 doi:10.20944/preprints202203.0200.v1

Article

Unsupervised Blink Detection Using Eye Aspect Ratio Values

Bharana Fernando ‡, Arjun Sridhar ‡, Shawhin Talebi ‡, John Waczak, and David J. Lary ‡

Hanson Center for Space Sciences, University of Texas at Dallas, Richardson TX 75080, USA
* Correspondence: aaf170130@utdallas.edu
‡ These authors contributed equally to this work.

Abstract: The eyes serve as a window into underlying physical and cognitive processes. Although factors such as pupil size have been studied extensively, a less explored yet potentially informative aspect is blinking. Given its novelty, blink detection techniques are far less available compared to eye-tracking and pupil size estimation tools. In this work, we present a new unsupervised machine learning blink detection strategy using existing eye-tracking technology. The method is compared to two existing techniques. All three algorithms make use of eye aspect ratio values for blink detection. Accurate and rapid blink detection complements existing eye-tracking research and may provide a new informative index of physical and mental status.

Keywords: Machine Learning; Eye Tracking; Blink Detection

Citation: Fernando, B.; Sridhar, A.; Talebi, S. Title. *Preprints* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

1. Introduction

It has been said that the eyes are a "window to the soul". Despite its colloquial nature, such a statement has an anatomical basis. The eyes play a key role in the central nervous system. For example, pupils dilate in response to elevated levels of autonomic arousal and mental effort, gaze converges and diverges based on attention [1–4]. Although these phenomena have been studied extensively, there is another aspect of the eyes that perhaps deserves more attention: blinking.

Human adults typically blink 15-20 times a minute. This action is physiologically necessary to keep the eye lubricated, but that is only required about two to four times a minute. On average, the duration of a blink can range from 100-400 ms, but it also depends on individual characteristics, fatigue level, and the time of day [5].

Effective classification of blinks has a wide variety of applications. For instance, it has been used to assess an individual's mental status. This includes the evaluation of fatigue, concentration levels, attention span, and cognitive load [6,7]. Another application of blink detection is to facilitate the removal of blinking artifacts from Electroencephalography (EEG) signals [8].

Although blinking can be controlled directly, it is often involuntary. Thus, autonomically regulated blink rates may be indicative of different cognitive states. When a person relaxes or concentrates on a visual object, their blinking rate decreases, whereas negative emotions and conversations with other people cause it to increase [5].

There are three observed types of blinks: spontaneous, reflex, and voluntary. The first two types of blinks are autonomic responses. Spontaneous blinks occur without external stimuli, while reflexive blinks depend on bright lights, loud noises, etc. Furthermore, the closing phase of the spontaneous blink is longer than the reflex blink, while the opposite is true for the opening phase, indicating that the eye seeks to autonomously protect itself at a moment's notice [9]. It has been found that the spontaneous eye blink rate correlates with dopaminergic activity and striatal dopamine receptor availability [10,11]. Many studies have shown that dopamine plays a role in attention [12], learning [13], goal-directed behavior [14], and time perception [15]. A quantification of dopamine can provide insight into these critical cognitive functions. Considering that spontaneous blinks are correlated with dopaminergic activity, they may serve as a convenient proxy.

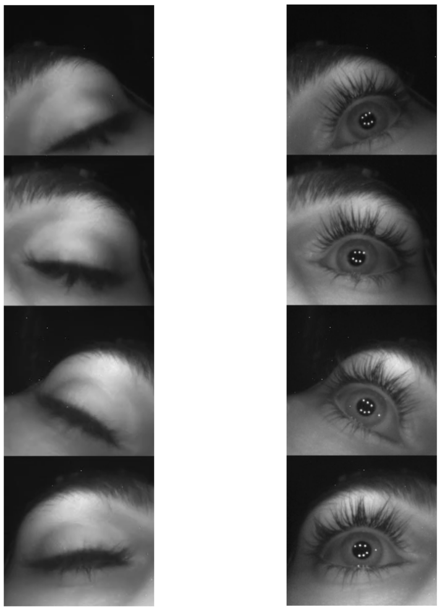
In this paper, we compare two existing blink classifiers to a novel blink detection technique that employs an unsupervised machine learning method. We find that our

 doi:10.20944/preprints202203.0200.v1

unsupervised method outperforms an existing supervised machine learning method, thus eliminating the need the need for a labeled training dataset.

2. Materials and Methods

Eye data were collected using the Tobii Pro Glasses 2 eye tracking system. In addition to gaze and pupil tracking, the glasses record infrared images of the left and right eyes from two angles, at a rate of 50 Hz. These images are stacked into a single frame and saved as a video. We will refer to this video as an eye-stream video. Example frames from an eye-stream video can be seen in Figure 1.



1.png

Figure 1. Example video frames from eye-stream video showing eyes-closed and eyes-open frames.

Although the Tobii Pro Glasses 2 collect over 30 biometric variables at 100Hz, it lacks a built-in blink detection feature. This is a major setback for a wide range of biometrics research toward attention, cognitive load, and dopamine, in which blinking may play a central role. Given the information rich eye-stream video, however, we demonstrate that reliable blink detection can be achieved.

2.1. Eye Aspect Ratio (EAR)

The eye aspect ratio (EAR) is the key concept on which all presented algorithms are built. An EAR value is a numerical representation of how open or closed the eye is [16]. The defining equation is a ratio between the height and width of the eye as calculated by:

$$EAR = \frac{||p2 - p6|| + ||p3 - p5||}{||p1 - p4||}, \tag{1}$$

Where, $p1$ through $p6$ are points contouring the eye, as seen on Figure 2. As the eye closes, the numerator (or height) will approach zero. Thus, a low EAR value may represent a blink occurring. The six points in the equation are facial landmarks automatically detected using the method from [17] as shown in the top panel of Figure 2.

Using the eye landmarks and equation 1, we can calculate the EAR value for each eye, and subsequently an average EAR value is considered, as eye blinking is synchronous [16]. However, the facial landmark detection method described above requires the entire face to be captured, but the Tobii Pro Glasses 2 only captures eye images as shown in Figure 1. To overcome this, the eyes were cropped from the eye-stream video and then superimposed onto a face image. Right panel of Figure 2 shows the result. Using the structure in Figure 2

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2022 doi:10.20944/preprints202203.0200.v1

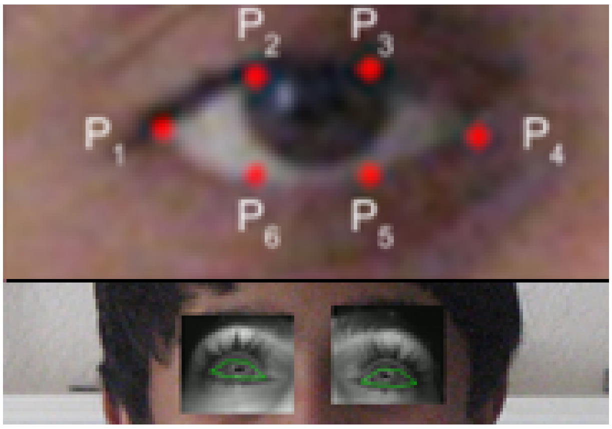
combined.png 

Figure 2. The top panel shows labeled eye landmarks that equation refeq1 utilizes. The bottom panel shows a cropped image of the eyes from the Tobii Pro glasses superimposed on a face image so that the eye landmarks can be detected.

and the EAR value equation, the algorithms discussed in the following sections attempt to detect blinks.

2.2. Baseline Method

The baseline method is adapted from [16], using the EAR values obtained from the previous section. This algorithm detects blinks by altering two parameters: a threshold EAR value and a number of consecutive frames below said threshold. If the average EAR value falls below the threshold for some number of consecutive frames, we classify this as a blink [16].

When the eyes are open, the average EAR value will remain relatively constant, but as they begin to close there is a sharp drop. Figure 3 shows the plot of the EAR values for each frame from the eye-stream video with a threshold of 0.2.

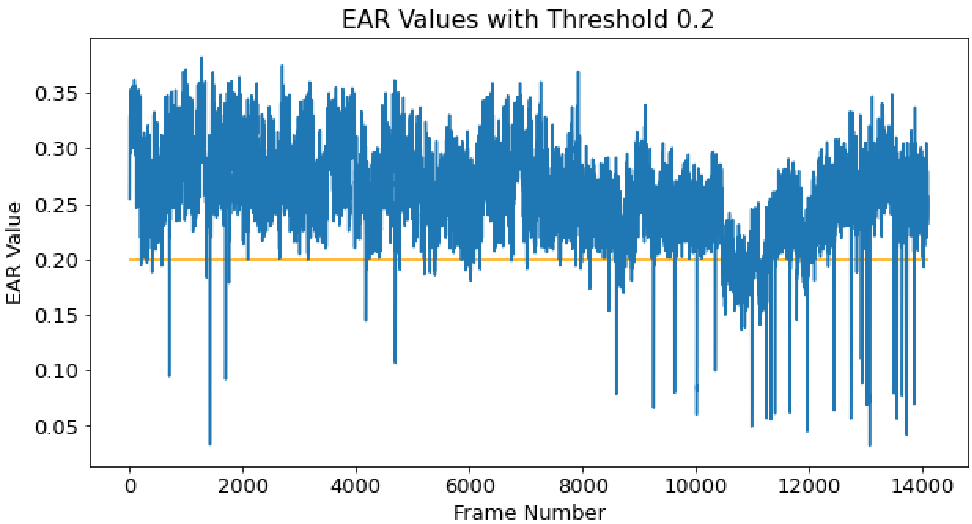


Figure 3. The plot of EAR Values for each frame using the structure in Figure 2 with a threshold of 0.2. A blink is detected if the EAR value is below 0.2 for 3 consecutive frames.

From Figure 3, an EAR value of 0.2 appears to be a reasonable threshold, and 3 consecutive frames were chosen to detect a blink. Although thresholding is a natural attempt to classify blinks, it is not robust to noise. As can be seen from Figure 3, EAR values are very noisy. Additionally, superimposing the eyes from the eye-stream video further amplifies noise when it comes to detecting the landmarks.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2022 doi:10.20944/preprints202203.0200.v1

This leads to several false positives in blink detection. EAR values may fall below the threshold for a consecutive number of frames even when the person is not blinking. Conversely, lowering the threshold produces true negatives, or missing out on blinks. Therefore, this method requires sophistication in terms of choosing a robust threshold and consecutive number of frames. The remaining two methods attempt to overcome this issue by using machine learning.

2.3. *Supervised Learning Method (Support Vector Machines)*

In this section, we present method offered in [16] which uses a supervised machine learning method, the support vector machine (SVM). It is a way of improving blink classification as it helps combat against the noise that is generated from the baseline method as described in the previous section.

The idea behind SVM is as follows: given a set of data points that are in different classes and represented in an n-dimensional space, the model will find the best line that separates the points into their respective classes. Thus, SVM is a good choice for dealing with noisy and high dimensional data, as it is not affected by local minima (e.g. if a person is squinting, the baseline method may categorize this as a blink) [16]. For more details on SVM we refer the reader to [18].

To implement this method, it is proposed that the SVM model use a 13 dimension feature vector, where each vector consists of the average EAR value for the current frame, the previous 6 frames, and the next 6 frames (as shown in Figure 4) [16].

5.png

N - 6	N - 5	N - 4	N - 3	N - 2	N - 1	N	N + 1	N + 2	N + 3	N + 4	N + 5	N + 6

Figure 4. The 13 dimension average EAR value window shown for frame N.

This 13 frame feature window allows for a larger temporal period, where noise can be accounted for, which results in significantly improved blink classification.

For our model, we used a dataset consisting of 9,042 static eye-stream images. The average EAR value was computed for each image, and then labeled as open or closed using a 0.1 threshold. Since these are static images, the number of consecutive frames is not needed. The structure of the right panel of Figure 2 was followed when obtaining the average EAR values, and Figure 4 was used to obtain the 13 dimension feature vector for each average EAR value. Then, the model was trained on these feature vectors, where it learned to classify the images as open or closed. To get the optimal hyperparameters for the model, we used the GridSearch Method with a 5-fold cross validation [19].

To evaluate the model, each frame from the eye-stream video contains the average EAR value feature vector with 13 values (using Figure 2 and Figure 4), and the model predicts whether the frame is open or closed using the feature vector as input. Thus, a blink will have occurred if there are successive frames with closed classifications.

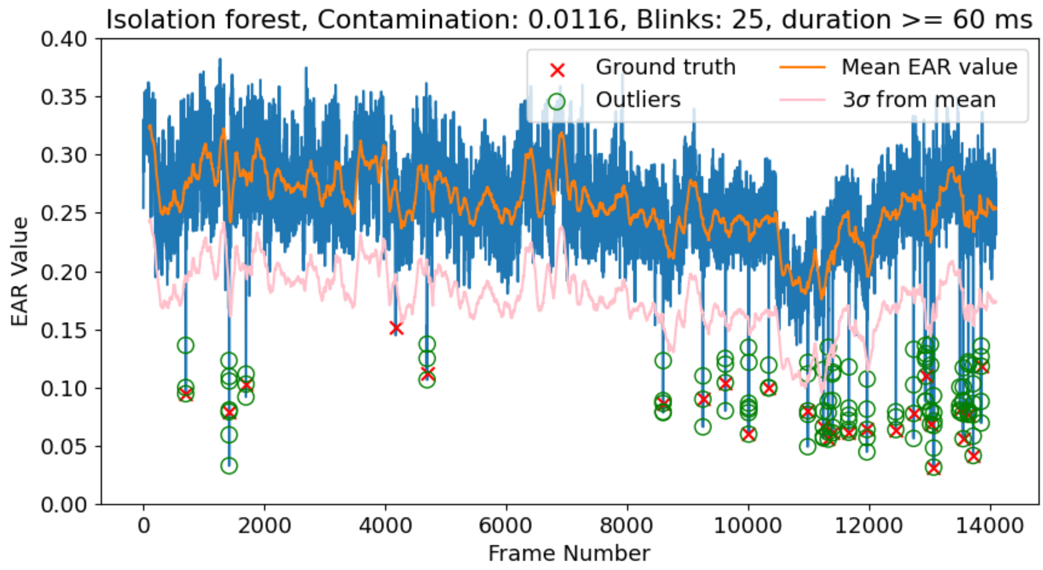
2.4. *Unsupervised Learning Method (Isolation Forest)*

The noisy nature of the EAR values makes it difficult to establish a threshold value to differentiate what is and is not a blink. Indeed, the EAR values in turn may be sensitive to the individual, motivating a departure from a static threshold. An unsupervised method is also presented as a means to establish a threshold for when a blink occurs.

This approach is based on interpreting blinks as an anomalous occurrence, since eyes are typically open while awake. This interpretation encourages the application of outlier detection algorithms to annotate blinks.

The particular technique chosen is the Isolation Forest algorithm, which is based on an ensemble of decision trees, where anomalies are said to have a *shorter path length* than normal points. The path length is determined by selecting a random data record and isolating into a partition. If the data point is “close" to other data points, then more partitioning, or a longer path length may be required. Conversely, isolated data points

 doi:10.20944/preprints202203.0200.v1



6.png

Figure 5. Isolation forest applied onto EAR values to determine 'outliers' (i.e. blinks) . *Contamination* is a pre-determined proportion of outliers. A blink is classified as such if at least three consecutive frames (time resolution of data is 20 ms) were determined to be outliers.

require fewer partitions, thus a shorter path length [20]. Furthermore, this method is well suited for multi-modal datasets and has low computational overhead.

Although the method does not require a labelled training dataset, it contains a parameter which the algorithm is quite sensitive to: contamination, i.e. the proportion of outliers in the dataset. To overcome a similar issue as with a static threshold for EAR values, contamination has the following automatic estimation. A simple moving average (SMA) of window size of a 100 frames is established. A first order contamination ratio is calculated as follows: the number of data points three standard deviations (3σ) away from the SMA mean to the total number of frames. This approach will enable the contamination value to adapt to the data as opposed to a static threshold. Results can be seen in Figure 6.

To avoid misclassifying sudden troughs in data as blinks, we require also that at least three consecutive frames are classified as outliers. The average duration of a blink is 100-400 ms, and given that the frame rate of the eye-stream is 50 fps, this methodology will be sensitive to eye movement longer than 60 ms.

3. Results

In this section, we evaluate the performance of the given three methods to blink detection. Codes for implementing the techniques were written in Python and are available at the GitHub repository [21].

To assess classifier performances, we applied the three methods to the same eye-stream video. The video contains a grayscale recording of a participant's eyes at two different angles for each eye, while naturally scrolling through their personal Twitter feed.

Blink ground truths are obtained manually from direct observation of each from on the eye-stream video. A blink was defined as a singular frame, i.e. the frame at which the eyes are *completely* closed. This frame was typically preceded by three to four frames in which the eyes were rapidly closing, and proceeded by a longer duration of opening the eyes. There were also a few instances where the eyes were closing but did not completely close, which were not labeled as blinks.

The classification methods are evaluated via True Negative (TN), False Negative (FN), False Positive (FP), and True Positive (TP) counts based on pairwise comparisons with the ground truth. Furthermore, the occurrence of a blink within a $\pm$ 6 frame is accepted. For

 doi:10.20944/preprints202203.0200.v1

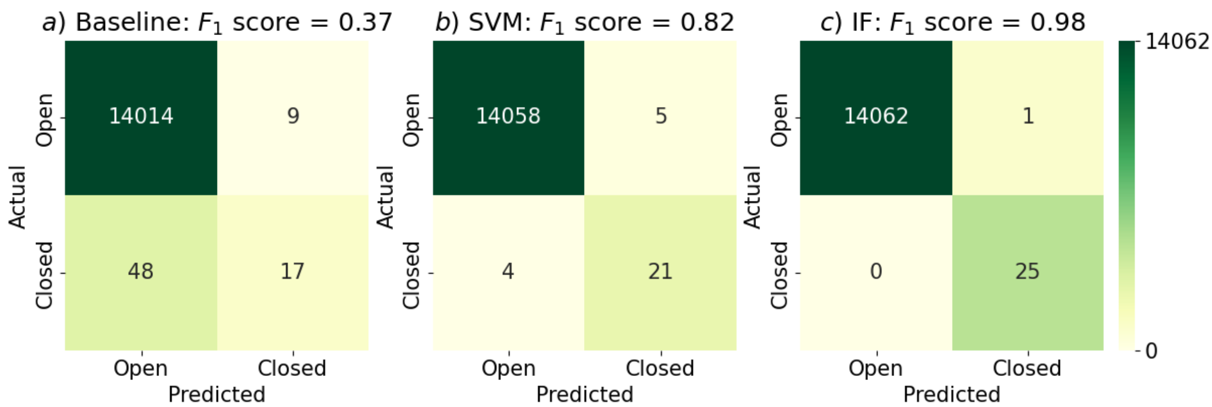


Figure 6. Confusion matrices of the results: **a)** *Baseline* method with a static EAR value threshold, **b)** Supervised prediction with *SVM*, **c)** Unsupervised outlier detection with *IF*. Prediction was accepted as a blink if within $\pm$ 6 frames from a ground truth frame. 26 blinks were observed manually.

example, if a ground truth blink is labeled at frame 700, any true blink flag between 693 - 707 will be recorded as a TP.

An F_1 score, the harmonic mean of precision and sensitivity, is also calculated for each method. Defined as:

$$F_1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)}, \tag{2}$$

It is a measure of relative performance, and is not skewed by the large number of TNs.

As evidenced from the baseline method, a static threshold on EAR values is insufficient to predict blinks accurately. The SVM method performs markedly better, with an F_1 score is 0.82. It should also be noted that due to the rolling nature of the SVM method, it makes 12 less predictions ($\pm$ 6) than other methods. IF method has performed best, with only one misclassification.

4. Discussion

EAR values provide an excellent pathway to detect blinks. However, they can be noisy and require more processing to accurately identify when a blink occurs. This also requires a consistent definition of what a blink is. For the purposes of this study, a blink was defined to be a singular instance of time, when the eyes were *completely* closed. Initially, this required a tedious manual labeling of the frames but through this study, we hope to automate the process of blink detection to a high degree of accuracy.

A static threshold on noisy EAR values (as demonstrated on the Baseline method) misclassified numerous non-blink frames as blinks. SVM method fared much better, but IF performed at the highest degree of accuracy and precision. Furthermore, the IF method does not require a labeled, ground-truth dataset.

This study encourages further analysis into types and phases of blinks and whether they can be reliably classified based on an observed blink duration. As striatal dopamine levels are related spontaneous blink rate, one may allocate a cognitive activation level based on blink frequency. Additionally, accurate blink detection will facilitate removal of blinking artifacts present in EEG data.

As a final note, although these algorithms may apply to other eye-tracking systems, only data from the Tobii Pro Glasses 2 is evaluated here (i.e. the eye-stream video). Therefore, the efficacy of these algorithms with other data requires further investigation. It is our hope that this work can be used to further our understanding of blinking and its association with cognition.

5. Conclusions

Blinks are a powerful, easily accessible, non-invasive way to identify information such as dopaminergic activity and stress indicators, thus it is important to establish an automated detection paradigm. In this paper, we proposed an unsupervised learning method built on top of an already existing technology to make detection more robust. The *Isolation Forest* method is an intuitive, lightweight approach that a) has shown to be more precise and accurate than other methods and b) is unsupervised, thus eliminating the need for a labeled dataset.

Author Contributions: Conceptualization, B.F., A.S., and S.T.; methodology, B.F., A.S., and S.T.; software, B.F. and A.S.; validation, B.F., A.S., and S.T.; formal analysis, B.F., A.S., and S.T.; investigation, B.F., A.S., and S.T.; resources, D.J.L.; data curation,B.F., A.S., and S.T; writing—original draft preparation, B.F., A.S., and S.T; writing—review and editing, B.F., A.S., S.T, J.W., and D.J.L.; visualization, B.F., A.S., and S.T.; supervision, D.J.L.; project administration, D. J. L.; funding acquisition, D. J. L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the following grants: The US Army (Dense Urban Environment Dosimetry for Actionable Information and Recording Exposure, U.S. Army Medical Research Acquisition Activity, BAA CDMRP Grant Log #BA170483). EPA 16th Annual P3 Awards Grant Number 83996501, entitled Machine Learning Calibrated Low-Cost Sensing. The Texas National Security Network Excellence Fund award for Environmental Sensing Security Sentinels. SOFWERX award for Machine Learning for Robotic Teams. Support from the University of Texas at Dallas Office of Sponsored Programs, Dean of Natural Sciences and Mathematics, and Chair of the Physics Department are gratefully acknowledged. The authors acknowledge the OIT-Cyberinfrastructure Research Computing group at the University of Texas at Dallas and the TRECIS CC* Cyberteam (NSF 2019135) for providing HPC resources that contributed to this research (https://utdallas.edu/oit/departments/circ/, accessed 02/22/2022.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data used https://zenodo.org/record/5874701#.YhPxrJPMJb8

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EEG	Electroencephalography
EAR	Eye Aspect Ratio
SVM	Support Vector Machine
IF	Isolation Forest

References

1. van der Wel, P.; van Steenbergen, H. Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin and Review* **2018**, *25*, 2005–2015. doi:10.3758/s13423-018-1432-y.
2. Bradley, M.M.; Miccoli, L.; Escrig, M.A.; Lang, P.J. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* **2008**, *45*, 602–607. doi:10.1111/j.1469-8986.2008.00654.x.
3. Kahnemann, D.; Beatty, J. Pupillary responses in a pitch-discrimination task. *Perception Psychophysics* **1967**, *2*, 101–105. doi:10.3758/BF03210302.
4. Beatty, J.; Kahneman, D. Pupillary changes in two memory tasks. *Psychonomic Science* **1966**, *5*, 371–372. doi:10.3758/BF03328444.
5. Soukupová, T.; Cech, J. Eye-Blink Detection Using Facial Landmarks. *Proceedings of the 21st Computer Vision Winter Workshop* **2016**, pp. 22–29.
6. Sakai, T.; Tamaki, H.; Ota, Y.; Egusa, R.; Inagaki, S.; Kusunoki, F.; Sugimoto, M.; Mizoguch, H. Eda-based estimation of visual attention by observation of eye blink frequency. *International Journal on Smart Sensing and Intelligent Systems* **2017**, *10*, 296–307. doi:10.21307/ijssis-2017-212.

7. Chen, S.; Epps, J. Using task-induced pupil diameter and blink rate to infer cognitive load. *Human-Computer Interaction* **2014**, *29*, 390–413. doi:10.1080/07370024.2014.892428.
8. Agarwal, M.; Sivakumar, R. Blink: A Fully Automated Unsupervised Algorithm for Eye-Blink Detection in EEG Signals. *2019 57th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2019* **2019**, pp. 1113–1121. doi:10.1109/ALLERTON.2019.8919795.
9. Espinosa, J.; Pérez, J.; Mas, D. Comparative analysis of spontaneous blinking and the corneal reflex. *Royal Society Open Science* **2020**, *7*. doi:10.1098/rsos.201016rsos201016.
10. Jongkees, B.J.; Colzato, L.S. Spontaneous eye blink rate as predictor of dopamine-related cognitive function—A review. *Neuroscience and Biobehavioral Reviews* **2016**, *71*, 58–82. doi:10.1016/j.neubiorev.2016.08.020.
11. Groman, S.M.; James, A.S.; Seu, E.; Tran, S.; Clark, T.A.; Harpster, S.N.; Crawford, M.; Burtner, J.L.; Feiler, K.; Roth, R.H.; Elsworth, J.D.; London, E.D.; Jentsch, J.D. In the blink of an eye: Relating positive-feedback sensitivity to striatal dopamine d2-like receptors through blink rate. *Journal of Neuroscience* **2014**, *34*, 14443–14454. doi:10.1523/JNEUROSCI.3037-14.2014.
12. Nieoullon, A. Dopamine and the regulation of cognition and attention. *Progress in Neurobiology* **2002**, *67*, 53–83. doi:10.1016/S0301-0082(02)00011-4.
13. Wise, R.A. Dopamine, learning and motivation. *Nature Reviews Neuroscience* **2004**, *5*, 483–494. doi:10.1038/nrn1406.
14. Goto, Y.; Grace, A.A. Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nature Neuroscience* **2005**, *8*, 805–812. doi:10.1038/nn1471.
15. Terhune, D.B.; Sullivan, J.G.; Simola, J.M. Time dilates after spontaneous blinking. *Current Biology* **2016**, *26*, R459–R460. doi:10.1016/j.cub.2016.04.010.
16. Cech, J.; Soukupova, T. Real-Time Eye Blink Detection using Facial Landmarks. *Center for Machine Perception, Department of Cybernetics Faculty of Electrical Engineering, Czech Technical University in Prague* **2016**, pp. 1 – 8, [arXiv:1011.1669v3].
17. Kazemi, V.; Sullivan, J. One millisecond face alignment with an ensemble of regression trees. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2014**, pp. 1867–1874. doi:10.1109/CVPR.2014.241.
18. Noble, W.S. What is a support vector machine? *Nature Biotechnology* **2006**, *24*, 1565–1567. doi:10.1038/nbt1206-1565.
19. Syarif, I.; Prugel-Bennett, A.; Wills, G. SVM Parameter Optimization using Grid Search and Genetic Algorithm to Improve Classification Performance. *TELKOMNIKA (Telecommunication Computing Electronics and Control)* **2016**, *14*, 1502. doi:10.12928/telkomnika.v14i4.3956.
20. Liu, F.T.; Ting, K.M.; Zhou, Z.H. Isolation Forest **2008**. pp. 413–422. doi:10.1109/ICDM.2008.17.
21. Arjun Sridhar.; Shawhin Talebi.; Ashen Fernando. mi3nts/tobiiBlinkDetection: Blink detection algorithms for Tobii Pro Glasses 2 data.