ORIGINAL ARTICLE

Feature Extraction from Regenerated EEG: A Better Approach for ICA Based Eye Blink Artifact Detection

Maliha Rashida, Mohammad Ashfak Habib *

Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chittagong-4349, Bangladesh

*Corresponding Author: Mohammad Ashfak Habib
Email: ashfak@cuet.ac.bd

Received: 15 October 2022 / Accepted: 04 February 2023

Abstract

Purpose: Independent Component Analysis (ICA) decomposition is a commonly used technique for eye blink artifact detection from Electroencephalogram (EEG) signals. Feature extraction from the decomposed ICs is a prime step for blink detection. This paper presents a new model of eye blink detection for ICA based approach, where the decomposed ICs are projected to their corresponding EEG segments (ReEEG), and feature extraction is performed on the ReEEG instead of the IC. ReEEG represents the eye blink activity more distinctly. Hence, ReEEG-based feature extraction is more potential in detecting eye blink artifacts than the traditional IC-based feature extraction.

Materials and Methods: This paper employs twelve EEG features to substantiate the superiority of ReEEG over IC. Support Vector Machine (SVM) is used as a classifier. A dataset, having 2638 clinical EEG epochs, is employed. All the considered twelve features are extracted from ReEEG and fed to SVM one at a time for blink detection. Then the obtained results are compared with an IC-based model with the same features.

Results: The comparison reveals the success of the proposed ReEEG-based blink detection approach over the traditional IC-based approach. Accuracy, precision, recall, and f1 scores are calculated as performance measuring metrics. For almost all features, ReEEG-based approach achieved up to 12.25% higher accuracy, 24.95% higher precision, 13.49% higher recall, and 12.89% higher f1 score than the IC-based traditional method.

Conclusion: The proposed model will be useful for researchers in dealing with the eye blink artifacts of EEG signals with more efficacy.

Keywords: Electroencephalogram; Eye Blink Artifact; Independent Component Analysis; Support Vector Machine; Feature Extraction.

DOI: https://doi.org/10.18502/fbt.v11i2.15336

1. Introduction

Electroencephalogram (EEG) is a popular technique in the field of medical diagnostics, as well as in research like neuroscience, cognitive science, etc. [1–3]. EEG-based Brain Computer Interface (BCI) is growing rapidly in recent years [4]. Controlling a wheelchair [5], controlling a robotic arm [6], neurorehabilitation [7], etc. are some examples of EEG-based BCI applications.

EEG signal is highly sensitive to artifacts that destroy the signal quality such that no meaningful information can be extracted from that artifact-contaminated EEG portion. These artifacts can be physiological (internal) or non-physiological (external). Physiological artifacts are the ones that are generated due to - eye movement, eye-blink, head movement, muscle movement, ECG pulse, etc. Non-physiological artifacts are created from line noise, electrode pop-ups, etc. [8]. Proper shielding and grounding of cables can prevent non-physiological artifacts. Physiological ones are the main challenge to handle. Among other physiological artifacts, the eye blink artifact is the most concerning, as it is unavoidable and occurs very frequently. Eye blink is a natural, biological phenomenon with an occurrence of 15 times per minute on average [9]. The frequency band of the eye blink signal overlaps with that of the EEG signal that makes the detection approach of blink artifacts more challenging than any other artifacts [10, 11]. Extensive research works aiming to recover the artifact-free EEG signal have been held for many years. Several methods and algorithms, such as regression, filtering, independent component analysis, wavelet transform, empirical mode decomposition, etc. are used widely for this purpose. More details of the methods are available in [12]. Independent Component Analysis (ICA), compared to the others, is the most popular and powerful technique in handling eye blink artifacts [12–14]. After decomposing the contaminated signal into its source components (ICs), the ICs responsible for blink artifact are identified and removed either manually or automatically.

Earlier, the automatic identification approach was performed based on a cross-correlation with Electrooculogram (EOG) reference channel. However, to avoid the EOG dependency for making the models feasible in a broader scope, feature-based detection approaches have become favorable. It reduces the dimensionality of input data as well. Kurtosis, skewness, entropy, scalp topography, Power Spectral Density (PSD), etc. are some of the widely used features to decide about blink contamination [15, 16]. In ICA-based approaches, the considered EEG feature is extracted from the decomposed ICs [17–24]. An IC represents a single EEG activity, and the feature values extracted from the decomposed IC characterize that activity.

Traditionally, a predefined threshold is set to separate the contaminated IC using the feature values. Nowadays, Machine Learning (ML) algorithms are incorporated instead of thresholding, and enhanced performance is achieved [14, 20, 25, 26]. Support Vector Machine (SVM) [10, 18, 26–28], Decision Tree (DT) [26], K-Nearest Neighbors (KNN) [26], Regression [20], K-Means Clustering [22, 29], etc. are commonly used in existing works. Although both supervised and unsupervised ML algorithms are applied, supervised is found to use more extensively.

A decomposed IC can be projected to its corresponding multichannel EEG signal. That is, a raw multichannel EEG can be decomposed into a set of ICs, and afterward, each IC can be converted to a new multichannel EEG. This later EEG is termed as ReEEG (regenerated EEG) in this study. Since a ReEEG comes from a single IC, it represents a single neural activity. This paper presents a model based on the concept that if feature extraction is performed on ReEEG instead of IC, it could characterize the neural activity more prominently. And, hence, the eye blink detection performance will improve significantly. Accordingly, the proposed model is prepared to take SVM as a classifier. Twelve EEG features are considered to test the model's performance. Each feature is extracted from the ReEEG and fed to the classifier one at a time. The detection performance of the proposed model is significantly higher for almost all features than that of an IC-based model.

The rest of this paper includes the materials and method in chapter 2, results and discussion in chapter 3, and finally chapter 4 concludes the paper.

2. Materials and Methods

2.1. Dataset

The dataset for the proposed work has been taken from an open-source repository namely Voluntary-Involuntary Eye-Blinks repository [30]. Twenty subjects participated in an experiment where both voluntary and involuntary (natural) eye blinks were recorded on 14 electrodes. The considered 14 electrodes are graphically shown in Figure 1. However, the involuntary portion alone is used for the proposed work. 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 signals of an involuntary eye-blink event. Each EEG recording is around 04 minutes long with a sampling rate of 256 Hz.

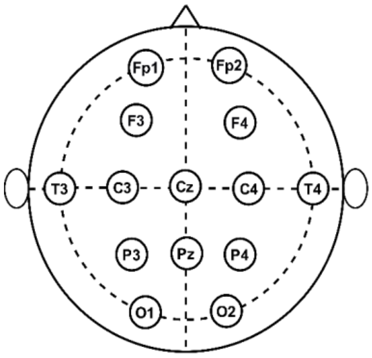


Figure 1. Considered electrodes of the dataset

2.2. Proposed Methodology

Figure 2 illustrates the graphical overview of the proposed model. Each block of the model is discussed sequentially in the next parts.

2.2.1. Epoching and Labeling

The raw EEG recordings with 14 channels and a 256 Hz sampling rate is taken as input to the model. Epoching is performed on the long EEG recordings maintaining 4-second duration. An equal number of blink and non-blink (clean) epochs has been extracted by the authors by visually inspected the associated EOG channel. The epochs are verified further by plotting the topographic map and then labeled as blink or non-blink epochs. Finally, a total of 2638 EEG epochs of 4-second duration is obtained containing an equal number (1319) from blink and non-blink categories.

2.2.2. ICA Decomposition and Generation of ReEEG

Each EEG epoch is decomposed into 14 independent components (ICs). Decomposition is performed by InfoMax ICA algorithm [31]. In a blink epoch, there must be one blink IC in its decomposed set of ICs. Such blink ICs were collected by visual inspection from all the generated ICs from 1319 blink epochs. Similarly, 1319 clean ICs were collected from the IC sets of 1319 clean EEG epochs. Thus, a total number (2638) of ICs was collected for further processing.

Later, each of the collected IC is gone through the Inverse ICA algorithm to be back-projected into its corresponding 14-channel EEG segment (ReEEG). The Inverse ICA is applied using runica function of EEGLAB toolbox [32]. The generated ReEEGs are used for feature extraction in the next step. It is noted that a normal EEG epoch (Figure 3a) represents all the brain activities that occurred within that time duration. On the contrary, a ReEEG is the outcome of a single IC (Figure 3b). As, one IC represents one brain activity, its corresponding ReEEG also represents only that single brain activity, which makes a ReEEG different from a normal EEG signal.

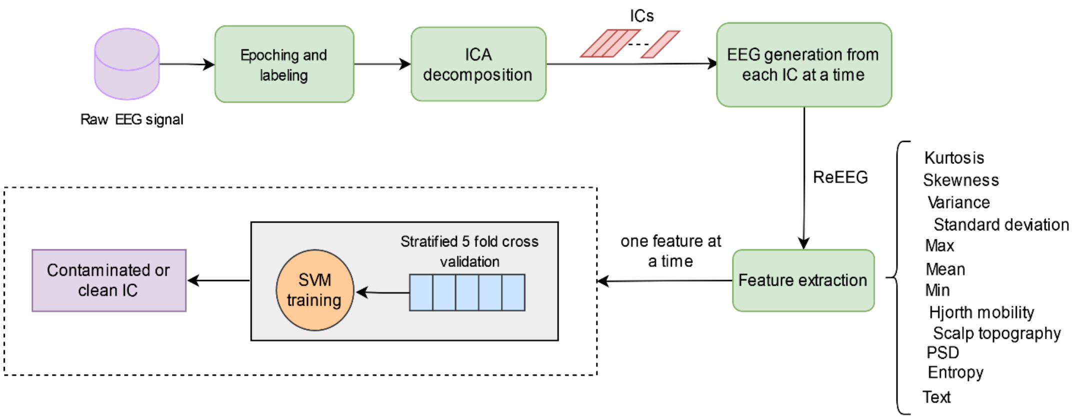
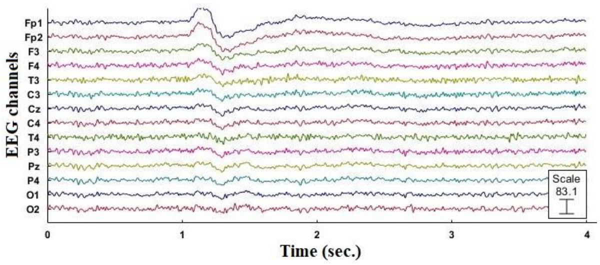
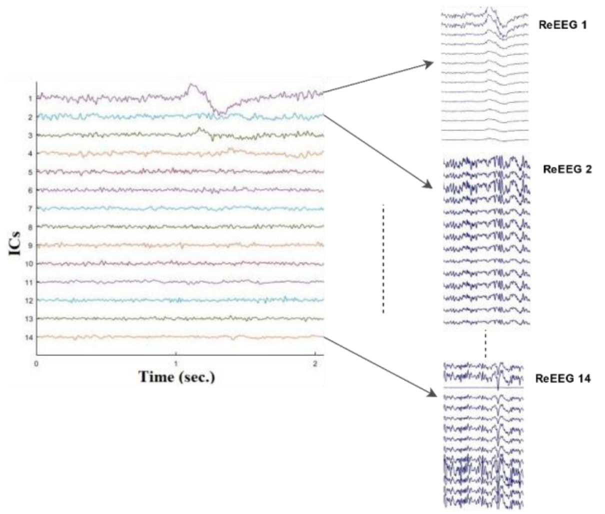


Figure 2. Graphical view of the proposed model



a



b

Figure 3. a): EEG epoch, b): Decomposed ICs and regenerated EEG (ReEEG) from each IC

2.2.3. Feature Extraction

Twelve EEG features are considered to evaluate the effectiveness of ReEEG-based feature extraction. These are the frequently used features found in relevant literature. The considered features are - kurtosis, skewness, entropy, scalp topography, standard deviation, variance, peak-to-peak amplitude, mean, power spectral density, max, min, Hjorth mobility. All these twelve features are extracted from the ReEEGs sequentially. More detail of the features can be found in Table 1.

2.2.4. Detection of Eye Blink Event

Support Vector Machine (SVM) is chosen as a classifier for this work because of its robustness and extensive use in the existing literature [10, 18, 26–28]. The extracted features from ReEEGs are fed to the SVM classifier, maintaining a stratified 5-fold cross-validation approach. Features are applied to the classifier one at a time and the individual detection outcomes for all the features are obtained and stored.

Table 1. Considered feature-details

Feature	Domain	Used in
Kurtosis	Time	10, 21, 23, 31–33
Skewness	Time	9, 25, 26, 32
Entropy	Entropy	25–27, 33
Scalp topography (S. topograph)	Spatial	17, 20, 34, 35
Standard deviation (Std dev)	Time	25, 27, 36
Variance	Time	10, 26
Peak-to -peak amplitude (P2P amplitude)	Time	10, 26
Mean	Time	25, 27
Power spectral density (PSD)	Frequency	20, 37
Max	Time	36
Min	Time	36
Hjorth mobility (H. mobility)	Time	23

2.2.5. System Evaluation Metrics

For evaluating the artifact detection performance of the model, accuracy, precision, recall, and f1 score are measured. The calculation of these four metrics is stated below.

Accuracy is the ratio of correctly classified observations to the total observations (Equation 1):

$$Accuracy = \frac{Correct\ observations}{Total\ observations} \quad (1)$$

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations (Equation 2):

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \quad (2)$$

Recall is the ratio of correctly predicted positive observations to all observations in the actual positive class (Equation 3).

$$Recall = \frac{True\ positive}{True\ positive + False\ negative} \quad (3)$$

The F1 score is the weighted average of precision and recall (Equation 4).

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

In this study, eye blink-contaminated IC is considered a true positive class, and clean IC is considered a true negative class.

3. Results and Discussion

Figure 3 shows an EEG epoch and the decomposition of the epoch into Independent Components (ICs). For better visualization, a smaller portion of a 4-second IC is shown here. The right portion of Figure 3b shows how an IC is back-projected to the corresponding Regenerated EEG (ReEEG).

The SVM classifier takes the features, extracted from ReEEG, and decides whether the respective ReEEG is blink-contaminated or clean. If a ReEEG is detected as blink-contaminated, that means the corresponding IC of that ReEEG is also contaminated.

The detection outcomes for individual features are measured in terms of accuracy, precision, recall, and f1 score. Table 2 summarizes the blink detection performance of the proposed model for individual features.

To better realize the success of ReEEG-based proposed model over the traditional IC-based model, another model is prepared maintaining the same setup and dataset. The only difference is - feature values are extracted from the Independent Components (ICs) instead of ReEEGs in this model. Table 3 summarizes the detection performance of this IC-based model. The outcomes of both models are compared and graphically illustrated in Figure 4a-4d.

Table 2. Detection outcomes of individual features for ReEEG-based model

Feature	Accuracy	Precision	Recall	F1 score
Entropy	0.93	0.95	0.91	0.93
Hjorth mobility	0.85	0.88	0.80	0.84
Kurtosis	0.85	0.91	0.77	0.84
Max	0.94	0.96	0.92	0.94
Mean	0.56	0.75	0.19	0.30
Min	0.93	0.94	0.93	0.93
Peak2peak	0.90	0.92	0.87	0.89
PSD	0.93	0.96	0.90	0.93
Skewness	0.71	0.91	0.47	0.62
Standard deviation	0.94	0.94	0.93	0.94
Variance	0.93	0.95	0.91	0.93
Topography	0.94	0.96	0.92	0.94

Table 3. Detection outcomes of individual features for IC-based model

Feature	Accuracy	Precision	Recall	F1 score
Entropy	0.89	0.93	0.84	0.88
Hjorth mobility	0.84	0.87	0.80	0.84
Kurtosis	0.85	0.91	0.77	0.84
Max	0.92	0.94	0.90	0.92
Mean	0.50	0.50	0.48	0.49
Min	0.87	0.88	0.86	0.87
Peak2peak	0.77	0.80	0.73	0.76
PSD	0.92	0.96	0.88	0.92
Skewness	0.71	0.92	0.46	0.62
Standard deviation	0.90	0.91	0.88	0.90
Variance	0.88	0.92	0.84	0.88
Topography	0.92	0.93	0.90	0.92

In Figure 4, for almost all the features, the SVM classifier showed enhanced performance while the feature extraction was held on ReEEG. Figure 5 shows the percentage of maximum improved performance for ReEEG over IC on the four-measuring metrics. In accuracy, the maximum value of improvement for ReEEG is 12.25% that is achieved by peak-to-peak amplitude. This same feature (peak-to-peak amplitude) obtained the maximum recall and f1 scores that are 13.49% and 12.89%, respectively. For precision, the maximum improvement is scored by mean that is 24.95%.

However, from Figure 4, it is found that mean, skewness, and PSD showed a little decrease in particular metrics; specifically mean decreased in recall and f1 score, skewness decreased in accuracy and precision, and PSD decreased in precision. One drawback of the mean is- it is outlier sensitive. This study used a clinical EEG dataset that may contain noises other than eye-blink. This could be the possible reason for the fall in performance for the mean.

Skewness tells about the asymmetry of signal distribution. The position of the outlier is exposed by skewness. For a particular blink contamination, the position of blink is the same in IC and in ReEEG. Thus, ReEEG has not much elaborated information here to provide the classifier. Hence, for skewness, no improved performance is achieved by ReEEG-based detection.

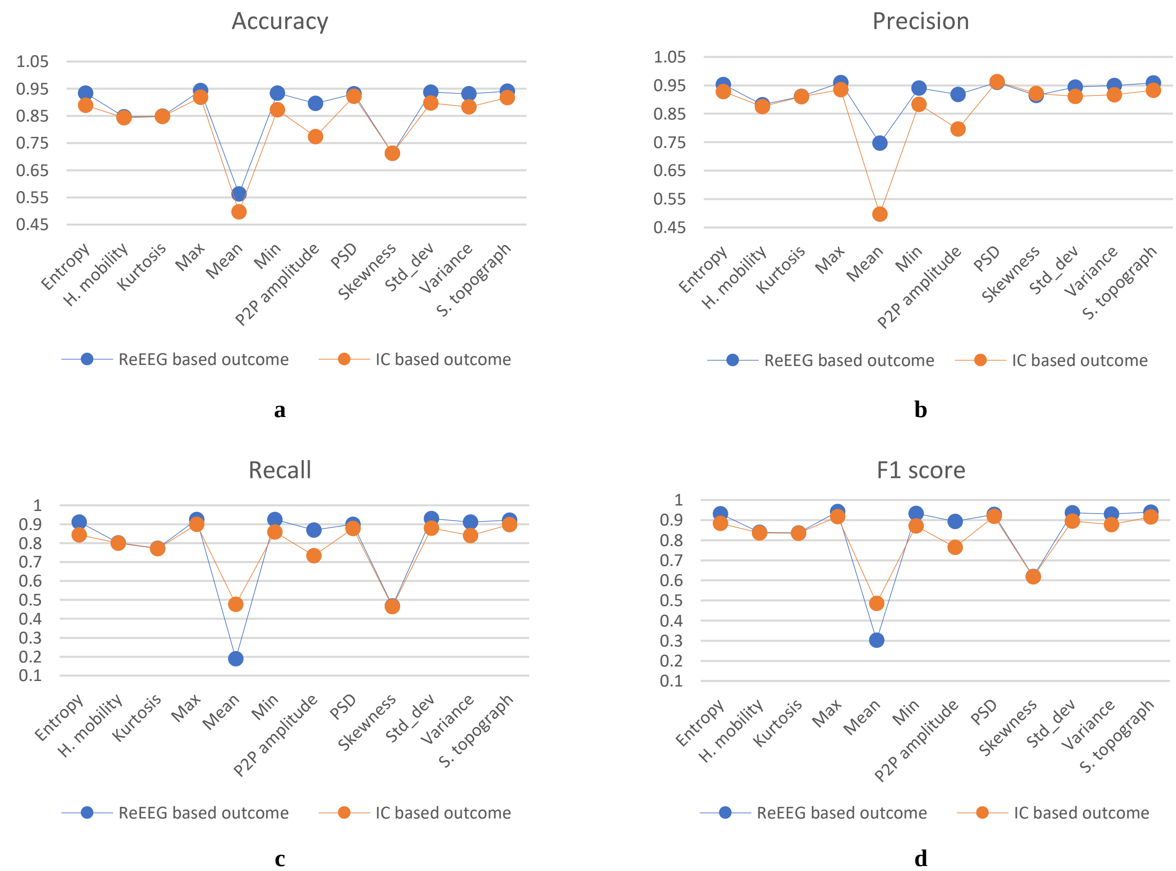


Figure 4. Comparative results of regenerated EEG-based, and IC-based model in terms of (a): accuracy, (b): precision, (c): recall, and (d): f1 score

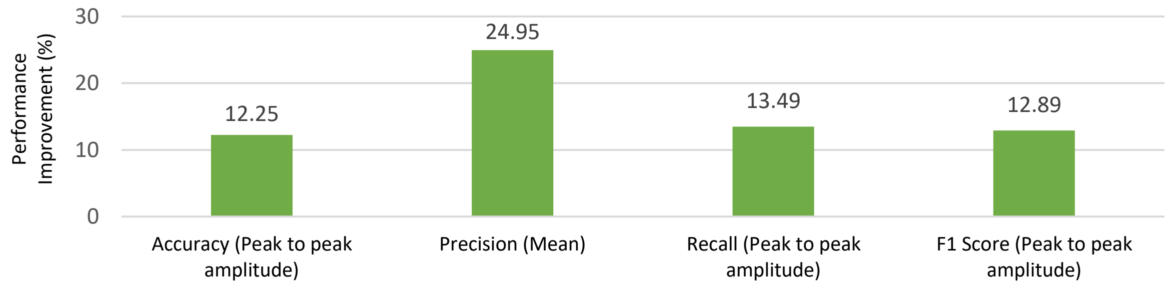


Figure 5. Maximum of performance improvement for ReEEG over IC

PSD distributes the power content of EEG as a function of frequency. Both IC and its corresponding ReEEG represent the same brain activity that lies in the same frequency band. Like skewness, ReEEG here has no extra information to provide the classifier. Therefore, ReEEG-based detection could not gain significant improvement in this case.

Among the twelve considered features, three showed slightly degraded or unimproved performance as discussed above. All the other features provided remarkably increased performance with ReEEG-based detection.

It is obvious that a multichannel ReEEG can characterize the eye blink event more prominently than a single IC. This study reveals the same by showing that the ReEEG-based blink detection approach outperformed the IC-based approach for most of the cases.

4. Conclusion

This paper presents a model with an innovative concept of feature extraction for ICA-based eye blink artifact detection approach. In ICA-based models, feature extraction is performed on the decomposed ICs of the

EEG signal. The proposed model shows that feature extraction, performed on regenerated EEG (ReEEG) instead of IC, will generate significantly increased performance in most cases. The proposed work employed twelve EEG features to evaluate the model and compared the performance with a traditional IC-based model. The comparative results reveal the success of the proposed model. It is expected that the proposed model will be helpful for the researchers in dealing with eye blink artifacts of EEG signals with more efficacy.

References

1- Carlo Miniussi and Gregor Thut, "Combining TMS and EEG offers new prospects in cognitive neuroscience." *Brain topography*, Vol. 22pp. 249-56, (2010).

2- Steven M Peterson, Emily Furuichi, and Daniel P Ferris, "Effects of virtual reality high heights exposure during beam-walking on physiological stress and cognitive loading." *PloS one*, Vol. 13 (No. 7), p. e0200306, (2018).

3- Ruxandra I Tivadar and Micah M Murray, "A primer on electroencephalography and event-related potentials for organizational neuroscience." *Organizational Research Methods*, Vol. 22 (No. 1), pp. 69-94, (2019).

4- Swati Aggarwal and Nupur Chugh, "Review of machine learning techniques for EEG based brain computer interface." *Archives of Computational Methods in Engineering*, pp. 1-20, (2022).

5- ZT Al-Qaysi, BB Zaidan, AA Zaidan, and MS Suzani, "A review of disability EEG based wheelchair control system: Coherent taxonomy, open challenges and recommendations." *Computer methods and programs in biomedicine*, Vol. 164pp. 221-37, (2018).

6- Rihab Bousseta, I El Ouakouak, M Gharbi, and F Regragui, "EEG based brain computer interface for controlling a robot arm movement through thought." *Irbm*, Vol. 39 (No. 2), pp. 129-35, (2018).

7- Anirban Chowdhury, Haider Raza, Yogesh Kumar Meena, Ashish Dutta, and Girijesh Prasad, "An EEG-EMG correlation-based brain-computer interface for hand orthosis supported neuro-rehabilitation." *Journal of neuroscience methods*, Vol. 312pp. 1-11, (2019).

8- Ibrahim Kaya, "A brief summary of EEG artifact handling." *Brain-computer interface*, (No. 9), (2019).

9- Suguru Kanoga, Atsunori Kanemura, and Hideki Asoh, "Multi-scale dictionary learning for ocular artifact reduction from single-channel electroencephalograms." *Neurocomputing*, Vol. 347pp. 240-50, (2019).

10- Souvik Phadikar, Nidul Sinha, and Rajdeep Ghosh, "Automatic eyeblink artifact removal from EEG signal using wavelet transform with heuristically optimized threshold." *IEEE Journal of Biomedical and Health Informatics*, Vol. 25 (No. 2), pp. 475-84, (2020).

11- Mohammad Shahbakhti, Maxime Maugeon, Matin Beiramvand, and Vaidotas Marozas, "Low complexity automatic stationary wavelet transform for elimination of eye blinks from EEG." *Brain sciences*, Vol. 9 (No. 12), p. 352, (2019).

12- Rakesh Ranjan, Bikash Chandra Sahana, and Ashish Kumar Bhandari, "Ocular artifact elimination from electroencephalography signals: A systematic review." *Biocybernetics and Biomedical Engineering*, Vol. 41 (No. 3), pp. 960-96, (2021).

13- Malik Muhammad Naeem Mannan, Muhammad Ahmad Kamran, and Myung Yung Jeong, "Identification and removal of physiological artifacts from electroencephalogram signals: A review." *Ieee Access*, Vol. 6pp. 30630-52, (2018).

14- Sari Sadiya, Tuka Alhanai, and Mohammad M Ghassemi, "Artifact detection and correction in eeg data: a review." in *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, (2021): *IEEE*, pp. 495-98.

15- Md Kafiul Islam, Amir Rastegarnia, and Zhi Yang, "Les méthodes de détection et de rejet d'artefact de l'EEG de scalp: revue de littérature." *Neurophysiologie Clinique*, Vol. 46 (No. 4-5), pp. 287-305, (2016).

16- Maliha Rashida and Mohammad Ashfak Habib, "Quantitative EEG features and machine learning classifiers for eye-blink artifact detection: A comparative study." *Neuroscience Informatics*, Vol. 3 (No. 1), p. 100115, (2023).

17- Priyalakshmi Sheela and Subha D Puthankattil, "A hybrid method for artifact removal of visual evoked EEG." *Journal of neuroscience methods*, Vol. 336p. 108638, (2020).

18- Raheel Zafar, Abdul Qayyum, and Wajid Mumtaz, "Automatic eye blink artifact removal for EEG based on a sparse coding technique for assessing major mental disorders." *Journal of Integrative Neuroscience*, Vol. 18 (No. 3), pp. 217-29, (2019).

19- Alejandro Villena, Lorenzo J Tardón, Isabel Barbancho, Ana M Barbancho, Elvira Brattico, and Niels T Haumann, "Preprocessing for lessening the influence of eye artifacts in EEG analysis." *Applied Sciences*, Vol. 9 (No. 9), p. 1757, (2019).

20- Thea Radüntz, Jon Scouten, Olaf Hochmuth, and Beate Meffert, "Automated EEG artifact elimination by applying machine learning algorithms to ICA-based features." *Journal of neural engineering*, Vol. 14 (No. 4), p. 046004, (2017).

21- Jun Feng Gao, Yong Yang, Pan Lin, Pei Wang, and Chong Xun Zheng, "Automatic removal of eye-movement and blink artifacts from EEG signals." *Brain topography*, Vol. 23pp. 105-14, (2010).

22- Salim Çınar, "Design of an automatic hybrid system for removal of eye-blink artifacts from EEG recordings." *Biomedical Signal Processing and Control,* Vol. 67p. 102543, (2021).

23- Mohamed F Issa and Zoltan Juhasz, "Improved EOG artifact removal using wavelet enhanced independent component analysis." *Brain sciences,* Vol. 9 (No. 12), p. 355, (2019).

24- Ajay Kumar Maddirala and Kalyana C Veluvolu, "ICA with CWT and k-means for Eye-blink Artifact Removal from Fewer Channel EEG." *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* Vol. 30pp. 1361-73, (2022).

25- Xiao Jiang, Gui-Bin Bian, and Zean Tian, "Removal of artifacts from EEG signals: a review." *Sensors*, Vol. 19 (No. 5), p. 987, (2019).

26- Elena Nedelcu, Raluca Portase, Ramona Tolas, Raul Muresan, Mihaela Dinsoreanu, and Rodica Potolea, "Artifact detection in EEG using machine learning." in *2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, (2017): *IEEE*, pp. 77-83.

27- Jianhui Wang, Jiuwen Cao, Dinghan Hu, Tiejia Jiang, and Feng Gao, "Eye blink artifact detection with novel optimized multi-dimensional electroencephalogram features." *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* Vol. 29pp. 1494-503, (2021).

28- K Yasoda, RS Ponmagal, KS Bhuvaneshwari, and K Venkatachalam, "Automatic detection and classification of EEG artifacts using fuzzy kernel SVM and wavelet ICA (WICA)." *Soft Computing,* Vol. 24pp. 16011-19, (2020).

29- Omid Dehzangi, Alexander Melville, and Mojtaba Taherisadr, "Automatic eeg blink detection using dynamic time warping score clustering." in *Advances in Body Area Networks I: Post-Conference Proceedings of BodyNets 2017*, (2019): *Springer*, pp. 49-60.

30- Suguru Kanoga, Masaki Nakanishi, and Yasue Mitsukura, "Assessing the effects of voluntary and involuntary eyeblinks in independent components of electroencephalogram." *Neurocomputing,* Vol. 193pp. 20-32, (2016).

31- Dominic Langlois, Sylvain Chartier, and Dominique Gosselin, "An introduction to independent component analysis: InfoMax and FastICA algorithms." *Tutorials in Quantitative Methods for Psychology*, Vol. 6 (No. 1), pp. 31-38, (2010).

32- Arnaud Delorme and Scott Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis." *Journal of neuroscience methods,* Vol. 134 (No. 1), pp. 9-21, (2004).