I.J. Information Technology and Computer Science, **2022, 6, 33-38**
Published Online on December 8, 2022 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2022.06.04

Comparing the Performances of Ensemble-classifiers to Detect Eye State

Kemal Akyol*
Faculty of Engineering and Architecture, Computer Engineering, Kastamonu University, Turkey
E-mail: kakyol@kastamonu.edu.tr
ORCID iD: https://orcid.org/0000-0002-2272-5243
*Corresponding Author

Abdulkadir Karacı
Faculty of Engineering, Software Engineering, Samsun University, Turkey
E-mail: abdulkadir.karaci@samsun.edu.tr
ORCID iD: https://orcid.org/0000-0002-2430-1372

Received: 14 July 2022; Revised: 06 September 2022; Accepted: 14 October 2022; Published: 08 December 2022

Abstract: Brain signals required for the brain-computer interface are obtained through the electroencephalography (EEG) method. EEG data is used in the analysis of many problems such as epileptic seizure detection, bipolar mood disorder, attention deficit, and detection of the sleep state of the vehicle driver. It is very important to determine whether the eye is open or closed, which is a substantial organ for the determination of the cognitive state of the person. The aim of this paper is to present a stable and successful model for detecting the eye states that are opened or closed. In this context, the performances of several ensemble classifiers were examined on the Emotiv EEG Neuroheadset dataset, which has 14 features excluding the target variable, 14980 records that have 8225 eye states opened and 6755 eye states closed. In the experiments, firstly the min-max normalization process was applied to the dataset, and then the classification performances of these classifiers were evaluated via a 5-fold cross-validation technique. The performance of each model was measured using accuracy, sensitivity, and specificity metrics. The obtained results show that the Random Forest algorithm is an acceptable level with 92.61% value of accuracy, 94.31% value of sensitivity and 91.36% value of specificity for detecting the eye state.

Index Terms: Eye state classification, Electroencephalography, Machine learning, Random Forest.

1. Introduction

Electroencephalography (EEG) data is used to examine brain activities. Machine learning is performed using these data and also brain-computer interface systems are designed [1]. One of the studies conducted with these systems based on EEG data is to determine the cognitive status of people. Detection of vehicle driving sleepiness can be given as an example. Whether the eye is open or closed has a key role in this problem. There are many studies [2–9] about EEG data analysis in the literature. Some studies conducted on the eye state are as follows: Wang et al. proposed an approach that includes incremental feature learning based on neural networks to describe the eye state. In their study, the authors first extracted features from the raw EEG data and then performed classification experiments on these features [10]. Bharati et al. conducted experiments on the EEG dataset with Naive Bayes polynomial, Logistics, Partial Decision Tree, K-nearest neighbor, Decision Table, and Support Vector Machines classifiers in Waikato Environment for Knowledge Analysis environment [11]. Kim et al. proposed a fuzzy rule-based approach that includes a genetic algorithm-optimized neural network [12]. Mridu et al. showed that Random Forest and sample-based classifiers such as IB1 and IBK offer better performance compared to other classifiers to predict eye status using EEG signals [13].

The main aim of this study is to compare the effectiveness of Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Gradient Boosting (GB) classifiers using a 5-fold cross-validation technique for detecting EEG signals. The performances of the classifiers were evaluated on the Emotiv EEG Neuroheadset dataset that consists of 14980 records, 8225 opened eye states and 6755 closed eye states.

The remainder of this study is organized as follows. Section 2 provides information about the dataset used in this study, the classifier algorithms, as well as the metrics used to evaluate the performances of the algorithms. Section 3 explains the model training and testing processes. Section 4 presents experiment and discusses results in detail. Finally, Section 5 introduces the conclusions.

2. Material and Method

2.1. Dataset

The dataset used in this study was composed by Roesler [14] using the Emotiv EEG Neuroheadset. This dataset can be downloaded from the UCI machine learning repository. There are 14 attributes that characterize the two-class eye state target variable. '1' indicates eye closed and '0' eye open state. The list of attributes in the dataset is summarized in Table 1.

Table 1. Attributes and their value ranges in the dataset

No	Attribute	Value
1	AF3	Numerical
2	F7	Numerical
3	F3	Numerical
4	FC5	Numerical
5	T7	Numerical
6	P7	Numerical
7	O1	Numerical
8	O2	Numerical
9	P8	Numerical
10	T8	Numerical
11	FC6	Numerical
12	F4	Numerical
13	F8	Numerical
14	AF4	Numerical
15	eyeDetection (Target class)	{0,1}

2.2. Ensemble Classifiers

Information about GB, XGBoost, and RF ensemble classifiers that are used to determine the eye-open and closed states is presented below, respectively.

The GB ensemble classifier is based on the forward distribution algorithm. Forward distribution is the idea of learning only one basis function and coefficient at a time and gradually approaching the optimal solution [15]. This algorithm is an ensemble learning algorithm based on iteratively generating predictive models [16]. The GB aims to minimize the loss function problem under the given training data and loss function condition [15].

XGBoost is a collection of decision trees based on gradient boosting. The peculiarity of XGBoost is the automatic use of CPU multi-threading for parallel processing. The XGBoost performs a second-order Taylor expansion of the loss function and adds a regular term to the loss function to find the optimal solution to compensate for the loss function's decay and complexity of the model and to avoid overfitting [17,18]. XGBoost is a scalable machine learning system for tree reinforcement and is widely used by data scientists to achieve cutting-edge results for many machines learning challenges [19].

RF is an algorithm based on a collection of decision trees developed by Breiman [20] to improve overall classification accuracy. The author developed this algorithm in order to eliminate the overfitting, that is, the memorization problem, encountered in the decision tree algorithm. This algorithm, which has very few classification errors, is very popular compared to other traditional classification algorithms [21,22].

2.3. Performance Evaluation

A 5-fold cross-validation technique is used for evaluating the performances of the classifier algorithms. As seen in Figure 1, different 20% of the dataset is reserved for the testing of the model, while the remaining parts are used in the training of the model for each fold. Thus, the performances of the classifier algorithms are discussed on 5 different test sets. Due to the nature of the k-fold cross-validation technique, the training and testing sets for each fold are different from each other.

Sensitivity, specificity and accuracy metrics given in between Equation 1 and Equation 3, respectively are quite often used to compare the performance of classifier models. Here, TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative. In other words, TN and TP denote the number of eyes open and eye closed samples correct classified, respectively. FP and FN denote the number of eye open and eye closed samples misclassified, respectively. Sensitivity (Sen) indicates the ratio of the numbers of positives that are correctly classified to the numbers of all positive samples, and specificity (Spe) indicates the ratio of the numbers of negatives that are correctly classified to all negative samples. Lastly, accuracy (Acc) indicates the general classification accuracy which is the ratio of the number of samples classified correctly to all samples.

$$Sen = \frac{TP}{TP + FN} \tag{1}$$

$$Spe = \frac{TN}{TN + FP} \tag{2}$$

$$Acc = \frac{TN + TP}{TN + FP + TP + FN} \tag{3}$$

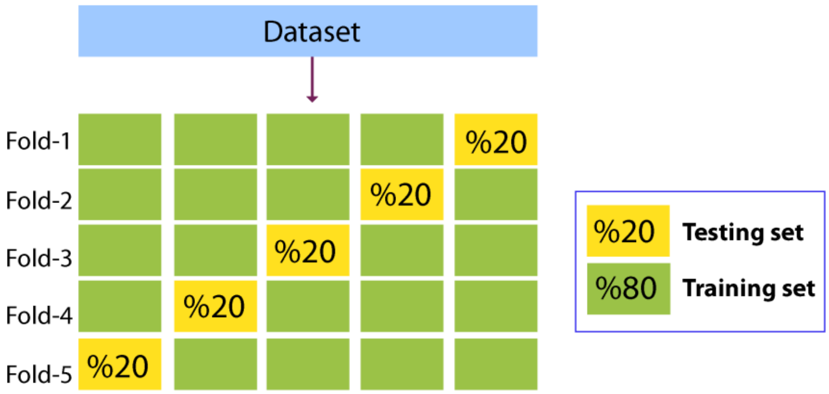

Fig.1. 5-fold cross validation technique

3. Experiments

Figure 2 presents the general framework of this study. Before the model training process, the min-max normalization process was applied to the dataset. RF, XGBoost, and GB ensemble learning classifiers were conducted to detect eye-open or eye-closed states from EEG signals within the framework of a 5-fold cross-validation technique. Thus, the training and testing phases for each classifier were performed 5 times. There are 11984 training samples and 2996 testing samples for each fold. For a fair comparison, the training and testing of classifiers were performed on the same training and testing sets for each fold. Also, the classifiers were trained with their default parameters. While Figure 3 shows the confusion matrices for each fold of the RF classifier that gave the best classification performance in the experimental studies, Figure 4 presents the overlapped confusion matrices that aggregated the confusion matrices obtained from each fold for each classifier. According to these overlapped confusion matrices, the RF algorithm that gave the best result misclassified 744 out of 6755 eyes closed samples and 363 out of 8225 eyes opened samples.

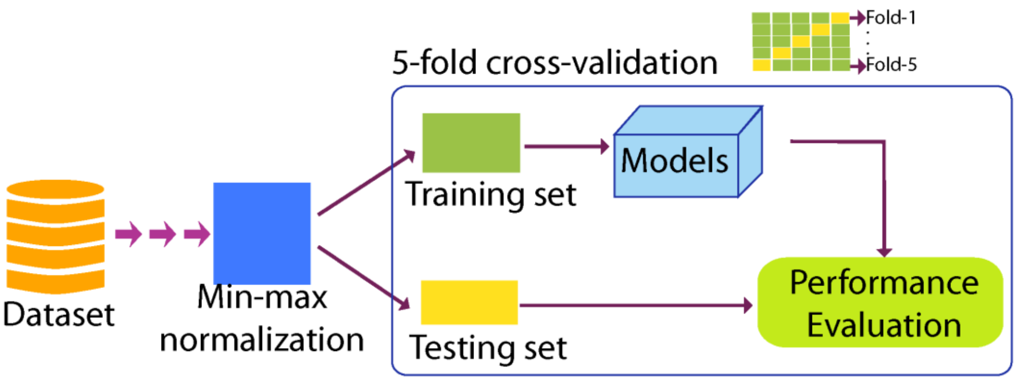

Fig.2. General block diagram of the study

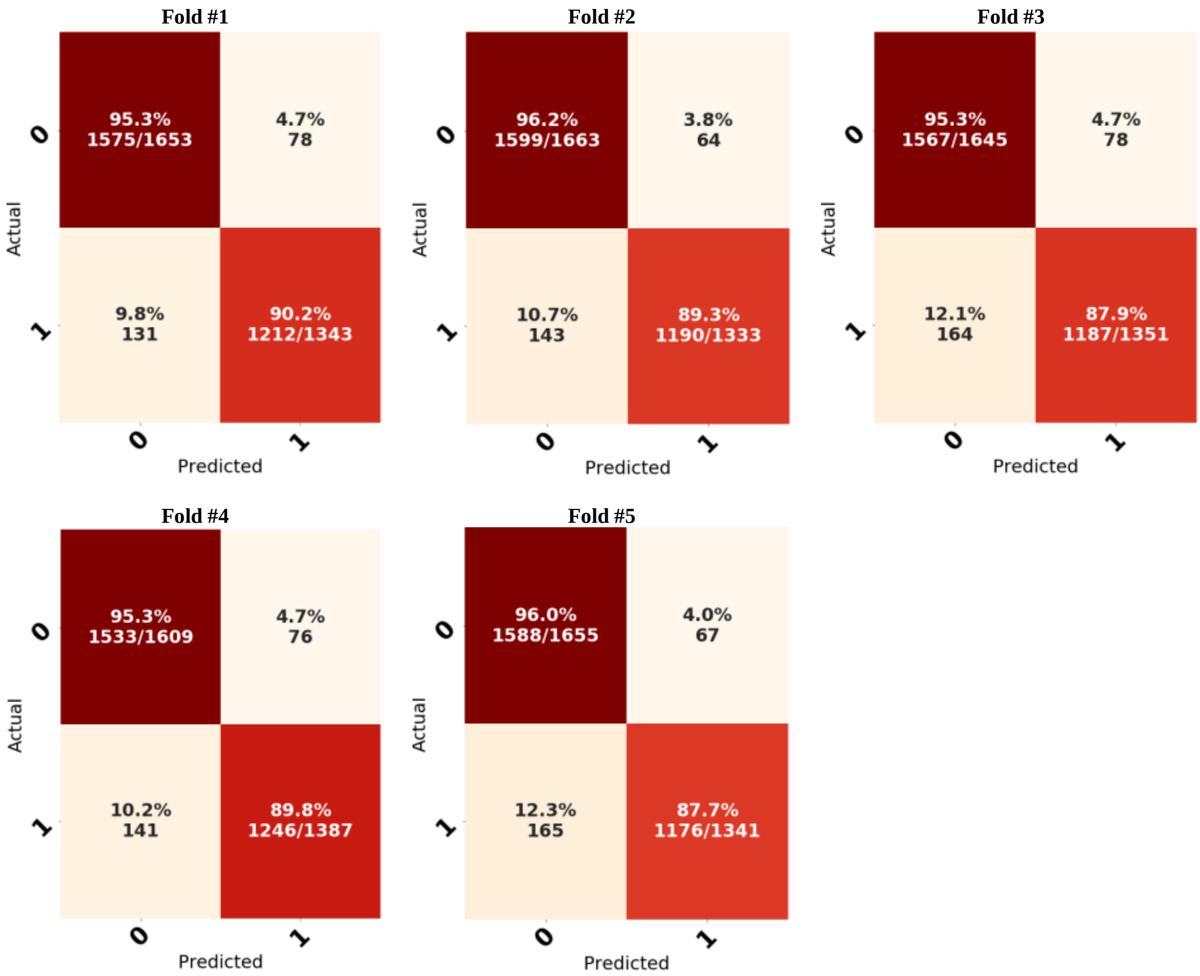

Fig.3. Confusion matrices obtained by Random Forest classifier within the framework of 5-fold cross-validation technique

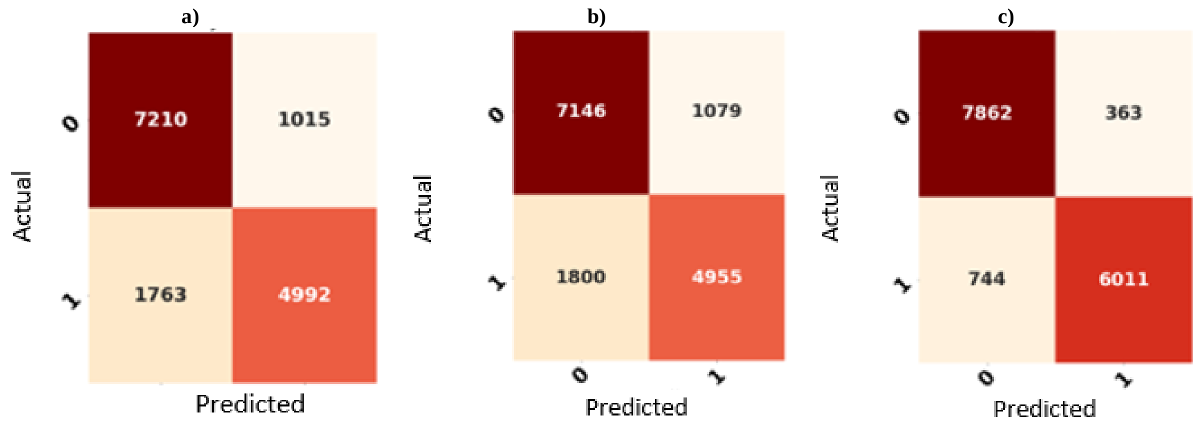

Fig.4. Overlapped confusion matrices of classifier algorithms; a) GB, b) XGBoost, c) RF

Table 2 summarizes the results presented by the three classifiers for each fold. TP, TN, FP and FN values, as well as accuracy, sensitivity and specificity values, are given in this table, respectively. For example, the last row in this table indicates the results of the RF classifier. The accuracy, sensitivity and specificity measures of this classifier for each fold were presented in the 5 rows. In addition, the average of the results presented by each classifier is also included in this table. According to experiments, RF presented considerably high overall classification performance with 92.61% classification accuracy, 94.31% sensitivity and 91.36% specificity values. Accordingly, the RF classifier has 11.16% and 11.83% better overall classification performance than GB and XGBoost classifiers, respectively.

Table 2. Results obtained with 5-fold cross-validation technique

Classifier	k value	TP	FN	TN	FP	Acc (%)	Sen (%)	Spe (%)
GB	1	1022	321	1433	220	81.94	82.29	81.7
	2	982	351	1447	216	81.07	81.97	80.48
	3	993	358	1442	203	81.28	83.03	80.11
	4	1051	336	1425	184	82.64	85.1	80.92
	5	944	397	1463	192	80.34	83.1	78.66
					Average	81.45	83.1	80.37
XGBoost	1	1000	343	1416	237	80.64	80.84	80.5
	2	986	347	1438	225	80.91	81.42	80.56
	3	982	369	1429	216	80.47	81.97	79.48
	4	1045	342	1402	207	81.68	83.47	80.39
	5	942	399	1461	194	80.21	82.92	78.55
					Average	80.78	82.12	79.9
RF	1	1212	131	1575	78	93.02	93.95	92.32
	2	1190	143	1599	64	93.09	94.9	91.79
	3	1187	164	1567	78	91.92	93.83	90.53
	4	1246	141	1533	76	92.76	94.25	91.58
	5	1176	165	1588	67	92.26	94.61	90.59
					Average	**92.61**	**94.31**	**91.36**

Note: **Bold** values indicate the best results

4. Conclusions

EEG is a test that measures brain electrical activities. One of the studies on the analysis of these activities is the determination of whether the eye is open or closed. In this study, eye state classification was performed by using EEG data obtained from brain activities. Within the framework of the 5-fold cross-validation technique that validates the performance of the models, the RF classifier correctly classified 6011 out of 6755 eyes closed samples and 7862 out of 8225 eyes opened. As result, the RF offered a higher classification performance than GB and XGBoost with an overall classification accuracy of 92.61%. In this context, it is thought that an expert decision support system based on the RF classifier that presents the best performance can be used in eye state determinations.

The limitation of this study is that the dataset has a 2-class with a limited number of samples. It is among the targets to work with multi-class EEG signals and also deep learning experiments in the future.

Acknowledgment

The authors wish to thank Roesler [14] for the eye state dataset.

References

[1] A. Vallabhaneni, T. Wang, B. He, "Brain-Computer Interface," *In: He, B. (eds) Neural Engineering. Bioelectric Engineering. Springer, Boston, MA.*, 2005, pp. 85–121.

[2] H. Javaid, R. Manor, E. Kumarnsit, S. Chatpun, "Decision Tree in Working Memory Task Effectively Characterizes EEG Signals in Healthy Aging Adults," *IRBM Innotive Research Biomedical Engineering*, 2021, In Press.

[3] B. Ari, N. Sobahi, Ö.F. Alçin, A. Sengur, U.R. Acharya, "Accurate detection of autism using Douglas-Peucker algorithm, sparse coding based feature mapping and convolutional neural network techniques with EEG signals," *Comput Biol Med*, 2022;, vol. 143, pp. 1-10.

[4] D. Sunaryono, R. Sarno, J. Siswantoro, "Gradient boosting machines fusion for automatic epilepsy detection from EEG signals based on wavelet features," *J King Saud Univ - Comput Inf*, Sci 2021. In Press.

[5] İ. Yıldız, R. Garner, M. Lai, D. Duncan, "Unsupervised seizure identification on EEG," *Comput Methods Programs Biomed*, 2022, vol. 215, pp. 1-9.

[6] W. Al-Salman, Y. Li, P. Wen, F.S. Miften, A.Y. Oudah, H.R.A. Ghayab, "Extracting epileptic features in EEGs using a dual-tree complex wavelet transform coupled with a classification algorithm," *Brain Res*, 2022, vol. 1779, pp. 1-15.

[7] A.N. Pusarla, B.A. Singh, C.S. Tripathi, "Learning DenseNet features from EEG based spectrograms for subject independent emotion recognition," *Biomed Signal Process Control*, 2022, vol. 74, pp. 1-12.

[8] C. Li, B. Wang, S. Zhang, Y. Liu, R. Song, J. Cheng, et al., "Emotion recognition from EEG based on multi-task learning with capsule network and attention mechanism," *Comput Biol Med*, 2022, vol. 143, pp. 1-11.

[9] Arjun, A.S. Rajpoot, M.R. Panicker, "Subject independent emotion recognition using EEG signals employing attention driven neural networks," *Biomed Signal Process Control*, 2022, vol. 75, pp. 1-11.

[10] T. Wang, S.U. Guan, K.L. Man, T.O. Ting, "EEG eye state identification using incremental attribute learning with time-series classification," *Math Probl Eng*, 2014, vol. 2014, pp. 1-10.

[11] S. Bharati, P. Podder, M. Raihan-Al-Masud, "EEG Eye State Prediction and Classification in order to Investigate Human Cognitive State," 2018 Int Conf Adv Electr Electron Eng ICAEEE 2018 2019. https://doi.org/10.1109/ICAEEE.2018.8643015.

[12] Y.M. Kim, C.Y. Lee, C.G. Lim, "Computing Intelligence Approach for an Eye State Classification with EEG Signal in BCI," *Software Engineering and Information Technology*, 2016, pp. 265–270.

[13] M. Sahu, N.K. Nagwani, S. Verma, S. Shirke,"Performance Evaluation of Different Classifier for Eye State Prediction Using EEG Signal," *Int J Knowl Eng*, 2015, vol. 1, pp. 141–145.

[14] O. Roesler, "UCI Machine Learning Repository: EEG Eye State Data Set," 2013. https://archive.ics.uci.edu/ml/datasets/EEG+Eye+State (accessed March 2, 2022).

[15] P. Qin, L. Zhao, Z. Liu,"State of health prediction for lithium-ion battery using a gradient boosting-based data-driven method," *J Energy Storage*, 2022, vol. 47, pp. 1-23.

[16] M. Sipper, J.H. Moore,"AddGBoost: A gradient boosting-style algorithm based on strong learners," *Mach Learn with Appl*, 2022, vol. 7, pp. 1-4.

[17] M. Wang, X. Li, M. Lei, L. Duan, H. Chen, "Human health risk identification of petrochemical sites based on extreme gradient boosting," *Ecotoxicol Environ Saf*, 2022, vol. 233, pp. 1-8.

[18] C. Bentéjac, A. Csörgő, G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artif Intell Rev*, 2021, vol. 54, pp. 1937–1967.

[19] L. Zhou, H. Fujita, H. Ding, R. Ma, "Credit risk modeling on data with two timestamps in peer-to-peer lending by gradient boosting," *Appl Soft Comput*, 2021, vol. 110, pp. 1-12.

[20] L. Breiman, "Random forests," *Mach Learn*, 2001, vol. 45, pp. 5-32.

[21] S. Georganos, T. Grippa, A.N. Gadiaga, C. Linard, M. Lennert, S. Vanhuysse, et al., "Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling," *Geocarto International*, vol. 36, no. 2, pp. 121–136.

[22] N. Farnaaz, M.A. Jabbar, "Random Forest Modeling for Network Intrusion Detection System," *Procedia Comput Sci*, 2016, vol. 89, pp. 213–217.

Authors' Profiles

Kemal Akyol received his B. sc. from Computer Science Department of Gazi University, Ankara/Turkey in 2002. He received his M. sc. degree from Natural and Applied Sciences, Karabuk University, Karabuk/Turkey and PhD degree from the same department. His research interests include machine learning, decision support systems and expert systems. He currently works at the Department of Computer Engineering, Kastamonu University.

Abdulkadir Karacı received the Doctoral degree in electronic and computer education (intelligent tutoring system, machine learning and fuzzy logic) from Gazi University, Turkey, in 2013. He is currently an Associate Professor at the Department of Software Engineering, Samsun University. His current research interests include machine learning, convolution neural network, intelligent tutoring system and programming languages.

How to cite this paper: Kemal Akyol, Abdulkadir Karacı, "Comparing the Performances of Ensemble-classifiers to Detect Eye State", International Journal of Information Technology and Computer Science(IJITCS), Vol.14, No.6, pp.33-38, 2022. DOI:10.5815/ijitcs.2022.06.04