2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
October 11-14, 2020. Toronto, Canada

Camera-based Driver Drowsiness State Classification Using Logistic Regression Models

Mohamed Hedi Baccour
Mercedes-Benz AG
Sindelfingen, Germany
mohamed_hedi.baccour@daimler.com

Frauke Driewer
Mercedes-Benz AG
Sindelfingen, Germany
frauke.driewer@daimler.com

Tim Schäck
Mercedes-Benz AG
Sindelfingen, Germany
tim.schaeck@daimler.com

Enkelejda Kasneci
University of Tübingen
Tübingen, Germany
enkelejda.kasneci@uni-tuebingen.de

***Abstract*—Drowsiness at the wheel is a major problem for traffic road safety. A drowsy driver suffers from decreased vigilance, increased reaction time and degraded decision-making ability, all of which have a huge impact on the driving performance. A driver monitoring system that warns the driver of his or her critical drowsiness state is a worthwhile contribution to traffic road safety. A drowsy driver typically exhibits some observable behaviors, such as eye blinking and head movements, that can be tracked using a camera. In this study, we analyze the potential of eye closure and head rotation signals, provided by a driver camera, to classify the driver's drowsiness state using logistic regression models. This analysis is based on a large dataset collected from 71 subjects in driving simulator experiments. A reliable and independent reference for drowsiness, however, is required in order to perform this analysis. For this purpose, we devise a methodology that merges several drowsiness monitoring approaches to construct a reliable reference for drowsiness. Furthermore, we describe our approach to extract eye blink and head rotation features. Ultimately, we design logistic regression classifiers and combine them using the one-vs-one binarization technique. Our approach achieves a global balanced validation accuracy of 72.7% on a three-class classification problem (awake, questionable and drowsy) by adopting a strict and rigorous evaluation scheme (i.e., leave-one-drive-out cross-validation).**

***Index Terms*—driving simulator, drowsiness, driver camera, driver monitoring system, ground truth construction, driver state classification, machine learning, logistic regression**

I. Introduction

A. Motivation

Driver drowsiness is a serious threat to road traffic safety. Drowsiness at the wheel leads to a substantial deterioration of the driving performance including increased reaction time, mitigated attention, diminished steering performance and impaired lane keeping ability. According to the National Highway Traffic Safety Administration, 4,111 fatalities from motor vehicle crashes in the US between 2013 and 2017 involved drowsy driving [1].

To eliminate this kind of risky driving and reduce the number of drowsiness-related motor vehicle crashes, many automobile manufacturers have equipped cars with driver-assistance systems that assess the driver's drowsiness and warn the driver of his or her critical state in case of detected drowsiness. Most of these systems monitor the steering pattern or the vehicle's position in the traffic lane to detect driving impairment and infer drowsiness. Contemporary vehicle-based technologies, however, have a limited ability to monitor drowsiness due to emerging car safety technologies such as lane keeping and lane centering assistance, which automatically steer and keep the vehicle in its lane.
A driver exhibits some observable behaviors indicating drowsiness, including slower eye blinks, drowsiness-related head movements such as reorienting the head from a leaning or tilting position and yawning, among other actions [2]. These observable behaviors, usually referred to as mannerism, can be unobtrusively tracked using a driver monitoring camera system and used to identify driver drowsiness. Such an approach also has the potential to analyze take-over readiness in the conditionally automated driving mode [3].

In this study, we investigate the potential of signals of a driver camera to detect driver drowsiness, specifically eye closure and head rotation signals. This study does not focus on how to extract such signals from camera images but, rather, on how to process these signals in order to detect drowsiness. A major challenge that arises when conducting such an investigation is the ground truth for drowsiness. Developing a driver's drowsiness monitoring system requires a reliable and independent measure for drowsiness against which the system can be checked, yet drowsiness is not a precisely or numerically defined quantity [2]. In this work, we devise a methodology that unobtrusively establishes a reliable reference for drowsiness by merging various drowsiness monitoring methods. We evaluate our approach utilizing a large dataset collected during simulated driving sessions with 71 participants.

B. Related work

The last decade has witnessed a rapid rise in the use of machine learning and deep learning algorithms to model driver drowsiness. This has been fostered not only by the growing availability of massive data sets, particularly because of substantial advances in computer vision, but also by the proliferation of user-friendly statistical modeling software. The most commonly used statistical modeling techniques in the context of drowsiness monitoring are logistic regression (e.g., [4] and [5]), support vector machines (e.g., [6] and [7]), artificial neural networks (e.g., [8] and [9]) and convolutional neural networks (e.g., [10] and [11]). The most widely adopted strategies to extract features and fit these statistical models include monitoring the driving performance (e.g., [8]), the blinking behavior (e.g., [6], [8] and [9]), the brain activity using electroencephalography (e.g., [12]), the heart activity using electrocardiography (e.g., [5]) and the breathing activity using thermal imaging (e.g., [7]). The most prominent measures to monitor the driving performance are the steering wheel angle and the standard deviation of lane position. While the former can easily be measured using a steering angle sensor, the latter requires an external lane monitoring camera, which is prone to measurement inaccuracies due to adverse weather conditions or poor lane markings. Both measures can be efficiently used for driver drowsiness monitoring as long as the driver actively steers the vehicle. Due to the growing use of automatic lane keeping systems in cars, these measures cannot be continuously monitored and such an approach has therefore only a limited functionality. Monitoring the brain's physiological signals as well as heart or breathing activity offers a very accurate assessment of a driver's drowsiness because physiological signals start to change in early stages of drowsiness [13]. Such an approach, however, typically requires intrusive sensors (e.g., electrodes), which makes it unsuitable for real-world automotive applications. Monitoring blinking behavior using a non-intrusive remote camera has established itself as a very promising approach monitoring driver drowsiness, especially in light of tremendous improvements in camera systems and computer vision. Regarding the ground truth for drowsiness in the scientific community, the reference for drowsiness is broadly obtained using self-reports (e.g., [4], [6] and [9]), video rating (e.g., [7] and [8]), the occurrence of line crossings (e.g., [14]) or by evaluation of brain activity signals (e.g., [15]).

C. Organization of the paper

In Section II, we describe the experimental setup and the collected data. In Section III, we introduce our methodology to establish a ground truth for drowsiness. Our approach to extract features from eye blinks and head movements is described in Section IV. In Section V, our method to classify the driver's drowsiness state is explained and the classification results are presented. Finally, a brief summary and discussion of the focus of future work are provided in Section VI.

TABLE I: Karolinska Sleepiness Scale

KSS	Description
1	Extremely alert
2	Very alert
3	Alert
4	Rather alert
5	Neither alert nor sleepy
6	Some signs of sleepiness
7	Sleepy, but no effort to keep alert
8	Sleepy, some effort to keep alert
9	Very sleepy, great effort to keep alert, fighting sleep

II. Data collection

As collecting data on very drowsy drivers is not feasible in real-world driving environments due to the high risk of traffic crashes, we carried out our experiments using the Mercedes-Benz driving simulator. Since it is a moving-base driving simulator, vehicle dynamics can be simulated in high fidelity and in a very realistic manner, enabling in-depth investigation of the driver and vehicle behavior on the road. Seventy-one healthy subjects (48 males, 23 females, 22–60 years old, $38 \pm 10\,\mathrm{years}$ (mean$\pm$standard deviation)) participated in this study. None of the participants was a professional driver and none reported to be sleep-deprived or suffering from sleep disorders. The participants were asked to avoid consuming coffee or energy drinks prior to their driving session so that it was more probable for them to become drowsy during the experiments.

The simulated route was a two-lane nighttime highway. The driving experiments (43–$204\,\mathrm{min}$, $115 \pm 36\,\mathrm{min}$, total duration=$136\,\mathrm{h}\ \ 8\,\mathrm{min}$) were conducted under highly monotonous driving conditions with visually unvarying scenery and low traffic density. Four driving experiments were conducted per day, which started at 8:15 a.m., 12:35 p.m., 4:35 p.m., and 8:30 p.m. The participants were asked to maintain the vehicle speed constant at $120\,{}^{\mathrm{km}}\!/\!_{\mathrm{h}}$. In some of the drives, the cruise control was used. Other advanced driver-assistance systems, such as the lane keeping assist or the lane departure warning, were deactivated so that driving performance measures, needed, in some cases, for further analysis of the driver's state, were not influenced. During the driving session, the participants were requested to assess their drowsiness level according to the Karolinska Sleepiness Scale (KSS) [16], see Table I. This request occurs every $15\,\mathrm{min}$ as a compromise between avoiding intrusive and activating feedback from the participant and high temporal resolution for the subjective assessment of drowsiness. While driving, the driver's face was filmed by a video camera and supervised by the experimenter sitting in the driving simulator control station. The experimenter continuously monitored the progression of drowsiness throughout the driving session and documented every noticeable sign of drowsiness. The dataset, with regard to blinking behavior and head movements, was collected using a camera mounted on the vehicle's dashboard directly in front of the driver. The camera uses infrared illumination and measures eye closure for each eye as well as head rotations

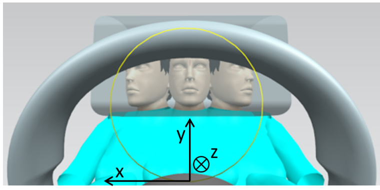

Fig. 1: Position and coordinate system of the driver camera

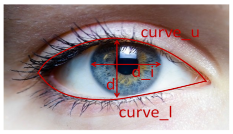

Fig. 2: Definition of $EyeClosure$

with respect to the three-dimensional coordinate system, as depicted in Fig. 1. The sampling frequency is $50\,\mathrm{Hz}$. The image processing algorithms of this camera do not fall within the scope of this work, as we only describe and use the signals provided by the camera. Such signals can be acquired with any commercially available camera. The most important signal in this study is $EyeClosure$, which is defined for each eye as follows:

$$EyeClosure = \max\left(1 - \frac{d}{d_I}, 0\right), \tag{1}$$

where d denotes the maximum vertical distance between the upper and lower eyelid and where d_I is the diameter of the iris, which is assumed to be constant for all drivers d_I=12 mm [17] (see Fig. 2). The camera also provides a confidence value for eyelid tracking because tracking is not always possible, e.g., when the driver's hand obstructs the camera's field of view. Furthermore, we make use of the head rotation signals $HeadRot_x$ and $HeadRot_z$. We refrain from using the signal $HeadRot_y$ in this study. This is due to the fact that head rotations with respect to the y-axis correspond mainly to side mirror checks and are, therefore, correlated with traffic density and not with a driver's drowsiness state.

III. Ground truth construction

A clear definition is helpful to obtain a measure for drowsiness. Despite advances in drowsiness research, there is still no consensus on how to unambiguously define drowsiness. Drowsiness is commonly referred to as "sleepiness", which is defined as the "inclination to sleep" [18]. Although defining this term typically necessitates a more complicated interdisciplinary consideration (e.g., neurobiological, physiological), the term drowsiness in the context of our work on driving refers to a combination of:

- observable driver behavioral cues considered a countermeasure to drowsiness,
- impaired driving performance indicators and
- a subjective feeling of drowsiness.

The self-estimations according to the KSS-scale, which were requested every $15\,\mathrm{min}$ during the experiments, represent a measure for the subjective feeling of drowsiness. The reliability of self-estimations in measuring drowsiness using KSS has been shown in previous studies (e.g., [19]). However, relying only on self-estimations is, in our view, not reliable enough. It is inconceivable that subjects who are unfamiliar with drowsy driving experiments always assess their drowsiness level correctly. Validating self-estimations by investigating additional drowsiness indicators in driver behavior leads us to obtaining a better ground truth.

TABLE II: HFC Drowsiness Scale

HFC	Description
1	Wide awake, vivid attention
2	Fresh and highly concentrated, focused attention
3	Attentive, but calm
4	Neither activated nor drowsy, reaction behavior without noticeable tendency
5	Somewhat dozy, digressive, but ready to respond
6	Signs of drowsiness, but effortlessly awake
7	Clearly drowsy, but mainly focused on the driving task
8	Fight against drowsiness, driving task is difficult, but largely perceptive
9	Absent-minded, listless, long periods without activity, microsleep is probable or occurs

Another practical approach to measuring drowsiness is video rating, which is particularly suitable for monitoring driver behavioral drowsiness-related cues. One commonly used scale for video rating is the HFC-Müdigkeitsskala® (HFC Drowsiness Scale), listed in Table II, which has been developed and adapted to the assessment of driver drowsiness [20]. Trained raters of the HFC Human-Factors-Consult GmbH [21] were instructed to perform the video rating by viewing one-minute segments of the driver's face, recorded prior to every KSS-request. Each segment was assigned a level of the HFC-Scale. In fact, each segment was rated independently by two raters and both ratings were subsequently averaged to obtain the final drowsiness level. It is possible, for this reason, for a segment to have a level of the scale with a decimal place. However, if the difference between two ratings for a video segment is greater than two, a third rater has to assess the driving session and the final drowsiness level is obtained by averaging the closest levels. Overall, 746 segments were rated in this way on the HFC-scale and each HFC value was compared to the corresponding self-rated KSS value. We found a global correlation between the self-rating and the video rating of $r = 0.533$ ($p < 0.001$) and an average individual correlation of $r = 0.674$. The heat map representation corresponding to the contingency table of the bivariate frequency distribution of the variables KSS and HFC is depicted in Fig. 3. The heat map representation provides a basic depiction of the interrelation between the two variables and exhibits the plausibility of video rating results as observations are most frequently distributed along the antidiagonal. The relationship between self-rating and video rating is additionally visualized using box plots as shown in Fig. 4. The tendency of the box plots helps to better crystallize the relationship between both scales, i.e., the KSS-scale and the HFC-scale. Based on

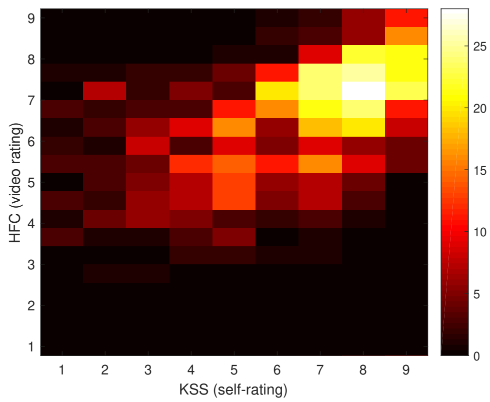

Fig. 3: Heat map representation of KSS and HFC

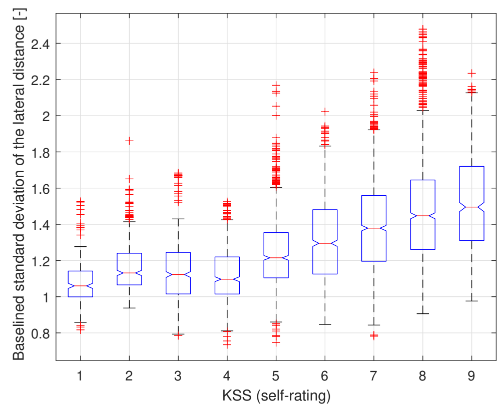

Fig. 5: Relationship between KSS and $StdLane$

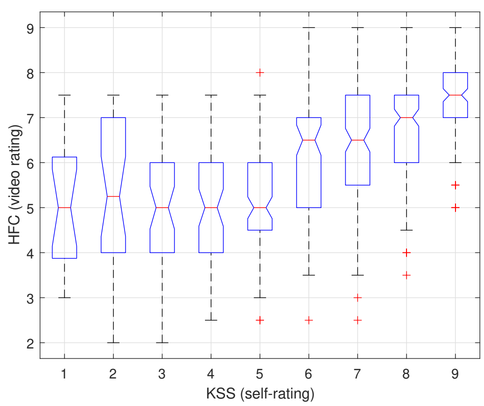

Fig. 4: Relationship between KSS and HFC

these box plots, we defined three drowsiness warning levels for both scales with three corresponding labels, as described in Table III. If the results of self-rating and video rating do not fall into the same drowsiness warning level, we have to investigate further drowsiness indicators. First, we made use of PERCLOS "PERcentage of eye CLOSure", which, in our work, is defined as the percentage of time that the eyes were 80% to 100% closed over a three-minute interval. PERCLOS has been demonstrated to be a reliable measure of drowsiness [22]. As reported in [23], three intervals are typically considered for PERCLOS when it is computed over a three-minute interval and are defined for three driver states as follows:

- $0\% \leqslant$ PERCLOS $< 9\%$: awake
- $9\% \leqslant$ PERCLOS $< 12\%$: questionable
- $12\% \leqslant$ PERCLOS: drowsy.

TABLE III: Drowsiness Warning Levels

	no warning	warning possible	warning necessary
KSS-scale range	1 - 5	6 - 7	8 - 9
HFC-scale range	1 - 5.5	6 - 6.5	7 - 9
Label	1	2	3

TABLE IV: Correlation Matrix

	PERCLOS	StdLane	SumExceed	HFC	KSS	$\text{KSS}_{\text{validated}}$
PERCLOS	1	0.275	0.182	0.459	0.358	0.419
StdLane	0.275	1	0.487	0.409	0.498	0.565
SumExceed	0.182	0.487	1	0,380	0,263	0,373
HFC	0.459	0.409	0.380	1	0.459	0.708
KSS	0.358	0.498	0.263	0.459	1	0.833
$\text{KSS}_{\text{validated}}$	0.419	0.565	0.373	0.708	0.833	1

Next, we examined two driving performance measures with respect to lane keeping behavior. These measures are the occurrence of lane departures denoted by $SumExceed$ (touching or exceeding of the left or right lane marking) and the baselined standard deviation of the lateral distance denoted by $StdLane$. The former indicates how often the driver unintentionally drifts out of his lane, while the latter provides a measure for the oscillations of the vehicle within the lane. The relationship between the KSS-scale and the baselined standard deviation of the lateral distance computed in a 10-minute sliding window is shown in Fig. 4. As the participants were not distracted during the driving sessions, both lane keeping measures mainly reflect drowsiness-related effects and, thus, serve well for the construction of ground truth.

In the event that, the results of self-rating and video rating do not fall into the same drowsiness warning level, analyzing the three aforementioned drowsiness indicators and examining comments from the experiment's supervisor allowed us to validate the label and, consequently, obtain a more reliable ground truth. A ground truth obtained in this way is referred to as $\text{KSS}_{\text{validated}}$. Fig. 6 shows the relationship between $\text{KSS}_{\text{validated}}$ and the video rating HFC using box plots. Clearly, the results of the video rating HFC are more plausible for $\text{KSS}_{\text{validated}}$ than for the self-rated KSS, which shows the positive effect of our validation procedure. The Pearson correlation coefficients of each pair of variables used in the generation of ground truth are summarized in the correlation matrix in Table IV. The correlation matrix shows that the degree of linear correlation of $\text{KSS}_{\text{validated}}$ to each variable is always higher than that of the self-rated KSS, which reveals the contribution of our validation procedure. In the remainder of this paper, $\text{KSS}_{\text{validated}}$ is used as ground truth for drowsiness.

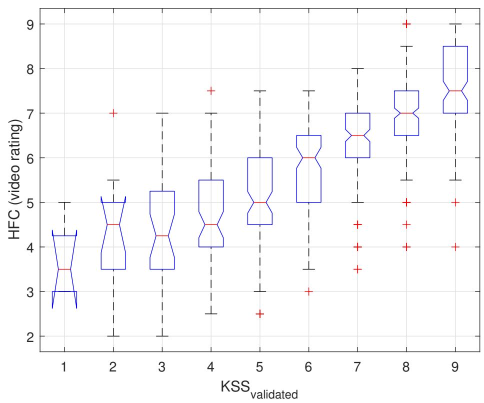

Fig. 6: Relationship between $\text{KSS}_{\text{validated}}$ and HFC

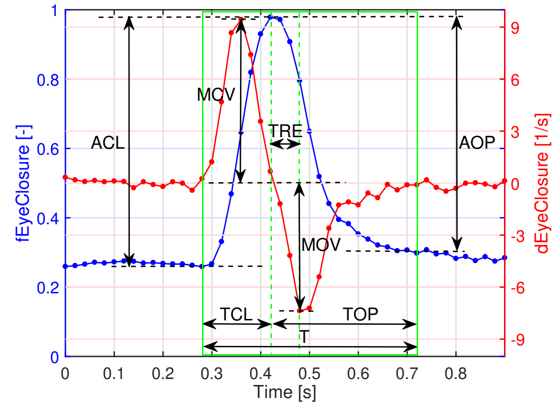

Fig. 7: Example of some extracted eye blink properties

IV. Feature Extraction

The purpose of this Section is to compute features that can both appropriately characterize the driver state and be used as inputs for a classifier. The features in this study are eye blink features extracted from the eye closure signal $EyeClosure$ and head movement features extracted from the signals $HeadRot_x$ and $HeadRot_z$. Before proceeding with the extraction of blink features, blinks must be reliably and accurately detected for each eye in the corresponding eye closure signal. This is a fundamental step of a camera-based drowsiness detection system because its performance relies heavily on the performance of the blink detection algorithm. Our blink detection algorithm is based on the filtered signal $fEyeClosure$ and its first derivative $dEyeClosure$. Both signals are computed from $EyeClosure$ using the Savitzky-Golay filter [24]. Our blink detection algorithm is thoroughly described and evaluated in our published work [25]. From each detected blink, blink properties (e.g., duration, amplitude, velocity etc.) are extracted for each eye separately. An example of some extracted blink properties in this work is depicted in Fig. 7. In this Fig., the time-based properties T, TCL, TOP and TRE denote the blink duration, closing duration, opening duration and the reopening duration, respectively. The amplitude-based properties ACL and AOP denote the closing and opening amplitude, respectively. The velocity-based properties MCV and MOV denote the maximum closing and opening velocity, respectively. As one single blink cannot individually reflect blinking behavior through its properties, blink features must be computed in a sliding time window by compressing the values of each property for many blinks using an appropriate statistical measure such as the arithmetic mean or the standard deviation. Finally, the two features $feature_l$, for the left eye, and $feature_r$, for the right eye, are averaged to obtain the final feature $feature_{uni}$, characterizing the blinking behavior of both eyes. The procedure for extracting the eye blink features is described in Fig. 8. The head movement features are extracted analogously, i.e., in a sliding time window. The feature $HeadNodding$ is the standard deviation of the signal segment $HeadRot_x$ within the time window.

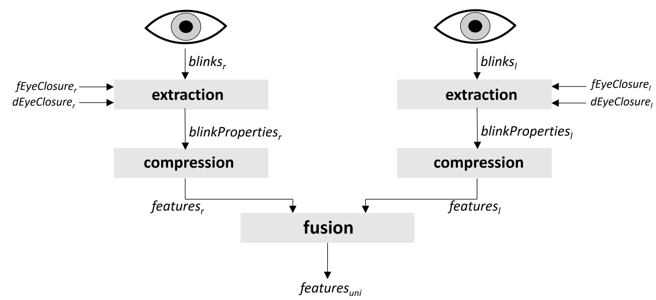

Fig. 8: Eye blink feature extraction approach

The features $HeadBobbing_{std}$ and $HeadBobbing_{msq}$ are the standard deviation and the mean square of the signal segment $HeadRot_z$ within the time window, respectively. In this work, the time window for all extracted features has a length of ten minutes and is shifted over the entire drive with a step of one minute. The driver's head movements and, particularly, the blinking behavior differ widely from person to person due to a range of factors including ethnicity, gender, humidity, luminosity, among others. These tremendous inter-individual differences are challenging when it comes to classifying the driver's drowsiness state. To tackle this challenge, a feature baselining is adopted, which consists in comparing a current feature to a historical baseline. The baseline for each feature is its value in the first ten minutes of the drive. The baselining is performed by dividing the current feature by its baseline. In this way, 35 baselined features are extracted overall. The only features selected for classification in our work are reported and listed with greater details in Table V, which also shows the Pearson correlation coefficient ρ between each feature and $\text{KSS}_{\text{validated}}$. In this Table, the asterisks ** and *** stand for a statistical significance level $p < 0.01$ and $p < 0.001$, respectively.

V. Driver State Classification

A. Theoretical background of logistic regression

In logistic regression, given a binary output variable $y \in \{0, 1\}$ and a feature vector $\boldsymbol{x} \in \mathbb{R}^{1\times p}$, the conditional prob-

TABLE V: Feature Details

Feature	Description	ρ
f	blink frequency	0.178^{***}
TCL_{mean}	mean closing duration	0.403^{***}
TOP_{mean}	mean opening duration	0.290^{***}
TRE_{mean}	mean reopening duration	0.533^{***}
ACL_{mean}	mean closing amplitude	0.035^{**}
AOP_{mean}	mean opening amplitude	0.093^{***}
MCV_{mean}	mean maximum closing velocity	-0.114^{***}
MOV_{mean}	mean maximum opening velocity	-0.145^{***}
$APMVCL_{mean}$	mean ratio of amplitude to max velocity of the closing phase	0.380^{***}
$APMVOP_{mean}$	mean ratio of amplitude to max velocity of the opening phase	0.336^{***}
$T80PT_{mean}$	mean ratio of the duration where the eye is more than 80% closed to the blink duration	0.413^{***}
$PoCE_{blink}$	percentage of closed eyes	0.230^{***}
Wf_{mean}	mean blink waveform area	0.443^{***}
L_d	eyelid aperture without blinks	-0.387^{***}
$asymmetry_{mean}$	mean blink asymmetry	0.049^{***}
$HeadNodding_{std}$	std of $HeadRot_x$	0.433^{***}
$HeadBobbing_{std}$	std of $HeadRot_z$	0.374^{***}
$HeadBobbing_{msq}$	mean square of $HeadRot_z$	0.343^{***}
TCL_{std}	std of the closing duration	0.371^{***}
TRE_{std}	std of the reopening duration	0.483^{***}
AOP_{std}	std of the opening amplitude	0.239^{***}
AOV_{std}	std of the average opening velocity	0.242^{***}
$APMVCL_{std}$	std of the ratio of amplitude to max velocity of the closing phase	0.348^{***}

ability $P(y=1|\boldsymbol{x})$ is modeled using a sigmoid function as follows

$$P(y=1|\boldsymbol{x}) = p(\boldsymbol{x},\beta_0,\boldsymbol{\beta}) = \frac{1}{1+e^{-(\beta_0+\boldsymbol{\beta}\cdot\boldsymbol{x})}}, \tag{2}$$

where $\beta_0 \in \mathbb{R}$ and $\boldsymbol{\beta} \in \mathbb{R}^{1\times p}$. The parameters $(\beta_0, \boldsymbol{\beta})$ of the logistic regression model are fitted using maximum likelihood estimation by maximizing the likelihood function

$$\max_{\beta_0,\beta_1,\ldots,\beta_p} L(\beta_0,\boldsymbol{\beta}) = \prod_{i=1}^{n} p(\boldsymbol{x}_i,\beta_0,\boldsymbol{\beta})^{y_i}\left(1-p(\boldsymbol{x}_i,\beta_0,\boldsymbol{\beta})\right)^{1-y_i}, \tag{3}$$

where n denotes the number of instances in the dataset. To solve (3), the partial derivatives of the likelihood function with respect to $\beta_0, \beta_1, \ldots, \beta_p$ must be computed, which is not trivial due to the products. Since the maximum of a function does not change with monotonic transformation and the log-function is a strictly increasing function, the log-likelihood function is maximized to solve (3). This turns the products in (3) into sums and yields

$$\begin{aligned} l(\beta_0,\boldsymbol{\beta}) &= \sum_{i=1}^{n} y_i \log\left(p(\boldsymbol{x}_i,\beta_0,\boldsymbol{\beta})\right) \\ &\quad + (1-y_i)\log\left(1-p(\boldsymbol{x}_i,\beta_0,\boldsymbol{\beta})\right) \\ &= -\sum_{i=1}^{n}\log\left(1+e^{\beta_0+\boldsymbol{\beta}\cdot\boldsymbol{x}_i}\right) + \sum_{i=1}^{n} y_i\left(\beta_0+\boldsymbol{\beta}\cdot\boldsymbol{x}_i\right) \\ &\to \max_{\beta_0,\beta_1,\ldots,\beta_p}. \end{aligned} \tag{4}$$

To solve (4), the partial derivatives with respect to each parameter $\beta_0, \beta_1, \ldots, \beta_p$ are set to zero, which yields the

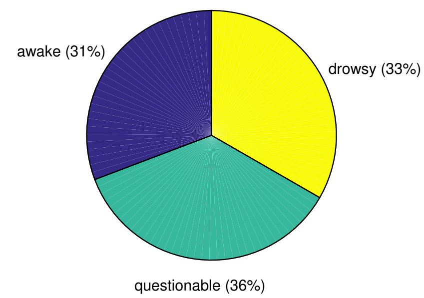

Fig. 9: Class distribution in the dataset

following nonlinear system of (p+1) equations

$$\begin{aligned} \frac{\partial l}{\partial \beta_j} &= -\sum_{i=1}^{n}\frac{1}{1+e^{\beta_0+\boldsymbol{\beta}\cdot\boldsymbol{x}_i}}e^{\beta_0+\boldsymbol{\beta}\cdot\boldsymbol{x}_i}x_{ij} + \sum_{i=1}^{n} y_i x_{ij} \\ &= \sum_{i=1}^{n}\left(y_i - p(\boldsymbol{x}_i,\beta_0,\boldsymbol{\beta})\right)x_{ij} = 0, \end{aligned} \tag{5}$$

with $j = 0, 1, \ldots, p$ and $x_{i0} = 1\ \forall i = 1, \ldots, n$. The nonlinear system of (p+1) equations (5) is solved numerically using the Newton-Raphson method [26].

B. Classification with logistic regression

In this study, driver state classification consists in classifying the driver state into one of three classes: awake (class 1) with $\text{KSS}_{\text{validated}} \in [1,5]$, questionable (class 2) with $\text{KSS}_{\text{validated}} \in [6,7]$ and drowsy (class 3) with $\text{KSS}_{\text{validated}} \in [8,9]$. The class distribution of this three-class classification problem is depicted in Fig. 9, which shows a nearly equal distribution of classes in the dataset. We convert this three-class problem into multiple two-class problems by means of the one-vs-one binarization technique. In this technique, three binary classifiers are trained for each pair of classes, i.e., (1,2), (1,3) and (2,3). To obtain the final predicted class, a max-wins voting strategy is adopted. In this voting strategy, every classifier assigns the instance to one of the two classes. Then, the vote for the class to which the instance is assigned is increased by one vote. Finally, the class with the most votes is the final predicted class of the combined three classifiers.

To avoid overfitting in the design of the logistic regression models, we adopt a *leave-one-drive-out* cross-validation strategy of N_{ds}=71 folds, where N_{ds} denotes the total number of the driving sessions. In this strategy, the model is fitted using the samples of N_{ds}-1 driving sessions and evaluated on the samples of the left-out driving session (unseen data). The conventional k-fold cross-validation approach, in which the data are randomly split into training and validation sets, was rejected in this work because both sets can contain samples of the same subjects. Therefore, the validation set cannot be considered as containing entirely unseen samples. This leads to an inflation of the classification rates and to erroneous conclusions about the classifier performance.

Furthermore, stepwise forward and backward feature selection

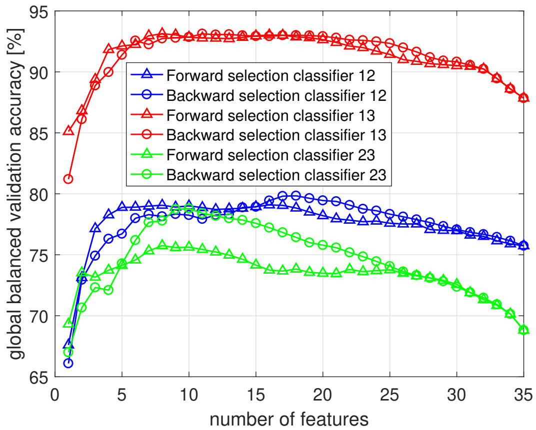

Fig. 10: Feature selection results

TABLE VI: Confusion Matrix

		predicted		
		awake	questionable	drowsy
given	awake	79.0%	20.2%	0.8%
	questionable	18.8%	65.3%	15.9%
	drowsy	2.7%	23.6%	73.7%

techniques are carried out to determine the most discriminative feature subset for each classifier and improve its prediction performance and generalization. The metric guiding the feature subset search is the global balanced validation accuracy. The logistic regression model yielding the highest global balanced validation accuracy from the stepwise forward and backward feature selection is selected to perform the binary classification for each pair of classes.

C. Results

The results of the stepwise forward and backward selection for each classifier are depicted in Fig. 10. As anticipated, the classifier$_{13}$ achieves the highest accuracy since it is manifestly easier to discriminate between the awake and drowsy class than to discriminate between the two classes within each other pair of classes. The parameters and selected features of the best classifier for each pair of classes in terms of global balanced validation accuracy are shown is Fig. 11, Fig 12 and Fig. 13. After combining the three classifiers and applying the one-vs-one binarization technique, we achieve a global balanced validation accuracy of 72.7% on this three-class classification problem (awake, questionable and drowsy). The confusion matrix obtained with this approach is given in Table VI.

VI. Conclusion

Through this work, we have contributed to the field of driver drowsiness monitoring. We have devised a methodology which establishes a ground truth for drowsiness by merging several drowsiness monitoring approaches and drowsiness measures, specifically self-reports KSS, video rating HFC, PERCLOS, the occurrence of lane departures $SumExceed$, the baselined standard deviation of the lateral distance $StdLane$ and the

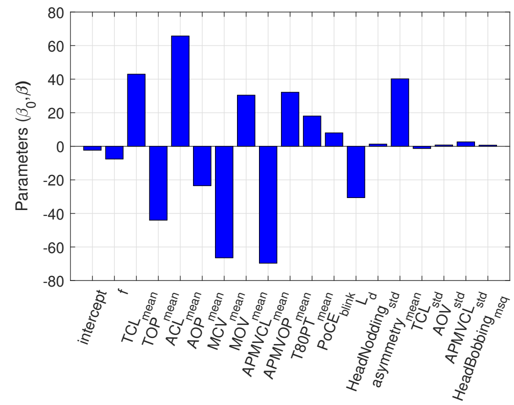

Fig. 11: Parameters and selected features of classifier$_{12}$

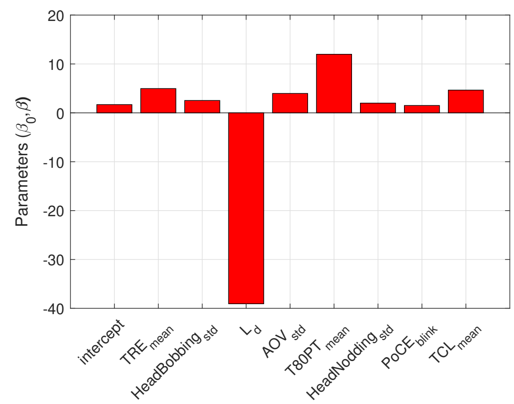

Fig. 12: Parameters and selected features of classifier$_{13}$

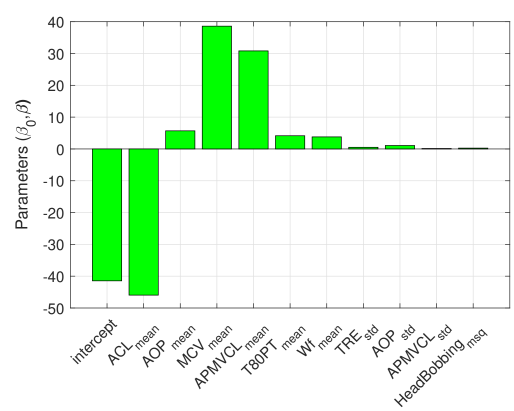

Fig. 13: Parameters and selected features of classifier$_{23}$

objective comments of the experimenter. Drowsiness is a multidimensional phenomenon. In its different forms, drowsiness is distinguished by distinct configurations of characteristics and, therefore, requires a multidimensional consideration of individual indicators to construct a reliable reference.

Furthermore, we have explored a remote camera's potential to classify the driver state using logistic regression models. The features extracted from the head rotation signals turned out to be relevant as they were always selected for the classification. Despite the simplicity and interpretability of the logistic regression models, they have achieved a satisfactory global balanced validation accuracy of 72.7% on three-class classification (awake, questionable and drowsy) by adopting a strict and rigorous evaluation scheme (i.e., leave-one-drive-out).

In future work, further statistical modeling approaches have to be applied to our data. Highly nonlinear approaches such as support vector machines or artificial neural networks can potentially provide better classification rates. This comes, however, at the expense of a less interpretable model, usually requiring an increased computing power in the vehicle. Finally, our findings have to be assessed in real-world environments. Although this study was conducted in a very realistic moving-base driving simulator, several factors in real-world driving scenarios including illumination, vibration in the vehicle and varying traffic density have to be examined to verify the validity of our methods.

REFERENCES

[1] https://www.nhtsa.gov/risky-driving/drowsy-driving

[2] W. W. Wierwille, L. A. Ellsworth, Evaluation of Driver Drowsiness by Trained Raters. Accident Analysis & Prevention, Vol. 26, Issue 5, 1994.

[3] C. Braunagel, W. Rosenstiel and E. Kasneci, Ready for Take-Over? A New Driver Assistance System for an Automated Classification of Driver Take-Over Readiness, IEEE Intelligent Transportation Systems Magazine, vol. 9, no. 4, pp. 10-22, 2017.

[4] A. Murata, Y. Fujii, K. Naitoh, Multinomial Logistic Regression Model for Predicting Driver's Drowsiness Using Behavioral Measures. Procedia Manufacturing, Vol. 3, pp. 2426-2433, 2015.

[5] M. Babaeian, N. Bhardwaj, B. Esquivel M. Mozumdar, Real Time Driver Drowsiness Detection Using a Logistic-regression-based Machine Learning Algorithm. IEEE Green Energy and Systems Conference (IGSEC), Long Beach, USA, pp. 1-6, 2016.

[6] P. Ebrahim, Driver Drowsiness Monitoring using Eye Movement Features Derived from Electrooculography. Dissertation University of Stuttgart, Germany, 2016.

[7] S. E. H. Kiashari, A. Nahvi, H. Bakhoda et al., Evaluation of Driver Drowsiness Using Respiration Analysis by Thermal Imaging on a Driving Simulator. Multimed Tools Appl, 2020.

[8] C. Jacobé de Naurois, C. Bourdin, A. Stratulat, et al., Detection and Prediction of Driver Drowsiness Using Artificial Neural Network Models. Accident Analysis & Prevention, Vol. 126, pp. 95-104, 2019.

[9] F. Friedrichs, B. Yang, Camera-based Drowsiness Reference for Driver State Classification under Real Driving Conditions. IEEE Intelligent Vehicles Symposium, San Diego, USA, pp. 101-106, 2010.

[10] Q. Massoz, J. G. Verly, M. Van Droogenbroeck, Multi-Timescale Drowsiness Characterization Based on a Video of a Driver's Face. Sensors, 18:2801, 2018.

[11] B. Reddy, Y. Kim, S. Yun et al., Real-Time Driver Drowsiness Detection for Embedded System Using Model Compression of Deep Neural Networks. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, USA, pp. 438-445, 2017.

[12] D. Wu, V. J. Lawhern, S. Gordon et al., Driver Drowsiness Estimation from EEG Signals Using Online Weighted Adaptation Regularization for Regression (OwARR). IEEE Transactions on Fuzzy Systems, vol. 25, pp. 1522-1535, 2017.

[13] A. Sahayadha, K. Sundaraj , M. Murugappan, Detecting Driver Drowsiness Based on Sensors: A Review. Sensors, 12(12), 16937-16953, Basel, Switzerland, 2012.

[14] Y. Liang, W. J. Horrey, M. E. Howard et al., Prediction of Drowsiness Events in Night Shift Workers During Morning Driving. Accident Analysis & Prevention, Vol. 126, pp. 105-114, 2019.

[15] C. François, T. Hoyoux, T. Langohr et al., Tests of a New Drowsiness Characterization and Monitoring System Based on Ocular Parameters. International Journal of Environmental Research and Public Health, Vol. 13, 2016

[16] T. Akerstedt, M. Gillberg, Subjective and Objective Sleepiness in The Active Individual. The International Journal of Neuroscience, 52(1-2):29-37, 1990.

[17] A. N. Bashkatov, E. A. Genina, E. V. Koblova, V. V. Tuchin, Estimation of Melanin Content in Iris of Human Eye. Saratov Journal of Medical Scientific Research, Vol. 5688, 2008.

[18] J. C. Stutts, J. W. Wilkins, B. V. Vaughn, Why Do People Have Drowsy Driving Crashes? Input from Drivers Who Just Did. AAA Foundation for Traffic Safety, 1999.

[19] K. Kaida, M. Takahashi, T. Akerstedt, A. Nakata, Y. Otsuka, T. Haratani, K. Fukasawa, Validation of The Karolinska Sleepiness Scale Against Performance and EEG Variables. Japan National Institute of Occupational Safety and Health, Kawasaki, Japan, 2006.

[20] H. Kolrep, M. Rimini-Döhring, A. Oehme, T. Jürgensohn, T. Altmüller, Wie sieht müde aus? – Entwicklung und Validierung einer Skala zur Müdigkeitsbewertung von Kraftfahrern. In Urbas, L., & Steffens, C. (Hrsg.). 6. Berliner Werkstatt Mensch-Maschine-Systeme. Zustandserkennung und Systemgestaltung. 13. 15. Okt. 2005, Berlin. Düsseldorf: VDI-Verlag, S. 65-70, 2005 (in German).

[21] https://human-factors-consult.de/en/

[22] W., Wierwille, L.A. Ellsworth, S. Wreggit, J. Fairbanks, L. Kirn, Research on Vehicle-based Driver Status/Performance Monitoring: Development, Validation, and Refinement of Algorithms for Detection of Driver Drowsiness. National Highway Traffic Safety Administration (NHTSA), 1999.

[23] R. J. Hanowski, The Drowsy Driver Warning System Field Operational Test: Data Collection Methods. NHTSA, 2008.

[24] A. Savitzky, M. J. E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry, pp. 1627-1639, 1964.

[25] M. H. Baccour, F. Driewer, E. Kasneci, W. Rosenstiel, Camera-based Eye Blink Detection Algorithm for Assessing Driver Drowsiness. IEEE Intelligent Vehicles Symposium (IV), Paris, France, pp. 987-993, 2019.

[26] C. R. Shalizi, Advanced Data Analysis from an Elementary Point of View. Cambridge University Press, 2019.