsiness detection for safe driving using chaotic features and statistical tests, J. Med. Signals Sens. 1 (2) (2011) 130.
[12] P. Estévez, C. Held, C. Holzmann, C. Perez, J. Pérez, J. Heiss, M. Garrido, P. Peirano, Polysomnographic pattern recognition for automated classification of sleep-waking states in infants, Med. Biol. Eng. Comput. 40 (1) (2002) 105–113.
[13] R.J. Barry, A.R. Clarke, S.J. Johnstone, C.A. Magee, J.A. Rushby, EEG differences between eyes-closed and eyes-open resting conditions, Clin. Neurophysiol. 118 (12) (2007) 2765–2773.
[14] Y. Di, X. An, F. He, S. Liu, Y. Ke, D. Ming, Robustness analysis of identification using resting-state EEG signals, IEEE Access 7 (2019) 42113–42122.
[15] D. González-Ortega, F. Díaz-Pernas, M. Antón-Rodríguez, M. Martínez-Zarzuela, J. Díez-Higuera, Real-time vision-based eye state detection for driver alertness monitoring, Pattern Anal. Appl. 16 (3) (2013) 285–306.
[16] A. Saghafi, C.P. Tsokos, M. Goudarzi, H. Farhidzadeh, Random eye state change detection in real-time using EEG signals, Expert Syst. Appl. 72 (2017) 42–48.
[17] K.K. Dutta, P. Manohar, S. Rajagopalan, F. Naaz, M. Lakshminarayanan, Eye state detection from electro-encephalography signals using machine learning techniques, in: 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), IEEE, 2022, pp. 1–7.
[18] Y. Dong, Y. Zhang, J. Yue, Z. Hu, Comparison of random forest, random ferns and support vector machine for eye state classification, Multimed. Tool. Appl. 75 (19) (2016) 11763–11783.
[19] A. Tahmassebi, J. Martin, A. Meyer-Baese, A.H. Gandomi, An interpretable deep learning framework for health monitoring systems: a case study of eye state detection using eeg signals, in: 2020 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2020, pp. 211–218.
[20] Z.-T. Liu, C.-S. Jiang, S.-H. Li, M. Wu, W.-H. Cao, M. Hao, Eye state detection based on weight binarization convolution neural network and transfer learning, Appl. Soft Comput. 109 (2021), 107565.
[21] L. Zhao, Z. Wang, G. Zhang, Y. Qi, X. Wang, Eye state recognition based on deep integrated neural network and transfer learning, Multimed. Tool. Appl. 77 (15) (2018) 19415–19438.
[22] C.-H. Han, G.-Y. Choi, H.-J. Hwang, Deep convolutional neural network based eye states classification using ear-EEG, Expert Syst. Appl. 192 (2022), 116443.
[23] M.M. Hassan, M.R. Hassan, S. Huda, M.Z. Uddin, A. Gumaei, A. Alsanad, A predictive intelligence approach to classify brain–computer interface based eye state for smart living, Appl. Soft Comput. 108 (2021), 107453.
[24] M. Varli, H. Yilmaz, Multiple classification of EEG signals and epileptic seizure diagnosis with combined deep learning, J. Comput. Sci. (2023), 101943.
[25] H.U. Amin, M.Z. Yusoff, R.F. Ahmad, A novel approach based on wavelet analysis and arithmetic coding for automated detection and diagnosis of epileptic seizure in EEG signals using machine learning techniques, Biomed. Signal Process Control 56 (2020), 101707.
[26] X. Liu, R. Shi, Q. Hui, S. Xu, S. Wang, R. Na, Y. Sun, W. Ding, D. Zheng, X. Chen, TCACNet: temporal and channel attention convolutional network for motor imagery classification of EEG-based BCI, Inf. Process. Manag. 59 (5) (2022), 103001.
[27] T. Keles, A.M. Yildiz, P.D. Barua, S. Dogan, M. Baygin, T. Tuncer, C.F. Demir, E.J. Ciaccio, U.R. Acharya, A new one-dimensional testosterone pattern-based EEG sentence classification method, Eng. Appl. Artif. Intell. 119 (2023), 105722.
[28] W. Al-Salman, Y. Li, A.Y. Oudah, S. Almaged, Sleep stage classification in EEG signals using the clustering approach based probability distribution features coupled with classification algorithms, Neuroscience Research 188 (2023) 51–67.
[29] Z. Mousavi, T.Y. Rezaii, S. Sheykhivand, A. Farzamnia, S. Razavi, Deep convolutional neural network for classification of sleep stages from single-channel EEG signals, J. Neurosci. Methods 324 (2019), 108312.
[30] N. Michielli, U.R. Acharya, F. Molinari, Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals, Comput. Biol. Med. 106 (2019) 71–81.
[31] S.U. Amin, M. Alsulaiman, G. Muhammad, M.A. Mekhtiche, M.S. Hossain, Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion, Future Generat. Comput. Syst. 101 (2019) 542–554.
[32] X. Li, H. Fan, H. Wang, L. Wang, Common spatial patterns combined with phase synchronization information for classification of EEG signals, Biomed. Signal Process Control 52 (2019) 248–256.
[33] R. San-Segundo, M. Gil-Martín, L.F. D'Haro-Enríquez, J.M. Pardo, Classification of epileptic EEG recordings using signal transforms and convolutional neural networks, Comput. Biol. Med. 109 (2019) 148–158.
[34] A. Asif, M. Majid, S.M. Anwar, Human stress classification using EEG signals in response to music tracks, Comput. Biol. Med. 107 (2019) 182–196.
[35] E. Perez-Valero, M.A. Lopez-Gordo, M.A. Vaquero-Blasco, EEG-based multi-level stress classification with and without smoothing filter, Biomed. Signal Process Control 69 (2021), 102881.
[36] L. Zunino, F. Olivares, A.F. Bariviera, O.A. Rosso, A simple and fast representation space for classifying complex time series, Phys. Lett. 381 (11) (2017) 1021–1028.
[37] O. Rosso, M. Martin, A. Plastino, Brain electrical activity analysis using wavelet-based informational tools, Phys. Stat. Mech. Appl. 313 (3–4) (2002) 587–608.
[38] O. Sagi, L. Rokach, Ensemble learning: a survey, Wiley Interdisc. Rev.: Data Min. Knowl. Discov. 8 (4) (2018) e1249.

[39] J.-S. Chou, D.-S. Tran, Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders, Energy 165 (2018) 709–726.
[40] C. Ying, M. Qi-Guang, L. Jia-Chen, G. Lin, Advance and prospects of AdaBoost algorithm, Acta Autom. Sin. 39 (6) (2013) 745–758.
[41] P. Bühlmann, Bagging, Boosting and Ensemble Methods, Handbook of Computational Statistics, Springer, 2012, pp. 985–1022.
[42] A.V. Konstantinov, L.V. Utkin, Interpretable machine learning with an ensemble of gradient boosting machines, Knowl. Base Syst. 222 (2021), 106993.
[43] A. Chatzimparmpas, R.M. Martins, K. Kucher, A. Kerren, Empirical study: visual analytics for comparing stacking to blending ensemble learning, in: 2021 23rd International Conference on Control Systems and Computer Science (CSCS), IEEE, 2021, pp. 1–8.
[44] M. Belgiu, L. Drăguţ, Random forest in remote sensing: a review of applications and future directions, ISPRS J. Photogrammetry Remote Sens. 114 (2016) 24–31.
[45] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140.
[46] T.-H. Lee, A. Ullah, R. Wang, Bootstrap Aggregating and Random Forest, Macroeconomic Forecasting in the Era of Big Data, Springer, 2020, pp. 389–429.
[47] M. Nilashi, R.A. Abumalloh, B. Minaei-Bidgoli, S. Samad, M. Yousoof Ismail, A. Alhargan, W. Abdu Zogaan, Predicting Parkinson's disease progression: evaluation of ensemble methods in machine learning, J. Healthcare Eng. 2022 (2022).
[48] M. Nilashi, R.A. Abumalloh, M. Alrizq, A. Almulihi, O. Alghamdi, M. Farooque, S. Samad, S. Mohd, H. Ahmadi, A hybrid method to solve data sparsity in travel recommendation agents using fuzzy logic approach, Math. Probl Eng. (2022) 2022.
[49] M. Nilashi, R.A. Abumalloh, S. Samad, M. Alrizq, S. Alyami, H. Abosaq, A. Alghamdi, N.A.M. Akib, Factors impacting customer purchase intention of smart home security systems: social data analysis using machine learning techniques, Technol. Soc. 71 (2022), 102118.
[50] M. Nilashi, B. Minaei-Bidgoli, A. Alghamdi, M. Alrizq, O. Alghamdi, F.K. Nayer, N.O. Aljehane, A. Khosravi, S. Mohd, Knowledge discovery for course choice decision in Massive Open Online Courses using machine learning approaches, Expert Syst. Appl. 199 (2022), 117092.
[51] M. Nilashi, E. Yadegaridehkordi, O. Ibrahim, S. Samad, A. Ahani, L. Sanzogni, Analysis of travellers' online reviews in social networking sites using fuzzy logic approach, Int. J. Fuzzy Syst. 21 (2019) 1367–1378.
[52] M. Nilashi, H. Ahmadi, L. Shahmoradi, O. Ibrahim, E. Akbari, A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique, J. Infect. Public Health 12 (1) (2019) 13–20.
[53] M. Nilashi, A. Ahani, M.D. Esfahani, E. Yadegaridehkordi, S. Samad, O. Ibrahim, N.M. Sharef, E. Akbari, Preference learning for eco-friendly hotels recommendation: a multi-criteria collaborative filtering approach, J. Clean. Prod. 215 (2019) 767–783.
[54] T. Kohonen, The self-organizing map, Proc. IEEE 78 (9) (1990) 1464–1480.
[55] T. Kohonen, Learning Vector Quantization, Self-Organizing Maps, Springer, 1995, pp. 175–189.
[56] O. Rösler, D. Suendermann, A first step towards eye state prediction using eeg, Proc. AIHLS 1 (2013) 1–4.
[57] M. Nilashi, P.F. Rupani, M.M. Rupani, H. Kamyab, W. Shao, H. Ahmadi, T.A. Rashid, N. Aljojo, Measuring sustainability through ecological sustainability and human sustainability: a machine learning approach, J. Clean. Prod. 240 (2019), 118162.
[58] M. Nilashi, H. Ahmadi, A. Sheikhtaheri, R. Naemi, R. Alotaibi, A.A. Alarood, A. Munshi, T.A. Rashid, J. Zhao, Remote tracking of Parkinson's disease progression using ensembles of deep belief network and self-organizing map, Expert Syst. Appl. 159 (2020), 113562.
[59] A. Ahani, M. Nilashi, W.A. Zogaan, S. Samad, N.O. Aljehane, A. Alhargan, S. Mohd, H. Ahmadi, L. Sanzogni, Evaluating medical travelers' satisfaction through online review analysis, J. Hospit. Tourism Manag. 48 (2021) 519–537.
[60] M.Y. Hu, G. Zhang, C.X. Jiang, B.E. Patuwo, A cross-validation analysis of neural network out-of-sample performance in exchange rate forecasting, Decis. Sci. J. 30 (1) (1999) 197–216.
[61] S. Borra, A. Di Ciaccio, Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods, Comput. Stat. Data Anal. 54 (12) (2010) 2976–2989.
[62] J. Pohjankukka, T. Pahikkala, P. Nevalainen, J. Heikkonen, Estimating the prediction performance of spatial models via spatial k-fold cross validation, Int. J. Geogr. Inf. Sci. 31 (10) (2017) 2001–2019.
[63] M. Nilashi, S. Samad, A. Ahani, H. Ahmadi, E. Alsolami, M. Mahmoud, H.D. Majeed, A.A. Alarood, Travellers decision making through preferences learning: a case on Malaysian spa hotels in TripAdvisor, Comput. Ind. Eng. 158 (2021), 107348.
[64] M. Ghane, M.C. Ang, M. Nilashi, S. Sorooshian, Enhanced decision tree induction using evolutionary techniques for Parkinson's disease classification, Biocybern. Biomed. Eng. 42 (3) (2022) 902–920.
[65] M. Nilashi, S. Asadi, R.A. Abumalloh, S. Samad, F. Ghabban, E. Supriyanto, R. Osman, Sustainability performance assessment using self-organizing maps (SOM) and classification and ensembles of regression trees (CART), Sustainability 13 (7) (2021) 3870.
[66] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Routledge, 2017.
[67] J.H. Friedman, Regularized discriminant analysis, J. Am. Stat. Assoc. 84 (405) (1989) 165–175.
[68] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[69] G.H. John, P. Langley, Estimating Continuous Distributions in Bayesian Classifiers, arXiv preprint arXiv:1302.4964, 2013.
[70] B. Aazhang, B.-P. Paris, G.C. Orsak, Neural networks for multiuser detection in code-division multiple-access communications, IEEE Trans. Commun. 40 (7) (1992) 1212–1222.
[71] M. Friedman, The use of ranks to avoid the assumption of normality implicit in the analysis of variance, J. Am. Stat. Assoc. 32 (200) (1937) 675–701.
[72] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Stat. (1979) 65–70.