Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2023) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

27th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2023)

Detection of Alcohol Inebriation from Eye Movements

Silvia Makowski[a,*], Annika Bätz[a], Paul Prasse[a], Lena A. Jäger[a,b], Tobias Scheffer[a]

[a]*Department of Computer Science, University of Potsdam, Potsdam, Germany*
[b]*Department of Computational Linguistics, University of Zurich, Zurich, Switzerland*

Abstract

Today, the most convenient way of estimating an individual's blood-alcohol concentration requires a breathalyzer device and intense user cooperation, which severely limits the scope of potential applications. We develop and study a machine-learning model that detects alcohol inebriation based on a person's eye gaze and eye closure. We investigate the relative contribution of individual features derived from eye gaze and eye closure to the model. In order to train and experimentally evaluate the model, we collect—and share—a new data set with participants in baseline and alcohol-intoxicated states. We find that the model can in fact detect the consumption of a moderate amount of alcohol; the accuracy grows significantly with increasing blood alcohol concentration. The most relevant features turn out to relate to the velocity and acceleration profiles during fixations and saccades. From our proof-of-concept study, we can conclude that contactless inebriation detection based on eye gaze is in fact possible, albeit data need to be collected on an industrial scale to reach practical applicability. Potential applications of contactless inebriation detection include the detection of impaired drivers or operators of other hazardous machinery as well as health-monitoring applications.

Keywords: Eye tracking; alcohol consumption; driver monitoring; machine learning

1. Introduction

Optical detection of alcohol inebriation holds the potential to improve the operational safety of vehicles and other hazardous machinery. Although the exact concentration of alcohol and other intoxicating substances can only be determined by blood tests, several biomarkers are known that are indicative of recent (e.g., ethyl glucuronide, EtG) and chronic alcohol consumption [24]. Breath tests have been shown to reliably estimate blood alcohol concentration (BAC) [1]. However, breathalyzers are expensive, include single-use tubes, and require user cooperation. Therefore, for wide-spread use, inebriation-detection systems would have to be limited to working with signals that can be extracted contactless and without explicit user cooperation.

* Corresponding author.
E-mail address: silvia.makowski@uni-potsdam.de

Driven by new regulations such as the EU General Safety Regulation 2019/2144 and safety rating programs [10], driver cameras are currently entering the automotive market. Driver cameras extract eye-closure features and the driver's eye gaze from the camera images in order to detect inattentive driving, drowsiness, micro-sleep, and sleep [12]. The European new car assessment program (EuroNCAP) furthermore envisages to award car makers a higher safety rating for introducing driver-monitoring systems that can detect driver impairment by alcohol or drugs [9].

In addition to enabling safety applications, the detection of inebriation could facilitate enhanced health-monitoring functionalities on mobile devices. A vast range of health-related apps such as Apple Health, Samsung Health, Fitbit, and Garmin Health allows users to review, share, and receive digital gratification for positive patterns of behavior. Smartphones are equipped with user-facing cameras, and in many cases depth sensors, that enable eye tracking.

Acute consumption of alcohol impairs brain regions that are involved in visual processing and oculomotor control: it reduces cerebellar activity [25], alters the functioning in cranial nerves [23] as well as motor and visual brain areas [5]. This can lead to an impairment of saccadic latency, velocity, and accuracy [15, 27]. On the cognitive level, alcohol intoxication can result in a narrower perceptive scope and impaired decision making [31, 19]. Maurage et al. [20] provide a comprehensive review of psychological research on the effect of alcohol on eye movements.

While there has been a considerable amount of studies investigating the effects of alcohol intoxication on eye movements, there has, to the best of our knowledge, not yet been any prior work on using eye movements as a predictor for alcohol intoxication. Indicators of alcohol intoxication can be detected in thermal images [16]; but it should be noted that thermal cameras are different from near-infrared cameras that are used for driver monitoring. A study that aims at detecting intoxication in facial images [35] is based on a data set in which participants have more cheerful facial expressions in their intoxicated state than in their sober state. It would be interesting to explicitly account for facial expression as a confounding factor. It has been found that alcohol reduces the entropy of gaze angles [30] during driving. Therefore, we include the gaze entropy as a baseline feature in our experiments.

In this paper, we present the first study to show that a machine-learning system can learn to detect acute consumption of moderate amounts of alcohol optically, based on features derived from fixational and saccadic eye movements as well as eye closure. We engineer a vast amount of features, explore the most relevant of them, measure their contribution to the overall accuracy, and discuss their relationship to known effects to alcohol consumption. We find that the inter-individual variability of the model's performance can be partially attributed to the BAC, and find some evidence that habitual alcohol consumption may offset some of the effects that give away alcohol inebriation. In order to promote research surrounding the topic of gaze-based inebriation detection, we also contribute a collection of gaze data of inebriated and sober participants to the research community.

2. Materials and Methods

This section formalizes the problem setting, then describes our data collection and quantitatively analyzes the data, introduces the intoxication-detection model, and elaborates the experimental setting.

2.1. Problem Setting

We study the problem of detecting whether an individual is inebriated, given an eye-gaze trace and eye-closure information. The input to the system consists of a time sequence of the following signals, over the observation period: (1) yaw and pitch angular velocities of the left and right eye, measured in degrees of visual angle per time step, (2) an eye-closure signal on a scale of zero (fully open, aperture of at least 12 mm) to one (fully closed), and (3) a discrete eye-state variable that indicates whether the pupil is covered by the eye-lid.

We measure false-positive and true-positive detection rates. Each time step of each evaluation sequence constitutes an instance; time steps progress with a stride of five seconds. A positive instance is a time step in which the participant's blood alcohol concentration (BAC) is 0.03% or higher. A negative instance is a time step in which the BAC is zero. We disregard time steps in which the BAC is greater than zero but less than 0.03% in the evaluation, since the intoxicating effects in this range are generally mild but inconsistent across individuals. A positive model output is triggered when the model output exceeds a decision threshold τ that allows us to adjust the trade-off between false-positive and false-negative rate. The attainable pairs of true and false positive rates are characterized by a ROC curve and aggregated into the area under the ROC curve *(AUC)*.

We conduct most experiments in an *open-population* setting, in which the evaluation is carried out for previously unseen individuals who were not part of the training data. This reflects the use case of an alcohol test in a traffic stop. In addition, we also explore the detection performance in a *closed-population* use case. In this setting, the evaluation is performed on yet unseen gaze sequences from the training users. This use case draws inspiration from a personal device or a driver-monitoring system that has to work for a specific user, or small group of users.

2.2. Data Collection

We collect the *Potsdam Binge / PVT* data set. The data set consists of binocular eye movements and eye-closure features of 44 participants, aged 18 to 47 with mean age 24 and from which 23 are female and 21 male. Participants have normal or corrected-to-normal vision (9 participants wearing glasses and 3 wearing contact lenses). All participants have given their written informed consent ant the study has been approved by the ethics committee of the University of Potsdam.

Each participant is recorded in three experimental sessions (one *alcohol* and two *baseline* sessions) with a time lag of at least one week between any two sessions. The order of the experimental sessions is varied across participants. Participants are asked to always appear well rested and sober to a recording session. (1) In the *alcohol* session, participants consume an alcoholic beverage with 26 ml of alcohol at the start of the session; this dose results in a blood alcohol concentration of below 0.05% for most individuals. (2) The procedure of the two *baseline sessions* is identical to the *alcohol* session except that no alcohol is consumed.

During each of the sessions, participants execute three Psychomotor Vigilance Tasks (PVT) of 10 minutes [6] (PC-based reimplementation), interrupted by two time intervals in which they perform tasks for other experiments. In the PVT task, a small red dot appears in the center of a black screen after random time intervals, drawn from within the range of 2 to 10 seconds. Participants are asked to press a button as soon as they recognize the dot. When the button is pressed, the color of the dot changes to green and disappears after one second. We chose this task because it requires sustained attention but no specific skills. In the alcohol session, we estimate the blood alcohol concentration before and after each 10-minute PVT recording block with a Dräger Alcotest 5820 breathalyzer, resulting in six test results per alcohol session. The BAC is then linearly interpolated across each 10-minute interval.

Some participants have volunteered to complete an alcohol use disorder identification test (AUDIT) [2] that gives us a quantitative measure of their level of habitual drinking. The test consists of ten questions covering the frequency and quantity of drinking, and the impact of drinking on the participant's health, professional, and social life.

2.2.1. Technical Setup

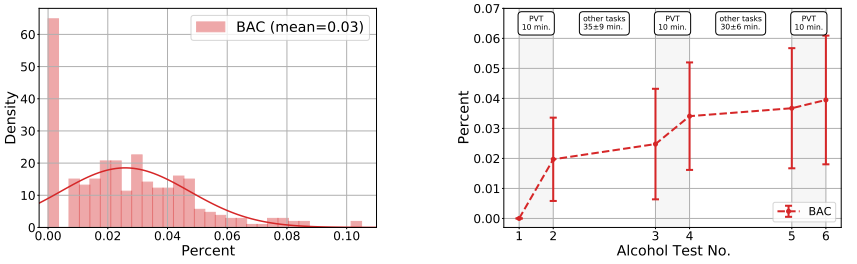
Participants' eye gaze is recorded with a desktop-mounted Eyelink Portable Duo Eyetracker (SR Research) at a sampling frequency of 2000 Hz and a vendor-reported spatial precision of 0.01°. Additionally, participants' faces are recorded with a video-camera, with a sampling frequency of 30 fps and an image resolution of 344×408 px. The camera records in the near-infrared spectrum and is sensitive to the infrared illumination of the eye tracker (880 nm). From these facial recordings we extract the continuous eye-closure signal that ranges between zero for an eye aperture of 12 mm or more and one for fully closed eyes, as well as a discrete variable that indicates whether the pupil is covered by the eyelid using a commercial driver-monitoring software (Asaphus Vision). During the experiment, participants sit at a height-adjustable table in front of a computer monitor (38×30 cm, 1280×1024 px) with their heads stabilized by a chin- and forehead rest. The eye-to-camera distance is 45 cm and the eye-to-screen distance is 63 cm.

2.2.2. Quantitative Data Analysis

Figure 1a shows the histogram of blood alcohol concentrations in the alcohol session across participants and measurements, while Figure 1b shows the development of the BAC over the time of recording, averaged over participants. The blood alcohol concentration starts at zero and grows to a mean value of 0.04%.

2.3. Intoxication Detection Models

Human eye movements alternate between fixations of around 250 ms during which input is obtained and saccades of around 50 ms which are fast relocation movements during which visual input is supressed. During fixations, involuntary micromovements always occur which, among other functions, prevent visual fading of the fixated image.



(a) Distribution of BAC across participants and measurements in the alcohol session.

(b) Development of mean BAC across participants in the alcohol session over time. Error bars show the standard deviation.

Fig. 1: Results of the breath alcohol test.

We segment the raw gaze data into saccades and fixations [28] and engineer a comprehensive set of features that describe motion, velocity, and acceleration profiles of saccadic and fixational eye movements as well as eye-lid movements during blinks. A complete list of eye-closure, saccadic and fixation base features is shown in Table 1.

For each of the base features in Tables 1, we calculate counts, mean, median, standard deviation, skewness, and kurtosis across all blinks, saccades, and fixations, respectively, in the input window. In addition, we determine counts of blinks, saccades, and fixations, counts of time steps with tracker loss and time steps that cannot be segmented into fixations and saccades, and the gaze entropy [30]. We also implement features that have been used for drowsiness detection in prior work; these features include the blink duration, the percentage of time for which the eyes are closed (PERCLOS), and the ratio of the maximum amplitude to maximum velocity of eyelid movement for the closing and reopening phases of blinks [34]. In total, we implement and extract 785 features.

We employ random forest [3] classifiers with varying feature subsets using the scikit-learn library [22]. Test instances which are above 0.00% but below 0.03% of BAC are disregarded (see Section 2.1). For the purpose of training models, we consider two settings. In the first setting, we treat input windows with a BAC of at least 0.01% as positive and all other input windows as negative instances. In the second setting, we discard training instances with a BAC of greater 0 but less than 0.03%, consistently with the handling of test instances.

2.4. Experimental Setting

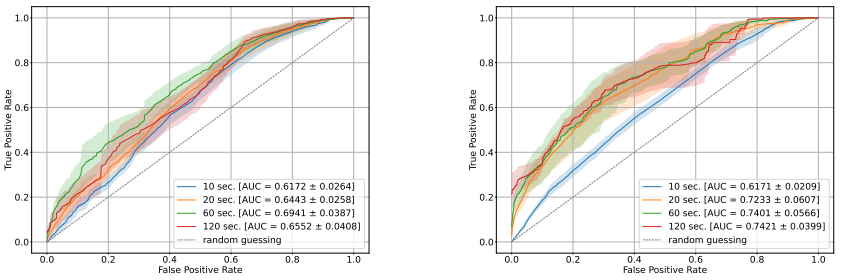
In the open-population setting, we use an outer loop of 5-fold cross validation that is stratified according to persons to ensure that no person appears in both the training and evaluation data at any time. In the closed-population setting we use a 5-fold cross validation that is not stratified according to persons. When studying the inter-individual variation in inebriation-detection accuracy, we conduct a leave-one-person-out cross validation with a nested 5-fold cross validation on the training data for hyper-parameter tuning.

In each cross-validation iteration, we optimize the hyper-parameters with a grid search; the search space is spanned by the number of trees (10, 20, 50, 100, 200, 500), the maximum depth (2, 4, 6, 8 and “none”), the maximum number of features considered for finding the best split (the square root or logarithm of the number of features), and the splitting criterion (Gini impurity, information gain). For hyper-parameters tuning, we conduct nested cross-validation with an inner 5-fold cross-validation on the training portion of the outer cross-validation.

In the open-population setting, the resulting hyper-parameters are 200 trees, depth 2, log2 as the maximal number of features, information gain as splitting criterion; in the closed-population setting the resulting values are 20 trees, depth 4, log2 as the maximal number of features, and the information gain criterion. After completing the hyper-parameter optimization, we keep these values fixed and perform recursive feature elimination in order to identify informative features. Starting with all available features and iterating between training and elimination of 10% of features with lowest weight, features can be ranked based on their level of elimination. Repeating this procedure across the outer 5-fold split described above, a mean rank is assigned to each feature. Models are subsequently trained on the highest ranked features. For the top 53 features, we observe the highest AUC on the evaluation data.

Table 1: Features: absolute values for counts, and mean, median, standard deviation, skewness, and kurtosis of other base features.

	Feature	Source
Eye-closure features	Time steps with eye state “open” (count)	Asaphus Vision
	Time steps with eye state “closed” (count)	Asaphus Vision
	Time steps with eye state “partially open” (count)	Asaphus Vision
	Time steps with eye state “not visible” (count)	Asaphus Vision
	Time steps with eye state “downcast” (closed or looking downward, count)	Asaphus Vision
	Time steps with eye state “not available” (count)	Asaphus Vision
	Number of blinks (count)	-
	Blink duration from start to maximum reopening velocity	[29]
	Blink duration normalized by mean duration	[29]
	Blink duration from maximum closing to maximum opening velocity	[34]
	Blink duration from onset of closing to full reopening	[34]
	Time interval between two adjacent blinks	[29]
	Lid-closure amplitude	[29]
	Lid-closure amplitude normalized by mean amplitude	[29]
	Maximum closure velocity during blink	[29]
	Maximum closure velocity during blink normalized by expected velocity	[29]
	Mean closure velocity during blink normalized by expected velocity	[29]
	Delay between full closure and onset of reopening	[29]
	Percentage of time with eyes closed	[34]
	Ratio of the maximum amplitude to maximum velocity of eyelid movement for the reopening phase	[34]
	Ratio of the maximum amplitude to maximum velocity of eyelid movement for the closing phase	[34]
	Percentage of time the eyes are fully closed for more than 10 ms	[34]
Saccadic features	Number of saccades (count)	-
	Saccade duration	[29]
	Saccade duration normalized by mean duration	[29]
	Time interval between two adjacent saccades	[29]
	Saccade amplitude	[29]
	Saccade amplitude normalized by mean amplitude	[29]
	Max velocity during saccade	[29]
	Max velocity during saccade normalized by expected velocity	[29]
	Mean velocity during saccade normalized by expected velocity	[29]
	25 features of the overall, horizontal and vertical acceleration profile of the saccade	[26]
	33 features for the overall, horizontal and vertical velocity profile of the saccade	[26]
	13 features for the curvature of a saccade	[7]
	Saccade angle	[26]
Fixational features	Number of fixations (count)	-
	Fixation duration in ms	[26]
	Horizontal and vertical position of fixation within centroid	[26]
	15 features for the overall, horizontal and vertical acceleration profile of the fixation	[26]
	15 features for distance, speed, and statistics for linear fit of drift during fixation	[26]
	17 features for the overall, horizontal and vertical velocity profile of the fixation	[26]
	Stationary gaze entropy (value)	[30]
	Gaze transition entropy (value)	[30]



(a) Open-population setting.

(b) Closed-population setting.

Fig. 2: Performance of a model with top 53 features, when training instances are split at threshold 0.01%. ROC curves for different input window sizes. Shaded bands show the standard error.

Table 2: Model performance for different input window sizes. AUC ± standard errors for a model with top 53 features, with and without using instances with a BAC label of above 0.00% and below 0.03% during training.

Training data	Input window size	AUC
Disregard training instances with 0 <BAC< 0.03%	60 seconds	0.6196 ± 0.0499
	20 seconds	0.5959 ± 0.0357
	10 seconds	0.6029 ± 0.0243
Split training instances at threshold 0.01%	60 seconds	**0.6941 ± 0.0387**
	20 seconds	0.6443 ± 0.0258
	10 seconds	0.6172 ± 0.0264

3. Results

3.1. Eye-Gaze and Eye-Closure Features are a Predictor for Alcohol Inebriation

We compare models that use the top 53 features determined by recursive feature elimination in the open-population setting. Table 2 compares the AUC of input window durations of 10, 20, and 60 seconds with a training data threshold at 0.1% and with training instances above 0 but below 0.03% discarded. The model performance increases with increasing input window size; this improvement comes at the cost of a longer testing duration. Based on a t-test, all six models are significantly better than random guessing ($p < 0.02$). The AUC appears to be highest for 60 seconds and for using all training instances instead of discarding low-BAC training instances, but the differences are not significant. Figure 2 (a) compares ROC curves for input windows of 10, 20, and 60 seconds. For all remaining experiments, we split training instances at a threshold of 0.01% and do not discard any instances.

3.2. Closed-Population Easier Than Open-Population Detection

Figures 2 (a) and 2 (b) compare ROC curves for probe durations between open- and closed-population settings. The AUC values are uniformly higher in the closed- than in the open-population setting; however, the differences do not reach statistical significance in a paired t-test. For 60 seconds, the AUC reaches 0.74 in the closed-population setting.

3.3. Acceleration and Velocity Profiles of Fixations and Saccades Indicate Inebriation

Table 3 shows the 25 most relevant features according to their SHAP value [18] after recursive feature elimination for an input window of 60 seconds in the open-population setting. A total of 15 out of the 25 highest-rated features characterize fixational acceleration and velocity profiles. A further six features relate to the count and characteristics of saccadic trajectories. Two features relate to the time steps with open eyes and the availability of the eye-tracking signal—when the tracker software fails to locate the iris and corneal glint, the eye gaze is not available.

Figure 3 shows a visualization of the impact of each feature to the model output according to their SHAP value. An inspection reveals that the visual stability is decreased in alcoholized participants: higher accelerations and velocities during fixations tend to lead to a higher likelihood of the positive (alcohol intoxicated) class. Furthermore, the saccadic behaviour is affected by alcohol intoxication: we can clearly see that a lower number of saccades in the input window increases the probability of the positive class. The impact of characteristics of the saccadic trajectories, however, is less apparent. A higher number of time steps without eye tracking signal and a lower number of time steps with open eyes increase the probability of a positive classification.

3.4. Omitting any Feature Group is Detrimental

In an ablation study, we investigate whether omitting feature groups results in a model deterioration in the open-population seting. Table 4 shows that removing either eye-closure, fixational, or saccadic features appears to have a moderate detrimental effect, but differences are not significant. Additionally, we study the effect of only using gaze entropy features [30]; this results in a significantly lower accuracy.

Table 3: Feature importance using mean absolute SHAP values. The table shows the 25 features with the highest mean absolute SHAP values.

No	Feature	SHAP value
1	Time steps with no eye tracking signal	0.01 ± 0.0
2	Median (over fixations) of median accelerations during fixation (deg/s^2) [26]	0.007 ± 0.0
3	Median (over fixations) of medians of vertical accelerations during fixations (deg/s^2) [26]	0.006 ± 0.0
4	Mean (over saccades) of mean vertical peak velocities during saccades (deg/s) [26]	0.006 ± 0.0
5	Median (over fixations) medians of velocities during fixations (deg/s) [26]	0.006 ± 0.0
6	Mean (over fixations) of kurtosis values of vertical velocities during fixations (deg/s) [26]	0.005 ± 0.0
7	Mean (over fixations) of kurtosis values of velocities during fixations (deg/s) [26]	0.005 ± 0.0
8	Mean (over saccades) of median velocities during saccades (deg/s) [26]	0.005 ± 0.0
9	Median (over fixations) of skewness values of velocities during fixations (deg/s) [26]	0.005 ± 0.0
10	Mean (over fixations) of skewness values of velocities during fixations (deg/s) [26]	0.005 ± 0.0
11	Time steps which are no saccade or fixation	0.005 ± 0.0
12	Standard deviation (over saccades) of median of vertical accelerations during saccades (deg/s^2)[26]	0.005 ± 0.0
13	Mean (over fixations) of skewness values of velocities during fixations (deg/s) [26]	0.004 ± 0.0
14	Median (over fixations) of means of vertical accelerations during fixations (deg/s^2) [26]	0.004 ± 0.0
15	Standard deviation (over saccades) of vertical peak velocities of saccades (deg/s^2) [26]	0.004 ± 0.0
16	Time steps with eye state “open”	0.004 ± 0.0
17	Median (over fixations) of mean accelerations during fixations (deg/s^2) [26]	0.004 ± 0.0
18	Median (over fixations) medians of vertical velocities during fixations (deg/s) [26]	0.004 ± 0.0
19	Median (over fixations) of standard deviations of vertical accelerations during fixations (deg/s^2)[26]	0.003 ± 0.0
20	Median (over fixations) of medians of horizontal velocities during fixations (deg/s) [26]	0.003 ± 0.0
21	Number of saccadees (count)	0.003 ± 0.0
22	Standard deviation (over saccades) of medians of velocities during saccades (deg/s) [26]	0.003 ± 0.0
23	Median (over fixations) of means of vertical drift velocities (deg/s) [26]	0.003 ± 0.0
24	Mean (over fixations) of standard deviations of vertical accelerations during fixations (deg/s^2)[26]	0.003 ± 0.0
25	Standard deviation (over saccades) of mean of vertical accelerations during saccades (deg/s^2) [26]	0.003 ± 0.0

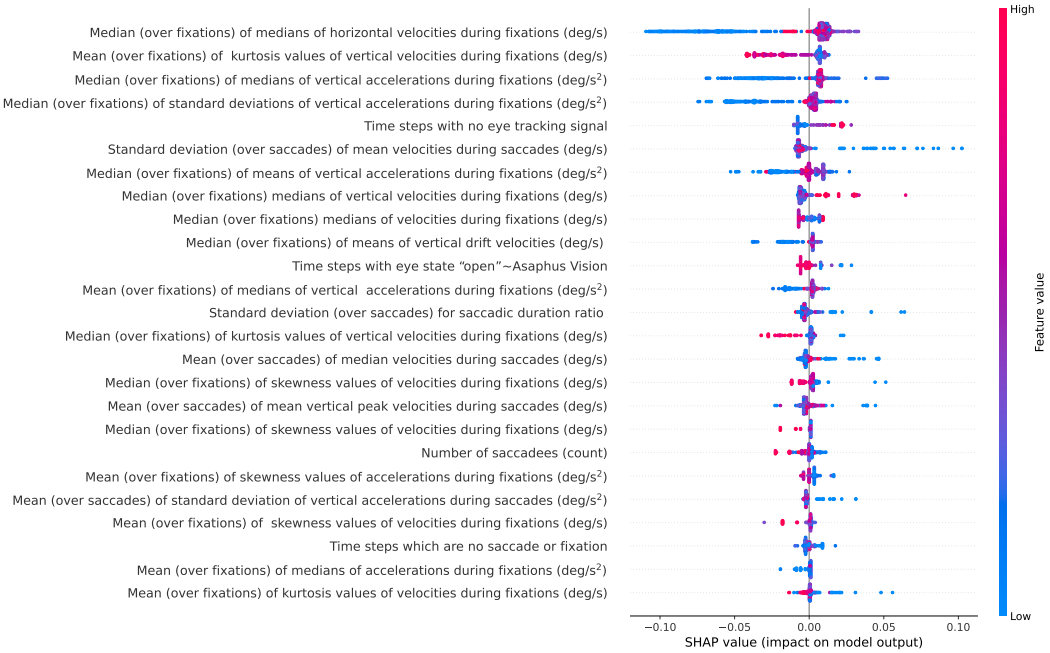
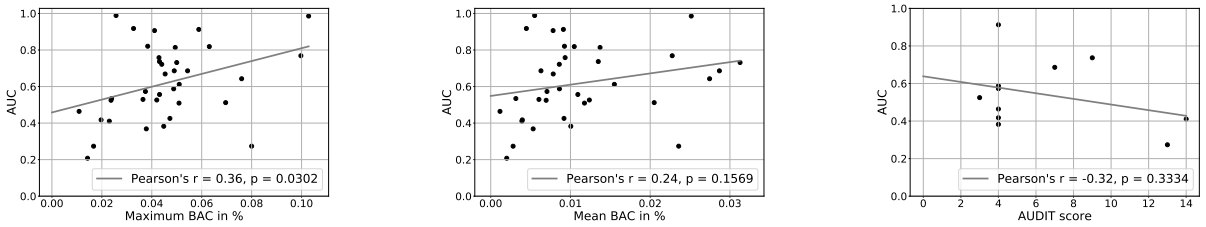


Fig. 3: SHAP summary plot. Visualization of the impact each feature has on the predicted probability of the positive (alcohol intoxicated) class, according to their SHAP value. The figure shows the 25 features with the highest mean absolute SHAP value, each dot represents a test instance.

Table 4: Ablation Study. AUC ± standard errors for models with different feature sets.

Feature set	AUC
All 53 features	**0.6941 ± 0.0387**
All 53 features without eye-closure features from Table 1	0.6470 ± 0.0265
All 53 features without saccade features from Table 1	0.6394 ± 0.0335
All 53 features without fixation features from Table 1	0.6744 ± 0.0210
Only gaze entropy features from Table 1	0.5530 ± 0.0128



(a) AUC over max BAC. (b) AUC over mean BAC. (c) AUC over AUDIT score.

Fig. 4: Relationship between the model AUC for an input window of 60 seconds and (a) the maximum and (b) the mean BAC measured in the alcohol session; and (c) the AUDIT score of the test participant. Each dot represents one test participant.

3.5. Detection Rates are Subject to High Inter-individual Variation

Participants' blood alcohol concentrations is estimated using a breathalyzer. Figure 4a shows the model's AUC over participants' maximum BAC, Figure 4b over participants' mean BAC during the recording session in the open-population setting. The linear correlation is statistically significant for Figure 4a ($p = 0.0302$) but not significant for Figure 4b ($p = 0.1569$). We conclude that the BAC explains some of the considerable inter-person variability in the model's performance that we observe in these figures. Figure 4c shows the model's AUC over the AUDIT-score that measures the degree of habitual drinking [2] in the open-population setting. The diagram suggests that there may be a negative correlation, but the correlation is not statistically significant ($p = 0.3334$).

4. Discussion

With 44 participants, the *Potsdam Binge / PVT data set* data set is not very large by machine-learning standards; it appears likely that a model trained on hundreds or thousands of participants would be considerably more accurate. It also appears plausible that a longer observation duration results in a more accurate detection, at the cost of a longer testing duration. Given that our study is motivated by applications such as monitoring drivers or operators of other hazardous machinery, we limit the observation periods in this study to 60 seconds.

The most indicative features characterize visual stability during fixations; intoxicated participants appear to have reduced visual stability during fixations. This matches our anecdotal observation that inebriated participants find it harder to fixate points during the tracker calibration procedure. A number of features characterize saccadic trajectories; this is consistent with earlier findings of reduced saccadic accuracy under the influence of alcohol [15, 27]. The eye tracker determines the eye gaze based on the locations of the pupil and the corneal reflection of a near-infrared illuminator—the *glint*. The single most relevant feature, the number of time steps without tracking signal, is related to the tracker's ability to locate the glint. Alcohol is secreted into tears and disturbs the ocular tear film [14]. We hypothesize that this may impact the tracker's ability to locate the corneal glint.

We observe that the inter-person variation of the model's performance is considerable. Whereas for some participants, intoxication can be detected with perfect accuracy, the model performs worse than random guessing for a significant number of individuals. The variation in maximum blood alcohol concentration explains some of this variability. The linear trend line suggests that a BAC in excess of 0.1% can be detected with high accuracy. The seemingly negative correlation between AUC and AUDIT score could be explained by an adaptation to the intoxicating effects of alcohol due to habitual consumption of alcohol. A larger population of participants who volunteer accurate information about their regular drinking behavior would be required to confirm this hypothesis.

Previous work [13] arrived at the conclusion that alcohol inebriation can be detected in facial images. However, this study was based on images in which participants appear more cheerful after imbibing than they are in their sober state. Since it is well known that neural networks can detect facial expressions [11], it appears imperative to account for the facial expression as a confounding factor.

Our experimental study includes several of the eye-lid and saccadic features that have previously been studied in the context of drowsiness detection. This includes PERCLOS, eye-closure amplitude ratios [34], and saccade durations,

velocities, and amplitudes [29]. Our study also includes velocity- and acceleration-profile features that have been used for oculomotoric biometric identification [26], gaze entropy that has been studied for intoxication detection [30], and saccadic curvature features that originate from cognitive neuroscience [7].

Alcohol consumption reduces the perceptive scope [31], and therefore the gaze transition entropy has been proposed as an indicator of intoxication [30]. However, a relatively fixed gaze is not exclusively tied to alcohol consumption, and our ablation study indicates that while gaze entropy performs just above chance level, the wide range of gaze features explored in this study leads to considerably higher detection rates.

Future research might investigate the use of additional features [17, 8] and of additional model classes. In a set of preliminary experiments, we found random forests to outperform linear models and neural networks. While linear models may be too limited in their expressiveness, we hypothesize that neural networks on raw input signals may outperform random forests on features in case training data become more abundant. For practical applications, using a single device to record the gaze and eye closure signal would be advantageous. Future research is needed to explore the use of low-frequency devices such as driver cameras which are already capable of extracting the driver's eye closure and eye gaze in order to detect distraction and drowsiness [12].

Alcohol consumption is a major contributor to the global burden of disease [4]. The associated health risks are known to be a complex multi-variational function of drinking behavior with considerable individual variation [32]. Health apps can help their users to adopt—and retain—healthier patterns of behavior [36]. Optical detection of inebriation could include alcohol consumption into the portfolio of trackable health factors. The WHO estimates that around 1.35 million traffic fatalities occur per year [21]; the NHTSA estimates that in the US, alcohol accounts for 28% of all traffic fatalities [33]. Regulators and safety rating programs are willing to take mitigating actions [9]. Technological advances towards widely usable inebriation detection therefore hold an enormous potential for societal benefit.

In order to promote and facilitate research on gaze-based intoxication detection, we will release the *Potsdam Binge / PVT data set*[1] to the research community upon acceptance of this manuscript.

5. Conclusions

We conclude that a machine-learning model based on participants' eye gaze and eye closure is able to detect moderate alcohol inebriation. With an AUC of 69% in the open-population setting and 74% in the closed-population setting across all participants, the model is significantly better than random guessing and therefore serves as a proof of concept; the AUC increases further for participants with higher BAC. The performance difference between open- and closed-population settings suggests that by adapting to a user's oculomotoric ideosyncrasies, alcohol detection on a personal device could reach a higher accuracy.

The most indicative features characterize visual stability during fixations; intoxicated participants appear to have reduced visual stability during fixations. However, omitting any group of features appears to be detrimental. We observe that the inter-person variation of the model's performance is considerable. This variation can only partially be explained by a variation in blood-alcohol concentration across participants.

Acknowledgements

This work was partially funded by the German Ministry of Education and Research (BMBF), grant 01|S20043.

References

[1] Aschbacher, K., Hendershot, C.S., Tison, G., Hahn, J.A., Avram, R., Olgin, J.E., Marcus, G.M., 2021. Machine learning prediction of blood alcohol concentration: a digital signature of smart-breathalyzer behavior. NPJ Digital Medicine 4, 1–10.

[2] Babor, T.F., Higgins-Biddle, J.C., Saunders, J.B., Monteiro, M.G., et al., 2001. The alcohol use disorders identification test. World Health Organization Geneva .

[3] Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

[1] https://osf.io/pvn2x/?view_only=2d81dc7a3c524767960553480a467dbf

[4] Buffington, C.K., 2007. Alcohol use and health risks: survey results. Bariatric Times 4, 1–21.

[5] Bühler, M., Mann, K., 2011. Alcohol and the human brain: A systematic review of different neuroimaging methods. Alcoholism: Clinical and Experimental Research 35, 1771–1793.

[6] Dinges, D., Powell, J., 1985. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behavior Research Methods, Instruments, & Computers 17, 652—-655.

[7] Doyle, M., Walker, R., 2001. Curved saccade trajectories: Voluntary and reflexive saccades curve away from irrelevant distractors. Experimental Brain Research 139, 333–344.

[8] Ebeid, I.A., Gwizdka, J., 2018. Real-time gaze transition entropy, in: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3204493.3208340, doi:10.1145/3204493.3208340.

[9] EuroNCAP, 2017. Euro ncap 2025 roadmap. EuroNCAP Technical Papers.

[10] EuroNCAP, 2021. European new car assessment programme assessment protocol – safety assist, safe driving v10.0. EuroNCAP Technical Papers.

[11] Gao, H., Yüce, A., Thiran, J.P., 2014. Detecting emotional stress from facial expressions for driving safety, in: 2014 IEEE International Conference on Image Processing (ICIP), IEEE. pp. 5961–5965.

[12] Halin, A., Verly, J.G., Van Droogenbroeck, M., 2021. Survey and synthesis of state of the art in driver monitoring. Sensors 21, 5558.

[13] Iamudomchai, P., Seelaso, P., Pattanasak, S., Piyawattanametha, W., 2020. Deep learning technology for drunks detection with infrared camera, in: 2020 6th International Conference on Engineering, Applied Sciences and Technology (ICEAST), IEEE. pp. 1–4.

[14] Kim, J., Kim, J., Nam, W., Yi, K., Choi, D., Hyon, J., Wee, W., YJ, S., 2012. Oral alcohol administration disturbs tear film and ocular surface. Ophthalmology 119, 965–971.

[15] King, A.C., Byars, J.A., 2004. Alcohol-induced performance impairment in heavy episodic and light social drinkers. Journal of Studies on Alcohol 65, 27–36.

[16] Koukiou, G., Anastassopoulos, V., 2015. Neural networks for identifying drunk persons using thermal infrared imagery. Forensic Science International 252, 69–76.

[17] Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., González Perilli, F., Pires, A., Vilaro, A., Villalobos, N., 2015. Gaze transition entropy. ACM Trans. Appl. Percept. 13. URL: https://doi.org/10.1145/2834121, doi:10.1145/2834121.

[18] Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems, pp. 4765–4774.

[19] MacDonald, T., Zanna, M., Fong, G., 1995. Decision making in altered states: Effects of alcohol on attitudes toward drinking and driving. Journal of Personality and Social Psychology 68, 973–985.

[20] Maurage, P., Masson, N., Bollen, Z., D'Hondt, F., 2020. Eye tracking correlates of acute alcohol consumption: A systematic and critical review. Neuroscience & Biobehavioral Reviews 108, 400–422.

[21] Organization, W.H., 2018. Global status report on road safety 2015. World Health Organization.

[22] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.

[23] Peragallo, J., Biousse, V., Newman, N., 2013. Ocular manifestations of drug and alcohol abuse. Current Opinion in Ophthalmology 24, 556–573.

[24] Peterson, K., 2004. Biomarkers for alcohol use and abuse: a summary. Alcohol Research & Health 28, 30.

[25] Quinet, J., Goffart, L., 2015. Cerebellar control of saccade dynamics: contribution of the fastigial oculomotor region. Journal of Neurophysiology 113, 3323–3336.

[26] Rigas, I., Friedman, L., Komogortsev, O., 2018. Study of an extensive set of eye movement features: Extraction methods and statistical analysis. Journal of Eye Movement Research 11.

[27] Roberts, W., Miller, M.A., Weafer, J., Fillmore, M.T., 2014. Heavy drinking and the role of inhibitory control of attention. Experimental and Clinical Psychopharmacology 22, 133.

[28] Salvucci, D.D., Goldberg, J.H., 2000. Identifying fixations and saccades in eye-tracking protocols, in: Proceedings of the 2000 Symposium on Eye tracking Research & Applications, pp. 71–78.

[29] Schleicher, R., Galley, N., Briest, S., Galley, L., 2008. Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? Ergonomics 51, 982–1010.

[30] Shiferaw, B.A., Crewther, D.P., Downey, L.A., 2019. Gaze entropy measures detect alcohol-induced driver impairment. Drug and Alcohol Dependence 204, 107519.

[31] Steele, C., Josephs, R., 1990. Alcohol myopia: Its prized and dangerous effects. Journal of the American Psychological Association 45, 921–933.

[32] Thakker, K.D., 1998. An overview of health risks and benefits of alcohol consumption. Alcoholism: Clinical and Experimental Research 22, 285s–298s.

[33] US Department of Transportation, 2021. Traffic Safety Facts. Research Note. National Highway Traffic Safety Administration.

[34] Wilkinson, V.E., Jackson, M.L., Westlake, J., Stevens, B., Barnes, M., Swann, P., Rajaratnam, S.M., Howard, M.E., 2013. The accuracy of eyelid movement parameters for drowsiness detection. Journal of Clinical Sleep Medicine 9, 1315–1324.

[35] Willoughby, C., Banatoski, I., Roberts, P., Agu, E., 2019. Drunkselfie: Intoxication detection from smartphone facial images, in: 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), pp. 496–501. doi:10.1109/COMPSAC.2019.10255.

[36] Zhao, J., Freeman, B., Li, M., et al., 2016. Can mobile phone apps influence people's health behavior change? an evidence review. Journal of Medical Internet Research 18, e5692.