Quantitative EEG Features and Machine Learning Classifiers for Eye-Blink Artifact Detection: A Comparative Study

Maliha Rashida (✉ maliha.2710@gmail.com)  
Chittagong University of Engineering and Technology  
https://orcid.org/0000-0002-4085-206X

Mohammad Ashfak Habib  
Chittagong University of Engineering and Technology

Research Article

Keywords: EEG, Eye blink artifact, Ocular artifact, Machine learning, Supervised learning

Posted Date: May 25th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1595643/v1

License: ☇ ⓒ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Ocular artifact, namely eye-blink artifact, is an unavoidable and one of the most destructive noises in EEG signals. Many solutions were proposed regarding the detection of the eye-blink artifact. Different subsets of EEG features and Machine Learning (ML) classifiers were used for this purpose. But no comprehensive comparison of these features and ML classifiers was presented. This paper presents the significance of twelve EEG features and five ML classifiers, commonly used in existing studies, for the detection of the eye-blink artifact. An EEG dataset, containing 2958 epochs of eye-blink, non-eye-blink, and eye-blink-like (non-eye-blink) EEG activities, is used in this study. The significance of each feature and classifier has been measured using accuracy, precision, recall, and f1-score. Experimental results reveal that scalp topography is the most potential among the selected features in detecting eye-blink artifacts. The best performing classifier is Artificial Neural Network (ANN) among the five classifiers. The scalp topography and ANN classifier combination performed as the most powerful feature-classifier combination. However, it is expected that the findings of this study will help the researchers to select the appropriate feature and classifier for their eye-blink detection studies in the future.

Introduction

Electroencephalography (EEG) is a non-invasive method of recording brain electrical potentials at several locations on the scalp [1]. EEG is a widely used tool in various research fields, including neural engineering, neuroscience, and biomedical engineering. In fact, it is the most common tool for diagnosing neurological disorders [2]. The brain-generated potentials are very much sensitive to noise signals that come from various sources other than the brain. These noise signals are termed as EEG artifacts. The major sources of EEG artifacts are - eye movement, eye blink, head movement, muscle movement, ECG pulse, and line noise [3, 4, 5, 6]. Among these, the eye blink artifact is one of the most destructive in nature. It distorts EEG recordings with such a high voltage that no meaningful information can be extracted from that contaminated portion of EEG recordings. As a result, potential information loss, and serious misinterpretation of EEG may occur that can lead to the wrong diagnosis of diseases and degrade BCI applications. Thus, properly detecting eye-blink artifacts from EEG signals and then removing these is a great concern to make the recorded EEG usable. Earlier, the eye-blink artifacts were detected by visual inspection of the Electrooculogram (EOG) channel [7]. EOG is a channel used to monitor the eye-activities by placing electrodes around the eyes [8, 9]. However, this manual detection process is time-consuming and difficult as well, especially for long-time EEG recordings. Besides, manual detection is not applicable for real-time EEG applications [10]. These limitations have led to developing the automatic approaches for eye-blink artifact detection. Researchers have been working on developing automatic ocular-artifact detection techniques for the last few decades and have achieved significant success. These works are commonly based on template or pattern matching and filtering techniques [11]. These techniques require the EOG channel to be given as reference input along with the EEG recordings. This constraint makes the techniques unsuitable for modern EEG-based applications because the modern wireless EEG systems have no built-in EOG channels. Besides, recording EOG signals by placing
electrodes next to the eyeball is inconvenient for both normal and clinical subjects. Therefore, researchers have focused on developing EOG channel-independent automatic systems for handling eye-blink artifacts. Several EEG feature-based automatic techniques were proposed that can identify eye-blink artifacts without incorporating the EOG channel. EEG feature corresponds to a distinctive piece of information that well-describes the nature of any event of EEG signal [12]. Several temporal, spatial, and frequency-based EEG features have been found using eye-blink detection approaches, such as kurtosis, skewness, variance, power spectral density (PSD), entropy, and scalp topography [13]. In [14], Shahbakhti et al. presented a stationary wavelet transform-based blink-elimination approach, where they introduced skewness as a criterion for recognizing blink-artifact from recorded EEG signals. As the amplitude of blink-contaminated EEG is comparatively high, it will produce a higher absolute value of skewness than that of clean EEG. Yadav et al. [15] employed kurtosis along with sample entropy as EEG features in their work. They combined Ensemble Empirical Mode Decomposition (EEMD) and Spatial Constraint ICA (SCICA) to remove the eye-blink artifacts. After performing SCICA followed by EEMD on the recorded EEG, the kurtosis, and the sample entropy of the Independent Components (ICs) were measured for detecting the contaminated ICs. These feature-based automatic techniques need to set a threshold value to confirm the blink artifacts. However, selecting an appropriate threshold is a challenging task. In recent years, machine learning (ML) algorithms have been incorporated to make this automatic and feature-based ocular artifact detection process more robust and accurate [8, 16–19]. The use of ML algorithms can remove the abovementioned problem of finding threshold values. In ML-based eye-blink detection techniques, several EEG features are selected and extracted from EEG recordings to train the ML models. Thus, the ML algorithms can decide whether an EEG segment is blink-contaminated or not. Supervised ML algorithms are preferably employed in this regard. Support Vector Machine (SVM) [19, 20, 21, 22, 23], Decision Tree (DT) [19], K-Nearest Neighbors (KNN) [19], Regression[17] etc. are commonly found using in related articles. However, a few studies investigated the use of unsupervised algorithms as well, such as K-Means Clustering [24, 25].

The above discussion clearly indicates that the success of an automatic eye-blink detection model mostly relies on selecting appropriate EEG features and ML classifiers. Hence, finding out the appropriate EEG feature that can highly describe the distinguishable property of eye-blink events is an important research issue. At the same time, employing a proper ML classifier will help to build a robust model to a greater extent. In this context, this study investigates the efficacy of twelve EEG features and five ML classifiers relevant to eye-blink artifact detection. Based on the investigation, a comparative and comprehensive discussion on their individual performances is held. So far, to our knowledge, this is the first initiative of comparative analysis on features and classifiers regarding EEG eye-blink artifacts.

**Methodology**

**Overview of the proposed method**

Figure 1 shows the general block diagram of the proposed work. The detail of each block is discussed below.
Dataset description

This study has explored two open-source repositories of EEG data, whose experimental protocols were almost identical and have common 14-EEG channels that were utilized in this study. EEG records taken from these two repositories have been merged to form the dataset of the proposed work. Figure 2 shows the 14 EEG channels considered in this work. There are 2958 EEG epochs of 14-channel and 04-second duration in our dataset, where an equal number (1479) of epochs comes from eye-blink and non-blink categories. The major portion of the dataset was taken from the first repository, which is 94.6% of the total dataset. This first repository contains EEG recordings of eye-blink and non-eye blink events collected from twenty healthy subjects. In a real-world scenario, the EEG of the clinical subjects may contain such brain activities (e.g., epileptic rhythm) that are morphologically close to the eye-blink signal. Hence, there is a high chance of removing these eye-blink-like events by wrongly detected as eye-blink-artifact which will result in the loss of important brain information. So, this type of event needs to be considered while building an eye-blink detection model to get a sophisticated and robust model. To incorporate an eye-blink-like-event, the proposed work introduced the second repository. This repository contains EEG recordings of epileptic subjects where many epileptic rhythms (eye-blink-like-events) are present in the interictal portion of the EEG signal. From this repository, 160 epochs of eye-blink-like-events have been extracted and added to the non-eye blink portion of our dataset. Around 10.8% of the total non-eye blink epochs of the dataset come from this repository. Figure 3 illustrates the graphical view of the dataset formation. More details of the two repositories can be found below.

The first repository is created by Kanoga et al. [26]. The signals were recorded at 14 scalp positions, namely Fp1, Fp2, F3, F4, T3, C3, Cz, C4, T4, P3, Pz, P4, O1, and O2 according to the international 10–20 placement. A vertical EOG signal was recorded as well. All the EEG and EOG data were band-pass filtered from 0.5 Hz to 60 Hz with a Butterworth filter. The sampling rate of the recordings is 256 Hz. 20 subjects (14 males and 6 females) participated in recording these EEG signals, whose mean age was 22.75 ± 1.45 years. Each subject took part in two types of experiments, one for voluntary eye-blink event and the other for involuntary (natural) eye-blink event. The recorded voluntary eye-blink signals are very clean and prominent. These do not have other brain activities than eye-blink, which is unrealistic. That is why this study prefers the involuntary portion of the repository so that the detection of natural eye-blink signals embedded in other brain activities can be examined. In the involuntary eye-blink experimental protocol, the subjects were seated in front of a laptop pc and were instructed to fix their eyes at a black cross-fixation point. Then 03 sounds ("A", "S", and "D") were presented in a randomized order with a 10–14 second gap. The task of the subjects was to press the keys ("A", "S", and "D") of the keyboard corresponding to the associated sound. Maintaining this protocol, three sessions, each including twenty trials, were held to record the EEG signal of involuntary eye-blink event. Each EEG recording is around 04 minutes long.

The second repository is created by Stevenson et al. [27]. This is a 19-channel EEG recording of 79 neonates who were admitted to the hospital due to suspicion of seizures. There are 79 EEG recordings of approximately one hour long. The sampling rate of the recordings is 256 Hz. The recordings were filtered
with a high-pass filter with a cut-off frequency of 0.5 Hz and a notch filter with a center at 50 Hz. According to the international 10–20 placement, all the EEG recordings were collected at 19 scalp positions, namely Fp2, F4, C4, P4, Fp1, F3, C3, P3, Fp2, F8, T4, T6, Fp1, F7, T3, T5, Fz, and Cz. From these 19 EEG channels, the 14 channels that were considered for this study (Fig. 02) were utilized.

**Epoching and labeling**

To prepare the raw 14-channel EEG signals usable for the experiment, epoching and labeling have been performed. Epoching is the segmentation of long-duration EEG signals into smaller chunks. In this study, each epoch is extracted as a 14-channel, 04-second duration EEG segment. Figure 4(a) shows a 14-channel 04-second duration EEG epoch containing an eye-blink artifact. After the epoching of the raw EEG, labeling of each epoch is performed. Labeling recognizes which category (here eye-blink and non-eye blink) an epoch belongs to. There was a recorded EOG signal associated with all the EEG recordings in the first repository. This EOG signal has been visually inspected by both the authors individually and epochs of blink and non-blink events have been extracted manually from the EEG recordings of this repository. Then the common epochs extracted by both the authors were selected as blink and non-blink epochs. After that, the associated EOG signal has been removed from all the epochs. 2798 epochs in total have been extracted in the abovementioned way, where 1479 epochs came from the eye-blink event and 1319 from the non-eye-blink event. Figure 4(b) and 4(c) are examples of blink and non-blink events. These are taken as 1-second segments for better visualization instead of a 4-second epoch. Both have the EOG signal at the topmost position. Epochs of the eye-blink-like event (epileptic rhythm) have been extracted from the second repository and added to the non-eye-blink epoch category. The second repository has no EOG channel along with the EEG recordings. Hence, topographic map was used here for extracting eye-blink-like epochs. Initially, both the authors visually inspected the EEGs and extracted epochs that deflected like eye-blink events. Then the common epochs, extracted by both the authors, were selected. Next, the topographic maps of all the selected epochs were generated to validate the correctness of the selection. Topographic maps were generated at the time point of maximum peak occurrence of the Fp1 channel for all the selected epochs. After that, the epochs whose corresponding topographic maps indicated having non-eye-blink activity were finalized to be added to the dataset of the proposed work. In this way, 160 non-eye blink (eye-blink-like) epochs have been extracted from the second repository. Figure 4(d) is an example of a non-eye blink (eye-blink-like) epoch of 1-second duration. The corresponding topographic map is shown at the bottom-left of the figure.

The total number of EEG epochs in the dataset becomes 2958, where 2798 epochs come from the first repository (1479 for blink, 1319 for non-blink), and the rest 160 (blink-like) epochs taken from the second repository (Fig. 3). Having an equal number of epochs (1479 epochs) from blink and non-blink categories has made the dataset balanced. Finally, all the eye-blink epochs were labeled as zero (0), whereas non-blink epochs were labeled as one (1). These labeled epochs were used for feature extraction in the next step.

**Feature extraction**
A feature is a piece of data that represents the characteristics of an event. EEG is a non-stationary signal; its’ properties change with time and frequency. Feature extraction aims to track these changes to effectively differentiate various events implicit in an EEG recording. Feature extraction can make a classifier model simpler and more comprehensible by reducing dataset dimensionality [28]. This study selected twelve features of temporal, spatial, entropy, and frequency domains to explore. Selected features are- kurtosis, skewness, variance, standard deviation, peak-to-peak amplitude, mean, max, min, normality test, entropy, scalp topography, and power spectral density. All the selected features were commonly used in earlier studies in identifying eye-blink artifacts. These twelve features have been extracted from our dataset and then fed to five ML classifiers one at a time to detect eye-blink artifacts. Table 1 describes the detail of the features, their formula, the domain these features belong to, considered EEG channels for calculation, and the list of the earlier works in which each feature was used.

Training and test set splitting

After feature extraction, the dataset has been split into training and test sets, maintaining an 80 – 20 percentage. As the dataset holds 2958 EEG epochs, after splitting, the training set got 2366 epochs and the test set got 592 epochs. After that, a 10-fold cross-validation approach is applied while training the classifier model.

Classifier training

To evaluate the potency of the selected features, a classifier is required to train using these features. Several ML algorithms have been found in earlier studies in handling eye-blink artifacts. From these, commonly used five classifiers have been selected for comparison in this work. Selected classifiers are- Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbors (KNN), Naïve Bayes (NB), and Artificial Neural Network (ANN). Classifier modeling is performed using the Scikit-learn library of the Python programming language. Each of the classifiers has been fed by one of the twelve features at a time.

Blink and non-blink EEG classification

After getting trained, classifiers could recognize EEG epochs as ‘blink epoch’ or ‘non-blink epoch’. Each of the five classifiers generates individual outputs for each of the twelve features. Thus, a total of 60 sets (twelve features × five classifiers) of output have been generated in this step.
### Performance measure

Four performance measuring metrics have been used to validate the implemented model; these are accuracy, precision, recall, and f1-score. Eye-blink artifact detection is a binary classification problem, where the two classes are – eye blink and non-eye blink. In this proposed model, the positive (target) class is eye-blink, and the negative class is non-eye-blink.

**Accuracy:** Accuracy is the ratio of correctly classified observations to the total observations.

\[
Accuracy = \frac{Correct\ observations}{Total\ observations}
\]

**Precision:** Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

\[
Precision = \frac{True\ positive}{True\ positive + False\ positive}
\]

**Recall:** Recall is the ratio of correctly predicted positive observations to all observations in the actual positive class.
Recall = \frac{Truepositive}{Truepositive + Falsenegative}

F1 score: The F1 score is the weighted average of precision and recall.

\[ F1\text{score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \]

Results

Table 2 shows an example of extracted feature values (scalp topography) of the dataset. A total of twelve such tables have been generated by extracting the corresponding twelve features. In Table 2, each row corresponds to an EEG epoch, whereas each column represents the feature-value of an EEG channel, and the last column labels each epoch as a blink (0) or a non-blink (1) category. Each of the five classifiers has been trained and tested with all the twelve features consecutively.

Comparison Of Features (Irrespective Of Classifiers)

Figure 5(a)-5(d) illustrate the comparative performances of the twelve EEG features based on accuracy, precision, recall, and f1-score as performance measuring metrics. The part of the figure for a particular performance measuring metric shows its’ highest scores, irrespective of classifiers, for all the twelve features.

For instance, the accuracy scores of scalp topography are (0.94, 0.88, 0.94, 0.92, and 0.97) % obtained with the five classifiers SVM, LR, KNN, NB, and ANN, respectively. Therefore, the best score, i.e., 0.97%, is shown in Fig. 5(a) as the accuracy score of scalp topography. Figure 5a shows the top accuracy scores of all the features. Here, scalp topography achieved the best score (0.97%) among all features. Max and peak-to-peak amplitude individually gained the second highest (0.96%) score close to the best score. The lowest score is 0.68% which is acquired by the feature mean.

According to the precision scores of Fig. 5(b), PSD, scalp topography, and max individually outperform by achieving 0.98%. The second highest score is 0.97% which is achieved by variance. Mean has been found as the worst feature scoring 0.73%. Several features individually achieved the top recall value (0.99%). These features are- scalp topography, variance, standard deviation, peak-to-peak amplitude, entropy, and max. The lowest recall value is 0.80%, which is gained by skewness.

While considering the F1 score, PSD, max, scalp topography, and peak-to-peak amplitude are the best features scoring 0.97%. The second-best feature is the standard deviation (0.96%). The worst feature is the mean by gaining 0.68%.

Comparison Of Classifiers (Irrespective Of Features)
For classifier comparison, in how many cases (features) a classifier achieved the highest score regarding a particular performance measuring metric is considered. In Fig. 6(a-d), the Y-axis value indicates the number of features for which a classifier scored best.

Figure 6a shows the comparison of classifiers based on accuracy. Here, ANN consistently scored top values for all the twelve features. SVM got the top accuracy score only with PSD, while other classifiers have no top accuracy score across any feature.

Figure 6b shows the classifier comparison under precision score. In this case, ANN got the top precision score across 07 features (PSD, max, mean, entropy, peak-to-peak amplitude, standard deviation, variance).

SVM achieved for six features, and LR, and NB achieved for 2 and 4 features, respectively. KNN has no top precision score with any feature.

While considering recall (Fig. 6c), KNN outperforms by getting the top score with 09 features. SVM and NB got top precision with 02 features, but LR and ANN have no success in this regard.

KNN again outperforms considering the F1 score. It achieved a top F1 score with 07 features. However, ANN, SVM, and NB achieved top for 5, 2, and 1 feature, respectively. LR has no achievement here.

Overall Result

To find out the best feature-classifier combination, the summation of the four performance measuring metrics scores of each feature-classifier combination is considered. The summation scale ranges from 0 to 4. Accordingly, scalp topography-ANN have found as the best combination, and their summation score is 3.87.

Figure 7 summarizes the overall results in a graphical view. The X-axis represents the features, whereas each column represents a classifier, and the stacked bars show the feature-classifier performances of the corresponding performance measuring metrics. Each Feature on X-axis takes five consecutive columns for the five classifiers- SVM, LR, KNN, NB, and ANN, respectively.

Discussion

Eye-blink artifact is the most deteriorative noise of the EEG signal. This artifact must be eliminated before processing of raw EEG signal. For identifying eye-blink artifacts, feature-based ML classification is an effective way. Hence, finding the most suitable EEG features that can prominently distinguish eye-blink artifacts is a crucial issue. At the same time, selecting the proper classifier that can effectively learn from the EEG features in recognizing eye-blink artifacts is also very important. In this article, a comparative study of EEG features and ML classifiers has been held to help the feature and classifier selection phase
while dealing with eye-blink artifacts of EEG signals. The performance of the features and classifiers was examined using a 14-channel EEG dataset with 2958 epochs of 04-second duration.

The findings of this study reveal that scalp topography is the most prominent feature among the selected twelve for describing the distinctive nature of eye-blink events. Maximum, power spectral density and peak-to-peak amplitude also show remarkable performances. On the contrary, mean and skewness are the weakest features in differentiating eye-blink artifacts. These both got very poor scores against most of the performance measuring metrics.

In the case of the classifiers, ANN outperformed other selected classifiers in recognizing eye-blink artifacts from EEG signals. It achieved consistently high scores for almost all performance measuring matrices. However, SVM also performs well while considering precision. KNN achieved a good score for recall or f1-score. However, LR and NB found as less efficient in classifying eye-blink artifacts.

While considering both features and classifiers, scalp topography-ANN is the best performing combination.

The presented work considers an eye-blink-like EEG event by including EEG recordings of epileptic rhythm in the dataset. However, it is planned to investigate more eye-blink-like events in future work. In addition, the performance of multiple EEG feature combinations will also be examined in future work.

**Conclusion**

This study aims to facilitate the detection of eye blink artifacts from EEG signals by performing a comparative analysis on twelve EEG features and five ML classifiers. The recognition rate of the eye-blink artifact mostly depends on selecting appropriate features that can distinguish eye-blink events mostly. At the same time, proper classifier selection is also a very important parameter for success. This study deals with the aforesaid issues of finding the appropriate features and classifiers for eye blink detection. It covers twelve features: kurtosis, skewness, variance, standard deviation, peak-to-peak amplitude, mean, max, min, normality test, entropy, scalp topography, and power spectral density (PSD). It is found that scalp topography is the most powerful feature in classifying eye blink artifacts. On the contrary, mean and skewness showed significantly lower recognition rates comparing others. However, other remaining features performed with a good recognition rate. In the case of the classifiers, ANN outperformed other classifiers. KNN and SVM were also found as efficient classifiers. LR and NB have performed weakly in detecting blink-artifact. The combination of scalp topography and ANN outperformed other combinations. However, it is expected that this study will work as a good reference for feature and classifier selection and will smooth the detection process of eye-blink artifacts from EEG signals in future research.

**Declarations**

**Compliance with Ethical Standards**
Funding: This research did not receive any funding.

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical Approval: This article does not contain any studies with human participants performed by any of the authors.

References


Table 1

Table 1 is available in the Supplementary Files section.

Figures
Figure 1

Block diagram of the proposed work
Figure 2

Electrode diagram of 14 channels (according to international 10-20 standard) considered in the proposed work
Figure 3

Dataset formation of the proposed work
Figure 4

EEG segments of this study (a): Sample EEG epoch of 04-second duration; (b), (c), (d): segments of eye-blink, non-eye-blink, eye-blink-like events (epileptic rhythm) of 01 second duration respectively
Figure 5

Performance comparison of the features: (a), (b), (c), (d) based on accuracy, precision, recall, and f1-score respectively
Figure 6

Performance comparison of the classifiers: (a), (b), (c), (d) accuracy, precision, recall, and F1 score, respectively.
Figure 7

Graphical view of the overall results of this study

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.jpg