

# Analysis of Eye Blinking EEG Artifacts: Exploring Clinical Significance

Meixian Wu Trinity College Dublin wume@tcd.ie

# ABSTRACT

This paper aims to emphasize the underlying significance of blinking behavior in EEG signals and its relationship to the diagnosis and identification of specific disorders. The study utilizes the UCI Machine Learning Storage EEG-Eye-State dataset to analyze the correlation between electrode contact positions during open or closed eye states and their corresponding data. The KMO test is employed to identify three sets of feature vectors. Factor analysis is applied with variance maximization for factor orthogonal rotation, followed by logistic regression to validate the predictive role of brain frontal lobe electrical signals in relation to blinking behavior. Furthermore, this paper reviews the neurophysiological mechanisms of spontaneous blinking and explores its association with dopamine-mediated cognitive behaviors and various psychiatric cognitive disorders. The paper also provides a review of feature extraction techniques for blink artifacts and commonly used machine learning classification algorithms. Additionally, the integration of machine learning algorithms with clinical research is discussed to elucidate the potential significance of blinking behavior in EEG signals.

# CCS CONCEPTS

• Applied computing  $\rightarrow$  Life and medical sciences.

## **KEYWORDS**

EEG, Eye blinking, BCI

#### **ACM Reference Format:**

Meixian Wu. 2023. Analysis of Eye Blinking EEG Artifacts: Exploring Clinical Significance. In 2023 15th International Conference on Bioinformatics and Biomedical Technology (ICBBT 2023), May 26–28, 2023, Xi'an, China. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3608164.3608212

# **1** INTRODUCTION

# 1.1 The Neurophysiological Mechanisms of Blinking

Blinking, a ubiquitous behavior, can be classified into three main types: spontaneous, voluntary, and reflex blinking. In healthy individuals, spontaneous blinking is primarily regulated by the innervation of the levator palpebrae superioris (LP) and the orbicularis oculi



This work is licensed under a Creative Commons Attribution International 4.0 License.

ICBBT 2023, May 26–28, 2023, Xi'an, China © 2023 Copyright held by the owner/author(s). ACM ISBN 979-8-4007-0038-5/23/05. https://doi.org/10.1145/3608164.3608212

(OO) muscles[1]. Investigation into the neurophysiological mechanisms of spontaneous blinking in rodent models has identified the spinal trigeminal complex as a key component of the spontaneous blink generator[12]. timulation of the corneal nerves, resulting from tear film disruption, leads to excitation of the spinal trigeminal complex, thereby controlling the average interblink interval (IBI) of spontaneous blinking[2]. Dopamine also exerts modulatory effects on spontaneous blinking, as evidenced by the administration of dopamine agonists like apomorphine, which effectively reduce the amplitude of both trigeminal reflex and spontaneous blinking. The underlying neural mechanisms involve the modulation of trigeminal reflex blink amplitude and excitability by basal ganglia dopamine levels. This modulation occurs through the inhibition of the superior colliculus by the substantia nigra pars reticulata, the excitation of the nucleus raphe magnus by the superior colliculus, and the subsequent inhibition of the spinal trigeminal complex by the nucleus raphe magnus[7].

Numerous studies have demonstrated that the rate of spontaneous blinks (EBR) serves as a non-invasive, indirect marker of central dopamine (DA) function, with higher EBR reflecting increased DA activity. Dopaminergic signaling is also believed to be implicated in various neural systems involved in cognitive processes. For instance, the neural circuits associated with reward effects and addictive behaviors, including the basal ganglia (striatum), limbic system, and prefrontal cortex, exhibit expression of dopamine receptor type 1 and dopamine receptor type 2[18]. Importantly, the EBR is highly correlated with dopaminergic activity in the frontal striatum, rendering it an optimal non-invasive index[10].

In addition, the use of central anticholinergic medications has been proposed to impact the cholinergic-dopaminergic system balance by reducing acetylcholine neurotransmission in the striatum. This modulation, in turn, influences the regulation of cognitive states and movement control[3]. Notably, in attention network tests (ANT), the administration of central anticholinergic drugs acting on muscarinic (M) receptors, such as promethazine, has been shown to induce changes in blink frequency, indicative of increased activity during cognitive activation conditions compared to the resting state[19]. Thus, blink rate is considered one of the markers of central cholinergic pathway activity.

Overall, a comprehensive understanding of the neurophysiological mechanisms underlying blinking provides insights into the intricate interplay between dopaminergic and cholinergic systems in modulating blink behavior. These findings contribute to the broader understanding of cognitive processes and shed light on the potential clinical implications of blink abnormalities in various neurological and psychiatric conditions.

# 1.2 The Origins and identification of blink artifacts in EEG signals

In the realm of electroencephalogram (EEG) signals, blink artifacts emerge from two primary sources. Firstly, the cornea, in conjunction with the retinal structure, establishes an electric dipole that generates an electric field propagating throughout the cranium. This phenomenon is a consequence of the swift motion of the eyelids during blinking, which engenders blink artifacts[6]. It is worth noting that blink artifacts not only engender distortion within the delta (and theta) frequency bands but also introduce high-frequency components into the alpha and beta bands of the EEG signal, which predominantly stem from neural origins[8].

Independent Component Analysis (ICA) is a widely employed blind source separation (BSS) technique. Its primary objective is to identify an optimal separation matrix by employing iterative algorithms such as gradient-based optimization methods, with the aim of enhancing the independence of the separated components. However, due to the inherent uncertainty regarding the order and amplitude of the ICA-separated signals, accurate discrimination between significant neural activity and artifacts poses a considerable challenge. To enhance the effectiveness and real-time capabilities of eye movement removal, researchers have introduced the concept of entropy or combined it with other methodologies[16]. For instance, the Informax (Information Maximization) algorithm, based on a single-layer feedforward architecture for maximizing information transmission, facilitates feature enhancement of the original EEG signal and optimizes the differentiation of EEG artifacts through ICA. Additionally, within the framework of the Wavelet Transform (WT) algorithm, the discrete wavelet transform allows for the conversion of signals from the time domain to the wavelet domain. This decomposition process yields wavelet coefficients at different scales and frequencies, enabling signal analysis and equivalence between continuous signals through sampling and reconstruction. Nonetheless, the determination of suitable thresholds remains a pivotal factor limiting the applicability of this method[14]. Another approach, the Hilbert-Huang Transform (HHT), computes the instantaneous frequency of each EEG signal by means of Intrinsic Mode Functions (IMFs) and eliminates signals outside the EEG frequency range. While this method is characterized by its data-independence and wide applicability, it suffers from high computational overhead and inadequate removal of EEG artifacts due to the overlapping nature of EEG and artifact signals in the frequency domain.

### 2 MATERIALS AND METHODS

### 2.1 Participant and stimuli

The participant was instructed to stay relaxed during the restingstate section and to ensure that each section lasted long enough for the eye-opening and closing behavior to be identified later.

### 2.2 Data Description

The EEG eye state dataset from UCI machine learning repository is employed for the experiments. All data were derived from one continuous EEG measurement with the Emotiv EEG neuroheadset (Emotive AEpoc+) The duration of the measurement was 117 seconds. There were 14977 valid data sets recorded in the corpus, with a total of 14 continuous numerical variables recorded from sensors AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4. Additionally, a categorical variable "eye blinking" was detected via a camera during the EEG measurement and added later manually to the file after analyzing the video frames.

### 2.3 Data Processing

2.3.1 *Removing outliers.* When the sample size is too large, shrinkage methods such as ridge regression or lasso regression can be used to avoid overfitting in the regression analysis, especially when there are many predictor variables. Shrinkage methods add a penalty term to the regression equation, which shrinks the regression coefficients towards zero, making the model more stable and less prone to overfitting. However, shrinkage methods do not necessarily make the data smoother. Instead, they provide a balance between bias and variance, resulting in better prediction accuracy.

2.3.2 Correlation coefficient analysis. The data were analyzed using a factor analysis algorithm. The basic logic of factor analysis is to construct a small number of representative factor variables from the original variables. Therefore a relatively strong correlation between the original variables is required. Therefore, a sexual inertia matrix is done on the data after removing the outliers to determine whether there are common features.(Figure 1)



Figure 1: Correlation matrix.

2.3.3 Correlation test. The KMO test (Kaiser-Meyer-Olkin test) measures the degree of common variance among variables, indicating whether they are suitable for factor analysis. The KMO values range from 0 to 1, with higher values indicating better suitability for factor analysis. A KMO value above 0.5 is generally considered acceptable, while values above 0.8 are considered very good.

Bartlett's sphericity test, on the other hand, tests whether the correlation matrix is significantly different from an identity matrix (indicating independence between variables). A significant result (P-value <0.05) indicates that the variables are not independent and are suitable for factor analysis. In this data, both tests indicate that the variables are suitable for factor analysis, with a KMO value of 0.84 (which is very good) and a significant P-value of 0 for Bartlett's sphericity test. The scree plot suggests that 3 factors should be retained, with variables. (Figure 2)



Figure 2: Scree plot.

2.3.4 Factor analysis. The study employed principal component variance maximization orthogonal factor rotation to simplify the structure of a variable set. By transforming the factor loading matrix through rotation, each column element was simplified to either 0 or 1 based on its column. The factor loading matrix represents the correlation between each variable and the factor, with the magnitude of the absolute value indicating the explanatory power of the variable for the factor. Factor scores were then computed for each electrode point at RC1, RC2, and RC3.

The analysis revealed significant positive correlations within the first factor (RC1) for AF3, FC6, F4, F8, and AF4, while F3 and T8 exhibited moderate positive correlations with correlation coefficients of 0.59 and 0.60, respectively. The second factor (RC2) showed strong positive correlations for T7, P7, O1, and O2, with a correlation coefficient of 0.69 with T8. The third factor (RC3) demonstrated a strong positive correlation between F7 and FC5.

2.3.5 Logistic regression. The logistic regression model was utilized to evaluate the influence of three types of EEG data, pertaining to varying electrode positions, on the states of eyes open or closed. The results revealed that both RC1 and RC3 exhibited significant predictive capabilities for determining the eye open and closed statuses within the logistic regression framework. Therefore, within the constraints of limited electrode point EEG data, it is postulated that the EEG signal in the frontal lobe possesses a discernible predictive effect on the dynamics of eye opening and closing states.(Figure 3)

	Estimate	Std.Error	Z value	Pr(> z )
(Intercept)	-0.21473	0.01683	-12.761	<ze-16***< th=""></ze-16***<>
RC1	0.39076	0.01862	20.984	<ze-16***< th=""></ze-16***<>
RC2	0.03795	0.01678	2.262	0.0237*
RC3	-0.27174	0.01764	-15.406	<ze-16***< th=""></ze-16***<>
Signif.codes: 0 '***' 0.001 '**' 0.01 '* 0.05 '.' 0.1" 1				

Figure 3: The result of logistic regression.

### 3 DISCUSSION AND CONCLUSION

The data unequivocally demonstrate that the neural substrates governing blink behavior are likely associated with the frontal lobe. This empirical finding corroborates the hypothesis positing a substantive relationship between blink behavior and dopaminergic pathways. The frontal lobe is widely acknowledged for its pronounced involvement in dopamine-mediated neural processes, with the prefrontal cortex of primates exhibiting a rich presence of D1 receptors. Functioning as the epicenter of executive functions encompassing response inhibition, motivation, attentional allocation, and working memory[20], the frontal lobe's integrity is intricately linked to dopamine depletion, age-related cognitive disorders[13], prognostic evaluation in schizophrenia patients, and neurodegenerative ailments such as Alzheimer's and Parkinson's diseases that manifest associations with working memory performance. Notably, individuals with Parkinson's disease exhibit modulated cognitive operations regulated by dopaminergic manipulation, thus underscoring the frontal executive effects of this neurotransmitter in memory discrimination tasks[22]. Moreover, aberrant spontaneous blink rates have been consistently documented across aforementioned pathologies characterized by perturbed dopamine metabolism. For instance, patients afflicted with schizophrenia evince augmented spontaneous blink occurrences ascribed to excessive dopamine in the mesocortical circuitry. Likewise, attentional deficits and executive function impairments stemming from dopamine depletion are accompanied by notable fluctuations in blink rates during sustained attention and transient fatigue episodes[17]. Furthermore, heightened dopamine activity inferred from elevated eyelid blink rates (EBR) in mild cognitive impairment (MCI) patients serves as a reliable indicator of dysregulated dopaminergic activity in the central nervous system, thereby constituting a prodromal biomarker along the continuum from healthy aging to pathological dementia[15]. Importantly, abnormal blink behaviors are not confined to neurological disorders alone; they have also been observed in patients diagnosed with craniofacial muscle tension disorders, as well as in conditions involving ocular muscle anomalies like Graves-Basedow thyroid ophthalmopathy or relapsing conjunctivitis. (Chatziralli, Kanonidou et al. 2010)Consequently, blink behavior is considered a physiological correlate of dopaminergic and cholinergic pathways, thus warranting its utilization as an ancillary metric in investigating behavioral abnormalities in psychological experiments focusing on executive function, working memory, and reward-related processes. Moreover, blink parameters hold considerable promise as clinical diagnostic and prognostic indicators across a wide array of pathological conditions.

In addition to logistic regression, various machine learning algorithms have been compiled and summarized by researchers for the classification of brainwave data. Linear regression assumes a linear relationship between known data points, distributed around a straight line represented by slope and intercept. The sum of squared residuals is commonly used to describe the magnitude of the distance between the ideal values and the fitted line. Logistic regression, employing the logistic function (i.e., Sigmoid function), maps predictions to the range of 0 to 1. Evaluation of regression fits often involves the adoption of metrics such as cross-entropy, ROC curves, and confusion matrices.While linear regression assumes linearity and normal distribution, these assumptions may not hold true when dealing with EEG and EOG data. As an alternative, logistic regression is employed due to its ability to handle non-linear relationships. However, other machine learning algorithms offer additional advantages for brainwave data analysis[9].

K-nearest neighbors (KNN), as a supervised learning algorithm, requires labeled data for training. It selects K nearest samples to a given data point and determines the class based on the majority class of the neighboring samples. The similarity between two samples is defined by their distance. KNN is a non-parametric and non-linear classifier, particularly suitable for larger training sets. For instance, in EEG signal detection for epileptic seizures, KNN can be applied by utilizing discrete wavelet transforms to decompose the signal and feeding statistical features into the KNN classifier for identifying specific seizure signals[21].K-means, on the other hand, is an unsupervised learning algorithm used for clustering. It aims to partition data into K clusters, with each data point belonging to the cluster with the closest centroid. The algorithm iteratively updates the centroid positions until convergence. KNN exhibits flexibility in training time, interpretability, and suitability for smaller datasets. It has practical applications, such as the detection of driver fatigue through blink behavior analysis[5].

Decision trees are flowchart-like structures, where internal nodes represent features, branches depict decision rules, and leaf nodes represent outcomes or class labels. The tree-building process continues until the entropy reaches zero, indicating that all data points belong to the same class. Random Forest, an ensemble learning method, combines multiple decision trees to make predictions. It randomly selects subsets of training data and features for each tree, and aggregates the predictions through majority voting or averaging. Random Forest is an ensemble learning technique in machine learning that effectively combines multiple models for prediction.

Support Vector Machines (SVM) are another commonly used algorithm for non-linear classification problems. Their objective is to find the optimal hyperplane that maximizes the classification confidence between two classes, aiming to maximize the margin between the classes. In a study comparing random forest, random ferns, and support vector machines for eye state classification tasks, random forest demonstrated superior time consumption[4].

Convolutional Neural Networks (CNN) are highly relevant in various applications. They address the issue of significantly increasing parameter numbers in fully connected neural networks by introducing the concepts of parameter reduction and information sharing. Local information sharing is achieved through convolutional operations, and pooling operations further reduce the output within each layer. Filters in CNN extract different features from the input, while the ReLU activation layer introduces non-linearity, enabling the network to learn more complex patterns. Fully connected layers connect the outputs from the previous layer to neurons and transform them into final outputs.While CNN exhibits excellent performance in eliminating blink artifacts, it demonstrates regional differences, with superior performance however, it is important to note that CNN's performance may vary across different brain regions, with the frontal lobe showing better results compared to the occipital lobe. Additionally, CNN is more suitable for offline operations due to its time-consuming nature, which presents a trade-off between time consumption and accuracy[11].

In the future, further advancements can be made in addressing blink artifacts in EEG signals by exploring the distinct characteristics of spontaneous blinks in different diseases. This can involve developing more targeted feature extraction and classification algorithms that can provide a quantifiable measure of blink behavior for disease diagnosis. Simultaneously, investigating the pathological mechanisms associated with blink behavior can contribute to a deeper understanding of the underlying disease processes. Overall, the selection of an appropriate classification algorithm should be guided by a thorough understanding of the data, the specific requirements of the application, and the desire to achieve accurate and meaningful results.

# REFERENCES

- M Aramideh, BW Ongerboer De Visser, PP Devriese, LJ Bour, and JD Speelman. 1994. Electromyographic features of levator palpebrae superioris and orbicularis oculi muscles in blepharospasm. *Brain* 117, 1 (1994), 27–38.
- [2] Leigh F Bacher. 2010. Factors regulating eye blink rate in young infants. Optometry and vision science: official publication of the American Academy of Optometry 87, 5 (2010), 337.
- [3] Nicolaas I Bohnen and Roger L Albin. 2011. The cholinergic system and Parkinson disease. Behavioural brain research 221, 2 (2011), 564–573.
- [4] Yanchao Dong, Yan Zhang, Jiguang Yue, and Zhencheng Hu. 2016. Comparison of random forest, random ferns and support vector machine for eye state classification. *Multimedia Tools and Applications* 75 (2016), 11763–11783.
- [5] Mariella Dreißig, Mohamed Hedi Baccour, Tim Schäck, and Enkelejda Kasneci. 2020. Driver drowsiness classification based on eye blink and head movement features using the k-NN algorithm. In 2020 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 889–896.
- [6] Thomas Elbert, Werner Lutzenberger, Brigitte Rockstroh, and Niels Birbaumer. 1985. Removal of ocular artifacts from the EEG-a biophysical approach to the EOG. Electroencephalography and clinical neurophysiology 60, 5 (1985), 455-463.
- [7] JW Gnadt, Shao-Ming Lu, B Breznen, MA Basso, Victor M Henriquez, and Craig Evinger. 1997. Influence of the superior colliculus on the primate blink reflex. *Experimental brain research* 116 (1997), 389–398.
- [8] Dirk Hagemann and Ewald Naumann. 2001. The effects of ocular artifacts on (lateralized) broadband power in the EEG. *Clinical Neurophysiology* 112, 2 (2001), 215–231.
- [9] Sven Hoffmann and Michael Falkenstein. 2008. The correction of eye blink artefacts in the EEG: a comparison of two prominent methods. *PloS one* 3, 8 (2008), e3004.
- [10] Bryant J Jongkees and Lorenza S Colzato. 2016. Spontaneous eye blink rate as predictor of dopamine-related cognitive function—A review. *Neuroscience & Biobehavioral Reviews* 71 (2016), 58–82.
- [11] Marcin Jurczak, Marcin Kołodziej, and Andrzej Majkowski. 2022. Implementation of a convolutional neural network for eye blink artifacts removal from the electroencephalography signal. *Frontiers in Neuroscience* 16 (2022), 34.
- [12] Jaime Kaminer, Alice S Powers, Kyle G Horn, Channing Hui, and Craig Evinger. 2011. Characterizing the spontaneous blink generator: an animal model. *Journal of Neuroscience* 31, 31 (2011), 11256–11267.
- [13] Teresa M Karrer, Anika K Josef, Rui Mata, Evan D Morris, and Gregory R Samanez-Larkin. 2017. Reduced dopamine receptors and transporters but not synthesis capacity in normal aging adults: a meta-analysis. *Neurobiology of aging* 57 (2017), 36–46.
- [14] V Krishnaveni, S Jayaraman, L Anitha, and K Ramadoss. 2006. Removal of ocular artifacts from EEG using adaptive thresholding of wavelet coefficients. *Journal* of neural engineering 3, 4 (2006), 338.
- [15] Aristea Ladas, Christos Frantzidis, Panagiotis Bamidis, and Ana B Vivas. 2014. Eye blink rate as a biological marker of mild cognitive impairment. *International Journal of Psychophysiology* 93, 1 (2014), 12–16.
- [16] Jui-Chieh Liao and Wai-Chi Fang. 2013. An ICA-based automatic eye blink artifact eliminator for real-time multi-channel EEG applications. In 2013 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 532-535.
- [17] Antonio Maffei and Alessandro Angrilli. 2018. Spontaneous eye blink rate: An index of dopaminergic component of sustained attention and fatigue. *International Journal of Psychophysiology* 123 (2018), 58–63.
- [18] Yuta Miyamoto, Sachiko Katayama, Naoki Shigematsu, Akinori Nishi, and Takaichi Fukuda. 2018. Striosome-based map of the mouse striatum that is conformable to both cortical afferent topography and uneven distributions of dopamine D1 and D2 receptor-expressing cells. *Brain Structure and Function* 223 (2018), 4275–4291.

Analysis of Eye Blinking EEG Artifacts: Exploring Clinical Significance

ICBBT 2023, May 26-28, 2023, Xi'an, China

- [19] Preshanta Naicker, Shailendra Anoopkumar-Dukie, Gary D Grant, David L Neumann, and Justin J Kavanagh. 2016. Central cholinergic pathway involvement in the regulation of pupil diameter, blink rate and cognitive function. *Neuroscience* 334 (2016), 180–190.
- [20] Donald R Royall, Edward C Lauterbach, Jeffrey L Cummings, Allison Reeve, Teresa A Rummans, Daniel I Kaufer, W Curt LaFrance, Jr, and C Edward Coffey. 2002. Executive control function: a review of its promise and challenges for clinical research. A report from the Committee on Research of the American Neuropsychiatric Association. The Journal of neuropsychiatry and clinical

neurosciences 14, 4 (2002), 377-405.

- [21] A Sharmila and PJIA Geethanjali. 2016. DWT based detection of epileptic seizure from EEG signals using naive Bayes and k-NN classifiers. *Ieee Access* 4 (2016), 7716–7727.
- [22] Nahid Zokaei and Masud Husain. 2019. Working memory in Alzheimer's disease and Parkinson's disease. *Processes of visuospatial attention and working memory* (2019), 325–344.