

Article

# A Framework for Instantaneous Driver Drowsiness Detection Based on Improved HOG Features and Naïve Bayesian Classification

Samy Bakheet<sup>1,2,\*</sup>  and Ayoub Al-Hamadi<sup>2</sup> 

<sup>1</sup> Department of Information Technology, Faculty of Computers and Information, Sohag University, P. O. Box 82533 Sohag, Egypt

<sup>2</sup> Institute for Information Technology and Communications (IKT), Otto-von-Guericke University Magdeburg, 39106 Magdeburg, Germany; ayoub.al-hamadi@ovgu.de

\* Correspondence: samy.bakheet@fci.sohag.edu.eg

**Abstract:** Due to their high distinctiveness, robustness to illumination and simple computation, Histogram of Oriented Gradient (HOG) features have attracted much attention and achieved remarkable success in many computer vision tasks. In this paper, an innovative framework for driver drowsiness detection is proposed, where an adaptive descriptor that possesses the virtue of distinctiveness, robustness and compactness is formed from an improved version of HOG features based on binarized histograms of shifted orientations. The final HOG descriptor generated from binarized HOG features is fed to the trained Naïve Bayes (NB) classifier to make the final driver drowsiness determination. Experimental results on the publicly available NTHU-DDD dataset verify that the proposed framework has the potential to be a strong contender for several state-of-the-art baselines, by achieving a competitive detection accuracy of 85.62%, without loss of efficiency or stability.

**Keywords:** driver drowsiness detection; HOG features; shifted orientations; NB classification; NTHU-DDD dataset



**Citation:** Bakheet, S.; Al-Hamadi, A. A Framework for Instantaneous Driver Drowsiness Detection Based on Improved HOG Features and Naïve Bayesian Classification. *Brain Sci.* **2021**, *11*, 240. <https://doi.org/10.3390/brainsci11020240>

Academic Editor: Anirban Dutta  
Received: 7 December 2020  
Accepted: 9 February 2021  
Published: 14 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Globally, an average of 3200 persons die each day around the world due to road traffic crashes (RTCs). It is estimated that driver-related dangerous behaviors such as drowsiness, drug and alcohol use, inexperience and psychological stress, are contributing factors in the vast majority of these crashes, and driver drowsiness is the most commonly reported reason among non-performance errors that accounted for such crashes [1]. For example, in the U.S., the National Highway Traffic Safety Administration (NHTSA) reported that drowsy driving was responsible for an estimated 3662 fatal crashes and 4121 fatalities from 2011 to 2015, which corresponds to 2.4 percent of all fatal crashes and 2.5 percent of all crash fatalities recorded in the U.S. during the same period [2].

The relationship between driver drowsiness and crash risk has been investigated extensively in numerous studies, with the purpose of identifying and quantifying the increased risk. For instance, in [3], Williamson et al. provides clear evidence for sleep homeostatic effects producing impaired performance and accidents. Additionally, it has been reported that drowsy drivers are more likely to be involved in sleep-related crashes or near-crashes than attentive drivers by nearly 4–6 times. Recently, in a case-control study of heavy-vehicle drivers, Stevenson et al. [4] found that cumulative sleep deprivation or sleep debt can also increase the likelihood of having a serious crash. Furthermore, in [5], Li et al. stated that drowsiness can seriously impair the ability to drive properly, as people find it difficult to maintain their attention on the task. In other words, lack of sleep can make drivers less alert and affect their coordination, judgement and reaction time while driving.

Over the past few years, several approaches and techniques that significantly contribute to reducing road trauma have been developed, such as training drivers in improved fatigue-

management practices—e.g., having the required rest breaks [6]. This depends on subjective measures consisting of self-assessment of one's drowsiness levels. A relatively recent study [7] found that drivers have some ability to identify their current state of being drowsy and likelihood of falling asleep. Despite self-assessment of drowsiness being a relatively good initial coping mechanism, it is not able to get rid of drowsiness-related road trauma completely, so complementary warning and safety systems urgently need to be developed.

Technological innovations hold great potential to drive a considerable reduction in the number of both injuries and fatalities associated with road accidents, through alerting drivers to their drowsy state prior to accidents. For example, a study by Blommer et al. [8] has reported a significant improvement in driver reaction times in lane-departure scenario, when a warning signal is issued. The authors also concluded that the manner in which such warning signals are issued is shown to be not absolutely crucial. In other words, visual, auditory and haptic warnings are all equally operative.

Drowsiness detection methodologies found in the literature are generally categorized into three main categories, namely, vehicle-based measurements, physiological measurements and computer vision techniques. In traditional vehicle-based measurement approaches, sudden or large corrections in traveling directions are typically detected through calculating the values of a number of driving behavior metrics, such as deviations from lane position [9] and movements of the steering wheel [10]. An alternative approach for detecting changes in driver alertness level is to monitor and trace specific internal signals, such as heart rate variability [11] or brain activity [12]. However, due to the need for multiple sensors, physiological measurement approaches often turn out to be less practically feasible in the real world, compared to both vehicle-based and computer vision approaches [13]. In this context, it is worth mentioning that the great disadvantage with physiological sensors is that they are obtrusive and will thus never be used in a production vehicle.

Computer vision techniques based on artificial neural networks (ANNs) have been successfully (and still being) applied to many road safety problems (e.g., traffic safety analysis of toll plazas [14] and identifying behavioral changes among drivers [15]). These techniques have also emerged as leading architectures for many visual recognition tasks [16]. In the past few years, driver drowsiness detection has drawn great attention from the computer vision and object recognition community. In [17], Park et al. presented an automated system for driver drowsiness detection using three pre-trained deep neural networks (i.e., VGG-FaceNet, AlexNet and FlowImageNet) and two ensemble strategies (independently averaged architecture and feature-fused architecture) to classify each frame in an input video sequence as drowsy or not. In a similar vein, in [18], the authors proposed a driver drowsiness detection method using a 3D deep neural network along with a boosting framework for semi-supervised learning to improve supervised learning.

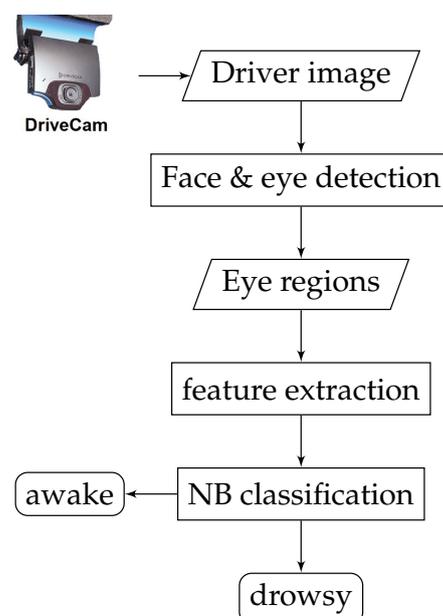
In [19], Rateb et al. presented a deep learning based approach for real-time drowsiness detection that can be monolithically integrated into android applications with a high accuracy rate. The primary contribution in this work involves the compression of a heavy baseline model to a lightweight model. Further, in this approach, a minimized network structure is designed to determine whether the driver is drowsy or not, depending on facial landmark (key point) detection. In [20], an eye blinking detection approach was presented using innovative color and texture segmentation algorithms, where the driver's facial features are obtained by a facial segmentation and neural network-based algorithm. The obtained features are then utilized for iris tracking and eye blinking detection. In their method, each eye closure longer than 220 ms is identified as drowsy or asleep.

In [21], Pauly and Sankar presented a method for drowsiness detection based on traditional histogram of oriented gradient (HOG) features and support vector machines (SVMs) for blink detection. The method was validated on their own dataset, achieving an overall drowsiness detection accuracy of 91.6%, by comparing the prediction of the developed system with that of a human observer. Moreover, in [22], a face-monitoring based framework for drowsy driver detection was proposed, where a concise face texture descriptor is utilized to identify the most discriminant drowsiness features. Similarly,

in [23], Singh et al. also proposed the use of HOG feature extraction and linear SVM classification to detect oncoming driver fatigue and issue a sufficiently early warning to help in preventing the accident. In this paper, the main contributions can be summarized as follows. First, a framework is developed for designing a vision-based system for instantaneous driver drowsiness detection. Secondly, we introduce an innovative feature descriptor depicting an improved version of HOG features based on binarized histograms of shifted orientations. The naïve Bayes (NB) classification model is then modified by adding the correlation between the data samples. This not only allows the NB algorithm to be a dependent hypothesis, but also enables the proposed framework to effectively tackle the problem of the classification of large-scale datasets. The remainder of this paper is structured as follows. In Section 2, we provide a detailed description of the proposed system for instantaneous driver drowsiness detection. Experimental evaluation results are then reported and discussed in Section 3. Finally, Section 4 provides concluding remarks with some thoughts for future work.

## 2. Proposed Architecture

In this section, the details of the proposed framework for driver drowsiness detection are introduced. A functional block diagram for the key framework steps is depicted Figure 1. The brief explanation of the methodology for detecting driver drowsiness using the proposed framework is as follows. The input drive image captured by a dashboard mounted camera is initially preprocessed by applying adaptive contrast-limited histogram equalization for reducing the fluctuations in lighting intensity and thus enhancing the overall brightness and contrast in the image. Then, the driver's face is detected by a cascaded adaBoost classifier based on Haar-like features [24]. For locating the eye-pair region, a simple and effective algorithm based on an improved active shape model (ASM) is applied. A set of potentially discriminative HOG features based on orientation-shifted histograms is extracted from the detected eye-pair regions, and finally fed into an NB classifier to predict the eye status. More details for each designed component in the detection framework are provided in the following subsections.



**Figure 1.** A functional block diagram of the proposed framework for driver drowsiness detection.

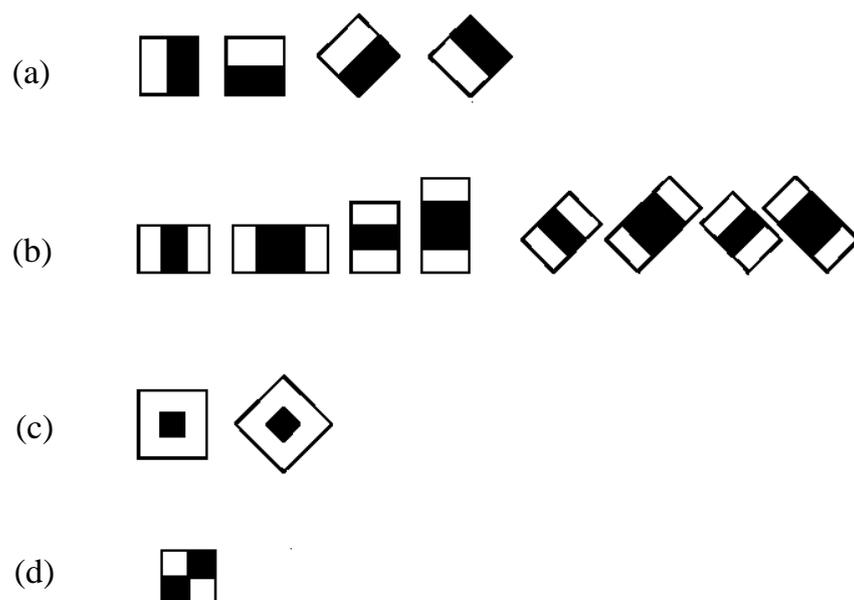
### 2.1. Image Preprocessing

Initially, the input driver image captured by an in-vehicle camera mounted on the dashboard is convolved with a 2D Gaussian blur filter over a  $3 \times 3$  pixel neighborhood and uniform standard deviation of 0.5 to get rid of (or suppress) disturbing noises and

unwanted background spots, while retaining spatially varying image structures. For light compensation, an adaptive contrast-limited histogram equalization algorithm [25,26] is then applied, with which each color channel is independently equalized to produce a better lighting-compensated image which further serves as input to the subsequent face-detection module. After the light compensation, the resolution of the light-compensated image can be reduced to increase the computational efficiency of the framework [27].

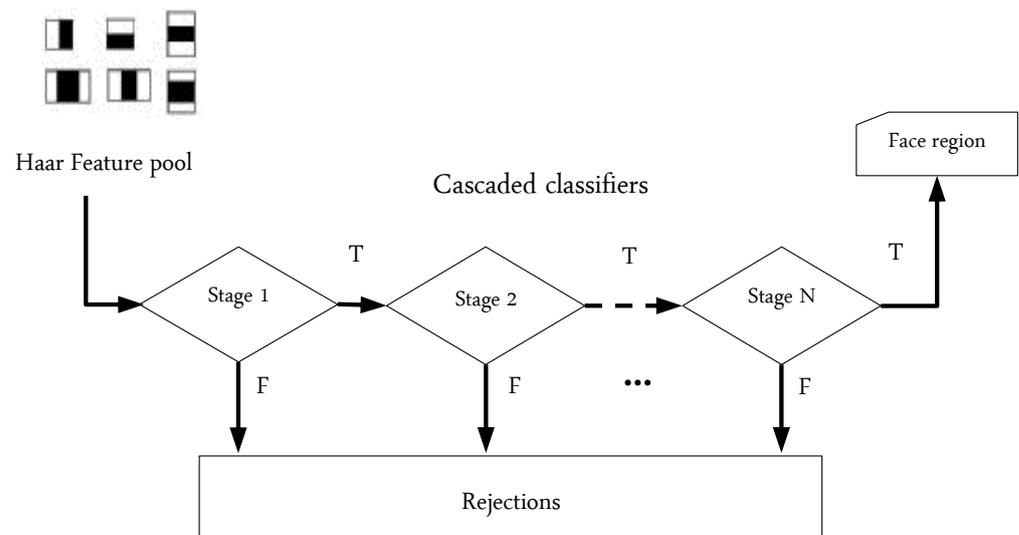
## 2.2. Eye Localization

As mentioned earlier, the first and very crucial step in developing and implementing an effective framework for drowsiness detection involves face detection and eye-pair localization (i.e., regions of interest—ROIs), which aims at locating the positions of the driver's eyes. In this work, a fast face detection algorithm was developed, where an improved cascaded adaBoost classifier based on an extended set of Haar-like features (see Figure 2) is utilized to automatically recognize the driver's face. In this algorithm, as a face can be located at any position and scale in the input image, the compensated image is first split into a number of rectangular regions. As it can be seen in Figure 2, the employed features are defined as different arrangements of bright regions and dark regions. In each case, the feature value corresponds to the difference between the sum of pixel intensities within the bright regions and the sum of pixel intensities within the dark regions. Due to the rapid training pace of improved Haar-like features, they have great potential for real-time face detection.



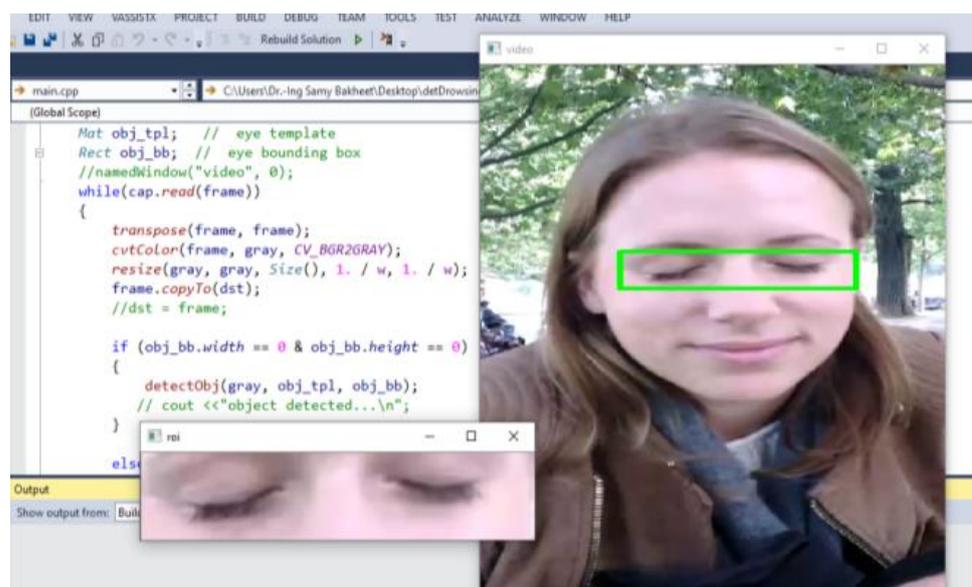
**Figure 2.** Extended Haar-like features: (a) edge features, (b) line features, (c) center-surrounded features and (d) a special diagonal line feature.

A cascaded adaBoost classifier basically is a strong (nonlinear) classifier built upon an ensemble of several weak (linear) classifiers; each is trained using the adaBoost algorithm. The face region is found when a candidate sample percolates through the cascaded adaBoost classifier. Nearly all face samples are allowed to pass through, while non-face samples are rejected. Waterfall-type classification using the adaBoost algorithm for face detection is shown in Figure 3.



**Figure 3.** The cascade structure of adaBoost classifier for face detection.

For detecting an eye-pair region, we use an improved active shape model (ASM) algorithm based on statistical learning models, which has the potential to extract relevant facial features rapidly and effectively. In this approach, the ultimate aim of active shapes is to match the model to a new image. To accomplish this goal, the ASM is trained on a set of specific points representing facial feature contours which has been manually labeled with facial feature interest points. Principal component analysis (PCA) is then performed to find out the main modes of variation in the training dataset. After establishing the ASM, the eye-pair regions in the face image are detected and localized, as shown in Figure 4. To reduce the difference between the model and the real contour, an iterative scheme using a cost function can be adopted to match models iteratively.



**Figure 4.** A sample snapshot of the resultant eye-pair region localization in our framework for driver drowsiness detection.

### 2.3. Feature Extraction

Due to their high descriptive power, robustness to illumination variation and simplicity and ease of implementation, HOG features originally initiated by N. Dalal and B. Triggs [28] have been extensively used (and still in use) in diverse domains of computer vision, such as face recognition, vehicle detection, video surveillance, image retrieval and disease diagnosis [29]. In this section, we show how to extract a modified variant of HOG features that resides in a relatively low-dimensional space and has significant discriminative power to properly characterize and quantify textures of eye and mouth regions for instantaneous driver drowsiness detection. In the HOG descriptor, the features are computed by taking orientation histograms of edge intensity in local eye regions. To achieve this objective, two fundamental computation units (i.e., cell and block) are locally defined. For each HOG feature, the block size is set to be  $2 \times 2$  cells each of size  $8 \times 8$  pixels, and blocks partially overlap—namely, each cell is covered by four blocks.

To extract the HOG features, the gradient and orientation values are first computed at each pixel location  $(x, y)$ , by the application of the 1D centered point discrete derivative mask with the filter kernel  $[-1, 0, 1]$ . For this goal, we initialize by calculating the magnitude  $\rho(x, y)$  and direction  $\gamma(x, y)$  for each pixel value as follows.

$$\begin{aligned}\rho &= \sqrt{L_x(x, y)^2 + L_y(x, y)^2} \\ \gamma &= \arctan\left(\frac{L_y(x, y)}{L_x(x, y)}\right)\end{aligned}\quad (1)$$

where  $L_x$  and  $L_y$  are the first-order Gaussian derivatives of the image patch luminance  $I$  in the  $x$  and  $y$  directions, respectively, which are computed at a scale of parameter  $\sigma$  (i.e., standard deviation) as follows.

$$L_{\xi} = I * \frac{\partial}{\partial \xi} \left( \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \right) \Big|_{\xi=x|y} \quad (2)$$

where  $*$  denotes 2D discrete convolution. The gradient magnitude  $\rho$  of each pixel in the cell is then voted into a specified number of angular bins (e.g., 8 bins) according to the orientation of the pixel's gradient. For every pixel in the orientation image, a histogram of orientations is built over a local spatial window (i.e., cell), such that the contribution of each pixel to an orientation bin is weighted by the gradient magnitude. More formally, the weight of each pixel, which is denoted by  $\alpha$ , is computed as follows:

$$\alpha = b + 0.5 - \frac{\gamma}{\pi} m \quad (3)$$

where  $b$  and  $m$  are the histogram bin to which  $\gamma$  belongs and the total number of bins in the histogram, respectively. To eliminate or reduce aliasing, we propose to update both values of two adjacent bins as follows:

$$\tilde{\gamma} = (1 - \alpha)\gamma, \hat{\gamma} = \alpha\gamma \quad (4)$$

Then, the weighted votes  $\tilde{\gamma}$  are accumulated into histogram bins over local spatial regions, so-called cells. The process to extract HOG features from a sample eye-pair image is shown in Figure 5.

In this work, we adopt an adaptive strategy to strengthen the description of HOG features, where similarities between image patches are utilized to capture the relatedness of local spatial regions. For this purpose, the orientation bins of the 8-bin HOG histogram created from a single cell are incrementally shifted by a factor  $\varepsilon$  ( $\varepsilon = 0, 1, \dots, 7$ ), resulting in a total of eight 8-bin histograms. Then, the binarized HOG feature quantities of two cell

regions  $c1$  and  $c2$  are directly computed by comparing the size relationships of the 8-bin orientation-shifted histograms:

$$b_{c1c2}(k, \epsilon) = \begin{cases} 1, & \text{if } v_{c1}(k) \geq v_{c2}((k + \epsilon)\%8). \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

where % denotes the modulo (division remainder) operator. It is worthy of pointing out that the extraction of HOG features based on binarized orientation-shifted histograms (see Figure 6) not only has potential to produce a more compact and robust version of HOG features, but also reduces greatly the computational time to a point compromising real-time execution. This, in turn, inevitably contributes greatly to faster and more accurate object detection.

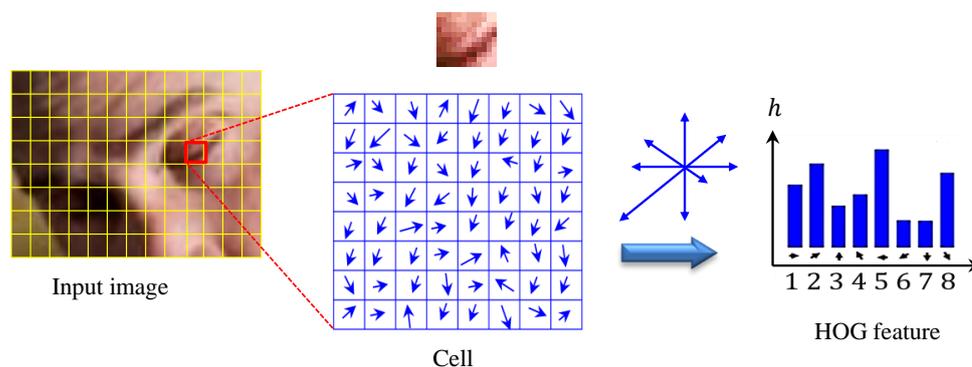


Figure 5. Extraction process of histogram of oriented gradient (HOG) features.

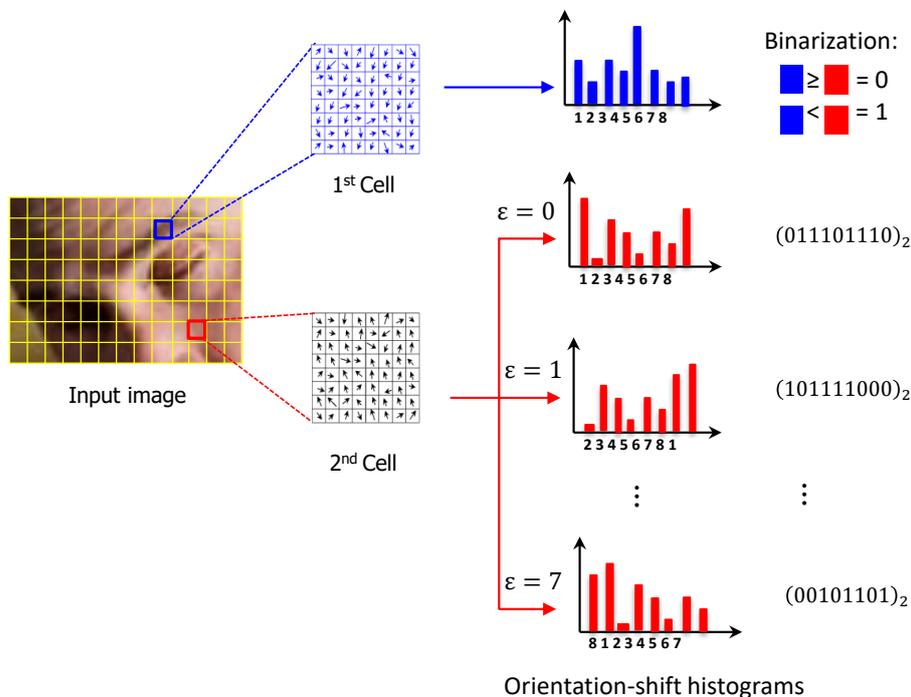
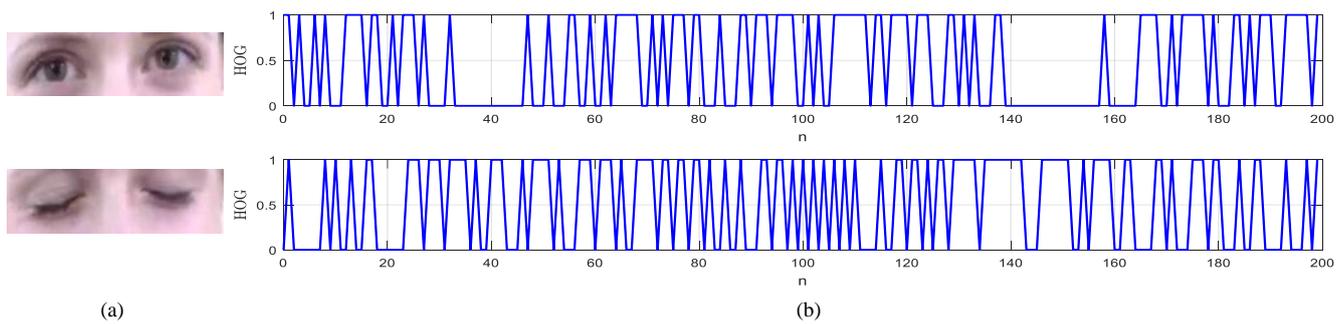


Figure 6. HOG feature extraction with a shift in the orientation.

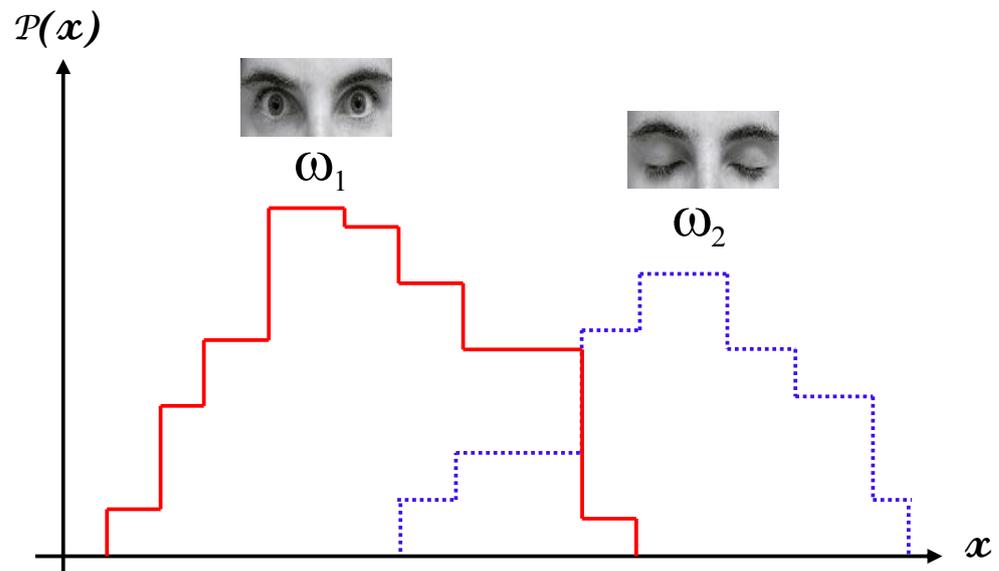
As an illustrative example, in Figure 7, we present 2D visualization plots for the developed HOG descriptor based on binarized orientation-shift HOG features extracted from two eye region snapshots.



**Figure 7.** 2D visualization plots for the improved HOG descriptor of the HOG features extracted from two eye region snapshots: (a) input eye region image and (b) binarized HOG descriptor based on orientation shift.

#### 2.4. Bayesian Feature Classification

In this section, we describe in detail the classification module based on an NB algorithm in the proposed system for instantaneous driver drowsiness detection. Strictly speaking, NB is a simple probabilistic model [30] based on applying Bayes' theorem [31] with strong independence assumptions among the attributes. Conventional probabilistic classifiers (e.g., NB) depend entirely on the conversion of data into probabilities for classification. As a representative example, in Figure 8, we show measurements of a given feature  $x$  for two classes  $\omega_1$  and  $\omega_2$ . As can be observed in the figure, the members belonging to the first class tend to have larger values than those of the second class; however, there is some degree of overlap between the two classes.



**Figure 8.** A histogram of feature values against their probability for two classes  $\omega_1$  and  $\omega_2$ .

It is visually obvious that at the two extremes of the range, it is a relatively easy task to predict the correct class for a given feature value, while carrying out the same task in or near the middle of the range is likely to be more challenging or even daunting. Formally speaking, let  $\mathcal{D}$  be a training dataset of pre-classified instances:

$$\mathcal{D} = \{(x, y) \in \mathbb{R}^n \times \{\omega_1, \dots, \omega_m\}\}$$

where  $x$  and  $y$  denote a feature vector of an input eye status and its true class label, respectively. In the current classification learning problem, the primary goal is to correctly assign the most probable of the available classes  $\omega = \{\omega_1, \dots, \omega_m\}$  for a given eye-status pattern represented by the feature vector  $x = (x_1, \dots, x_n)^\top$ , where  $x_i$  is the value of the  $i$ -th attribute.

For establishing optimal class labeling for unseen eye-status patterns, the maximum a posteriori (MAP) decision criterion is applied to achieve minimal misclassification rate:

$$\omega_{\text{MAP}}^* \equiv \arg \max_{\omega_j \in \omega} p(\omega_j | \mathbf{x}) \quad (6)$$

The MAP decision rule implies that the feature vector  $\mathbf{x}$  is assigned to class  $\omega^*$  where  $p(\omega^* | \mathbf{x}) > p(\omega_j | \mathbf{x})$ ,  $\omega_j \neq \omega^* \in \omega$ . To determine this class, estimates for the conditional probabilities  $p(\omega_j | \mathbf{x})$  are needed. In order to achieve this objective, we appeal to Bayes' theorem, which states that:

$$p(\omega | \mathbf{x}) = \frac{p(\mathbf{x} | \omega)p(\omega)}{p(\mathbf{x})} \quad (7)$$

where the above probabilities are defined as:

$p(\omega)$  : independent probability of  $\omega$  (i.e., prior probability);

$p(\mathbf{x})$  : independent probability of  $\mathbf{x}$  (i.e., evidence);

$p(\mathbf{x} | \omega)$  : conditional probability of  $\mathbf{x}$  given  $\omega$  (i.e., likelihood);

$p(\omega | \mathbf{x})$  : conditional probability of  $\omega$  given  $\mathbf{x}$  (i.e., posterior probability).

Upon applying the Bayes theorem, the MAP class given in Equation (6) can be computed as:

$$\begin{aligned} \omega_{\text{MAP}}^* &\equiv \arg \max_{\omega_j \in \omega} p(\omega_j | \mathbf{x}) \\ &= \arg \max_{\omega_j \in \omega} \frac{p(\mathbf{x} | \omega_j)p(\omega_j)}{p(\mathbf{x})} \\ &= \arg \max_{\omega_j \in \omega} p(\mathbf{x} | \omega_j)p(\omega_j) \end{aligned} \quad (8)$$

Notice that the quantity  $p(\mathbf{x})$  is omitted from Equation (8), since the data probability is constant (and independent of class) and thus can be safely excluded from calculations. Now all classes are assumed to be equally probable a priori (i.e., assuming a uniform prior);  $p(\omega_j) = p(\omega_k) \forall \omega_j, \omega_k \in \omega$ . Consequently, the calculations of the posterior distributions are greatly simplified:

$$\omega_{\text{ML}}^* = \arg \max_{\omega_j \in \omega} p(\mathbf{x} | \omega_j) \quad (9)$$

In this case, the so-called the maximum likelihood (ML) estimate that maximizes the likelihood of the training data is found. Recalling the MAP rule from Equation (8), we observe that Bayesian classification model depends both on the joint probability and the prior probability:

$$p(\omega | \mathbf{x}) \propto p(\mathbf{x} | \omega)p(\omega) = p(x_1, \dots, x_n | \omega)p(\omega) \quad (10)$$

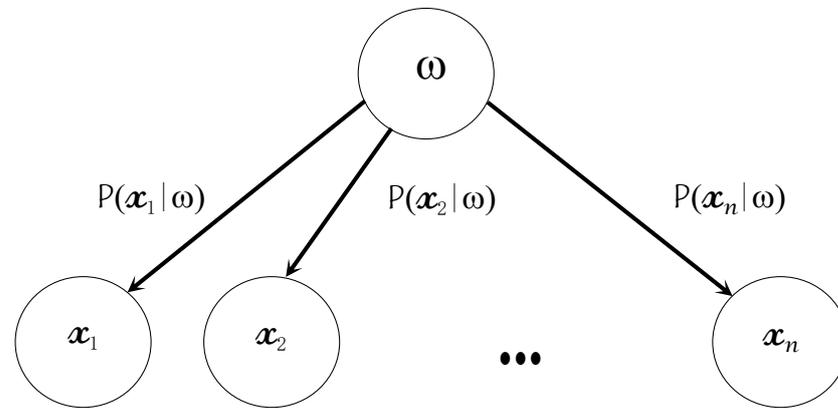
Intuitively, the inherent difficulty arising here is involved in learning the joint probability  $p(x_1, \dots, x_n | \omega)$ . To overcome this difficulty and make calculations increasingly tractable, we use the well-known "naïve Bayes independence assumption," which states that the probabilities of each attribute are conditionally independent of each other. This allows the joint probability to be conveniently written as a product of conditional probabilities:

$$\begin{aligned} p(x_1, \dots, x_n | \omega) &= p(x_1 | x_2, \dots, x_n; \omega)p(x_2, \dots, x_n | \omega) \\ &= \vdots \\ &\approx p(x_1 | \omega)p(x_2 | \omega) \dots p(x_n | \omega) \\ &= \prod_{i=1}^n p(x_i | \omega) \end{aligned} \quad (11)$$

At this point, it should be pointed out that even though the NB assumption is almost always violated in practice, NB learning is remarkably effective in practice. The substitution of the

joint probability from Equation (11) in Equation (8) generates the NB model (depicted in Figure 9),

$$\omega_{NB} = \arg \max_{\omega_j \in \omega} \left( p(\omega_j) \prod_{i=1}^n p(x_i | \omega_j) \right) \quad (12)$$



**Figure 9.** Naïve Bayes model with the assumption of conditional independence.

The above conditional probability term in Equation (9) can be estimated as relative frequency, simply by dividing each frequency  $n_c$  by the total number of opportunities. This approach can lead to significantly poor estimates, even when  $n_c$  is extremely small. To tackle such a challenging task, the conditional probabilities are estimated as follows:

$$\hat{P}(x_i | \omega_j) = \frac{n_c + m\alpha}{n + m} \quad (13)$$

where  $n_c$  and  $n$  are the number of instances for which  $\omega = \omega_j$  and  $x = x_i$ , and the total number of training instances for which  $\omega = \omega_j$ , respectively. The parameter  $\alpha$  denotes an a priori estimate (in calculations, it is set as  $\alpha = \frac{1}{t}$  for  $t$  possible values of  $x_i$ ) and  $m \geq 1$  is a weight that is specified a priori. With regard to the a priori probabilities, they can in principle be estimated by simply counting the proportion of classes in the training dataset:

$$\hat{P}(\omega_j) = \frac{\#(\omega_j)}{\sum_j \#(\omega_j)} \quad (14)$$

where  $\#$  denotes the frequency with which a certain class occurs within the available training data. Therefore, the MAP decision rule is equivalently written as:

$$\omega_{NB} = \arg \max_{\omega_j \in \omega} \left( \hat{P}(\omega_j) \prod_{i=1}^n \hat{P}(x_i | \omega_j) \right) \quad (15)$$

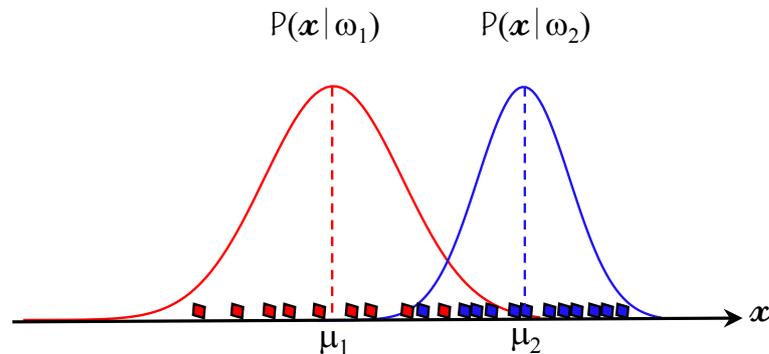
In cases of continuous-valued features, the class-conditional probabilities (likelihoods)  $p(x|\omega)$  are well-modeled by a Gaussian distribution  $\mathcal{N}(\mu, \sigma^2)$ :

$$\hat{P}(x_i | \omega_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{-\frac{1}{2} \left( \frac{x_i - \mu_{ij}}{\sigma_{ij}} \right)^2} \quad (16)$$

where  $\mu_{ij}$  and  $\sigma_{ij}^2$  are the mean and variance of the  $i$ -th feature of an eye-state example in the  $j$ -th class  $\omega_j$ , respectively (see Figure 10).

Based on the above analysis and derivations, the fundamental steps of the NB classification algorithm developed for automated driver drowsiness detection are outlined in Algorithm 1. In regard to the parameter dimensionality of the NB classification model,

given a set of  $N$  data points (or feature vectors of eye-status patterns) and a model with  $r$  parameters for the probabilities  $p(x_i)$ , it is not difficult to discern that the employed NB model will have only a set of  $mrN + (m - 1)$  parameters, where  $m$  is the number of classes.



**Figure 10.** Class-conditional probability distributions of features.

---

**Algorithm 1:** Naïve Bayes (NB) classification algorithm.

---

```

1. NB_Learn (eye_status_samples):
  foreach a target value  $\omega_j$  of an eye-status instance do
     $\hat{P}(\omega_j) \leftarrow$  estimate  $p(\omega_j)$ ;
    foreach feature value  $x_i$  of a feature  $x$  do
       $\hat{P}(x_i|\omega_j) \leftarrow$  estimate  $p(x_i|\omega_j)$ ;
2. classify_new_eye_status( $z$ )
    $\omega_{NB} = \arg \max_{\omega_j \in \omega} \hat{P}(\omega_j) \prod_{z_i \in z} \hat{P}(z_i|\omega_j)$ 

```

---

### 3. Experimental Results

In this section, various experimental results are presented and discussed, which are aimed at verifying the superiority of the proposed framework for driver drowsiness detection over competing state-of-the-art approaches in the literature. According to previously reported approaches [32–34], there are very few public datasets currently available for comprehensive performance evaluations of different approaches for driver drowsiness detection, particularly those with driver attention information from real-world driving scenarios [35]. On the other hand, it is especially difficult and most dangerous to build a realistic dataset for driver drowsiness detection in real situations that can be used to train the proposed framework comprehensively. For this reason, to verify the effectiveness of the proposed framework for driver drowsiness detection, we conducted extensive experiments on a public dataset, namely, the NTHU Driver Drowsiness Detection (NTHU-DDD) video dataset that is the only publicly-available dataset offering annotations for drowsiness, head, eye-pair and mouth status.

The academic NTHU-DDD dataset [36] collected by NTHU Computer Vision Lab at National Tsing Hua University was first introduced during the 2016 Asian conference on computer vision (ACCV) on driver drowsiness detection from video. In the dataset collection, the video streams were acquired by a high-speed camera under active infrared (IR) illumination, at a spatial resolution of  $640 \times 480$  pixels in AVI format. The total duration of video streams in the entire dataset was almost nine and a half hours. The dataset consists of a total of 36 subjects of various ethnicities who were recorded twice (with and without glasses/sunglasses) under a wide range of challenging simulated driving scenarios, such as normal driving, slow blink rate, yawning, falling asleep and bursting out laughing, under both daytime and nighttime illumination conditions. During video recording, the recruited subjects were asked to sit on a car chair and play a racing game

with a simulated driving wheel and pedals; in the meantime they were also asked to perform certain facial expressions. All video sequences were captured in a simulated environment under five different scenarios, namely, "BareFace," "Glasses," "Sunglasses," "Night-BareFace," and "Night-Glasses." The sequences of the first three scenarios have a frame rate of 30 fps, while those of other scenarios have 15 fps. In Figure 11, sample snapshots of NTHU drowsy driver detection (NTHU-DDD) dataset are shown.



**Figure 11.** Sample snapshots of the public NTHU-DDD dataset [36].

In our experiments, the entire NTHU-DDD dataset was split into a training set and an independent test set for the evaluation of the proposed framework. The training set consisted of 356 video samples from 18 subjects, while the test set contained 20 video samples from four subjects. We further divided the training set into videos from four subjects (for validation) and those from the remaining 14 subjects (for training). All selected test videos were resampled to be 15 fps to ensure statistical consistency between training and test data.

In order to quantitatively evaluate the drowsiness detection of the proposed framework, accuracy and F1-score (i.e., the harmonic mean of precision and recall) were calculated for each simulated driving scenario, where precision and recall are defined as follows:

$$\begin{aligned}
 \textit{precision} &= \frac{TP}{TP + FP} \\
 \textit{recall} &= \frac{TP}{TP + FN} \\
 \textit{F1 - score} &= 2 \times \frac{\textit{precision} \times \textit{recall}}{\textit{precision} + \textit{recall}}
 \end{aligned} \tag{17}$$

where TP (true positive) is the number of the correct drowsiness predictions, FP (false positive) is the number of incorrect drowsiness predictions (type I errors), TN (true negative) is the number of correct non-drowsiness predictions and FN (false negative) is the number of incorrect non-drowsiness predictions (type II errors). Detailed detection results of the proposed framework in terms of accuracy and F1-score for each simulated driving scenario in the NTHU-DDD dataset are provided in Table 1.

**Table 1.** Detailed results of the detection performance of the proposed framework on the public NTHU-DDD dataset.

Scenario	Drowsiness F1-Score (%)	Non-Drowsiness F1-Score(%)	HOG AC (%)	advHOG AC (%)
Bareface	90.99	87.19	86.21	89.35
Glasses	85.07	72.10	78.36	80.31
Sunglasses	81.74	67.14	75.69	76.30
Night-BareFace	93.49	85.37	87.64	90.64
Night-Glasses	87.93	93.68	88.24	91.48
Average	87.84	81.09	83.19	85.62

From the results shown in the above table, the following interesting observations can be made. First, perhaps the most remarkable fact emerging from the table is that the proposed framework achieved an average accuracy of 85.62% for drowsiness detection, which is very encouraging and agrees well with prior results reported in the literature. Furthermore, in light of these results, one can argue that a high level of accuracy in the drowsiness detection task along with significantly low computational costs would greatly contribute to the feasibility and robustness of the proposed framework for real-time traffic monitoring. Additionally, in order to assess the competitive performance of the proposed approach, we provide an experimental performance comparison of the framework against several state-of-the-art methods [17,37–40] in terms of accuracy of drowsiness detection. Table 2 provides a summary of this comparison. In light of the comparison, it is contended that the proposed framework exhibits superior results compared with other state-of-the-art approaches, while guaranteeing the deadline of the real-time traffic monitoring. In that vein, it is also worth mentioning that all the methods considered in the above comparison (Table 2) used the same dataset and almost similar experimental setups. Therefore, the comparison is apt to be most valid and revealing.

In closing, we thus conclude that the experimental results have shown convincing evidence that the proposed system has the potential to improve the performance of driver drowsiness detection systems, without losing the real-time guarantee; the system is capable of maintaining real-time processing of 24 fps. As a final concluding point, it is also worthwhile to mention that all the algorithms in this framework were implemented in Microsoft Visual Studio 2015 with OpenCV vision library version 4.1.2 for the graphical processing functions. All experiments (including tests and evaluations) were conducted on a PC with an Intel(R) Core(TM) i7 CPU-3.07 GHz processor, 8GB RAM, running Windows 10 Professional 64-bit operating system. As it might be expected, the achieved results demonstrate that the presented system can operate reliably and efficiently, achieving real-time performance operation for video sequences, due to the use of highly efficient algorithmic implementations in OpenCV library in combination with custom C++ functions.

**Table 2.** Quantitative comparison with other recent state-of-the-arts on the NTHU dataset

Method	Accuracy (%)
Proposed Method	85.62
CNN-LSTM [41]	75.80
Scale-Pruned 3D-CNN [39]	78.48
seqMT-DMF [37]	83.44
MSTN [38]	82.61
Joint-Shape RF [42]	88.18
3D-DCNN [18]	87.46
Human [17]	80.83

#### 4. Discussion and Conclusions

This paper has proposed an effective framework for instantaneous driver drowsiness detection, using an adaptive variant of HOG features for eye region representation and an NB model for classification. Quantitative evaluations on the publicly-available NTHU-DDD dataset have consistently demonstrated not only the significant superiority of the proposed system over several recent state-of-the-arts, but also its ability to make real-time predictions about whether drivers are drowsy or fatigued. A possible limitation of the framework relates to the generalizability of drowsiness detectors that have been developed based on the NTHU dataset, since video footage obtained in a moving car, in realistic traffic scenes, with sleepy drivers, is an entirely different thing than people in a laboratory who act sleepy or fatigued (as in the case of the NHTO dataset). We suspect that an additional limitation might arise from neglecting the effect of repetitive ambient change during driving and its effect of drowsiness detection rate in experiments. As prospects for future work, the intentions are twofold. On the one hand, we aim at enhancing the

validity and applicability of the presented methodology and extending implementations to integrate more diverse datasets to investigate the scalability of the proposed framework. On the other hand, we intend to extend the basic architecture of the framework for use in embedded boards or microcomputing systems to reduce operating and financial expenses and improve computational costs, without noteworthy performance degradation.

**Author Contributions:** Conceptualization, S.B.; methodology, S.B.; software, S.B.; validation, S.B.; formal analysis, S.B.; project administration, A.A.-H.; funding acquisition, A.A.-H. Both authors have read and agreed to the published version of the manuscript.

**Funding:** This work is funded by the Federal Ministry of Education and Research of Germany (BMBF) (Robo-Lab no 03zz04x02b, HuBa no. 03ZZ0470 367, RoboAssist no. 03ZZ0448L) within the Zwanzig20 Alliance 3Dsensation.

**Data Availability Statement:** Data sharing is not applicable to this article.

**Acknowledgments:** The financial support from the BMBF is gratefully acknowledged. We are also indebted to the anonymous referees for their insightful comments and valuable suggestions that have enriched the quality of the present work.

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

## References

1. Singh, S. *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*; Technical Report; (Traffic Safety Facts Crash•Stats. Report No. DOT HS 812 115); National Highway Traffic Safety Administration: Washington, DC, USA, 2015.
2. National Center for Statistics and Analysis. *Drowsy Driving 2015*; (Crash•Stats Brief Statistical Summary. Report No. DOT HS 812 446); National Highway Traffic Safety Administration: Washington, DC, USA, 2017.
3. Williamson, A.; Lombardi, D.A.; Folkard, S.; Stutts, J.; Courtney, T.K.; Connor, J.L. The link between fatigue and safety. *Accid. Anal. Prev.* **2011**, *43*, 498–515. [[CrossRef](#)]
4. Stevenson, M.R.; Elkington, J.; Sharwood, L.; Meuleners, L.; Ivers, R.; Boufous, S.; Williamson, A.; Haworth, N.; Quinlan, M.; Grunstein, R.; et al. The role of sleepiness, sleep disorders, and the work environment on heavy-vehicle crashes in 2 Australian States. *Am. J. Epidemiol.* **2014**, *179*, 594–601. [[CrossRef](#)]
5. Li, Z.; Li, S.E.; Li, R.; Cheng, B.; Shi, J. Online Detection of Driver Fatigue Using Steering Wheel Angles for Real Driving Conditions. *Sensors* **2017**, *17*, 495. [[CrossRef](#)] [[PubMed](#)]
6. Fletcher, A.; McCulloch, K.; Baulk, S.D.; Dawson, D. Countermeasures to driver fatigue: A review of public awareness campaigns and legal approaches. *Aust. N. Z. J. Public Health* **2005**, *29*, 471–476. [[CrossRef](#)]
7. Williamson, A.; Friswell, R.; Olivier, J.; Grzebieta, R. Are drivers aware of sleepiness and increasing crash risk while driving? *Accid. Anal. Prev.* **2014**, *70*, 225–234. [[CrossRef](#)] [[PubMed](#)]
8. Blommer, M.; Curry, R.; Kozak, K.; Greenberg, J.; Artz, B. Implementation of controlled lane departures and analysis of simulator sickness for a drowsy driver study. In Proceedings of the 2006 Driving Simulation Conference Europe, Paris, France, 4–6 October 2006.
9. Lawoyin, S. Novel Technologies for the Detection and Mitigation of Drowsy Driving. Ph.D. Thesis, Virginia Commonwealth University, Richmond, VA, USA, 2014.
10. Sayed, R.; Eskandarian, A. Unobtrusive drowsiness detection by neural network learning of driver steering. *Proc. Inst. Mech. Eng. J. Automob. Eng.* **2001**, *215*, 969–975. [[CrossRef](#)]
11. Tsuchida, A.; Bhuiyan, M.; Oguri, K. Estimation of drivers drive drowsiness level using a neural network based ‘Error Correcting Output Coding’ method. In Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems, Madeira, Portugal, 19–22 September 2010; pp. 1887–1892.
12. LM, K.; Nguyen, H.; Lal, S. Early driver fatigue detection from electroencephalography signals using artificial neural networks. In Proceedings of the International IEEE Conference of the Engineering in Medicine and Biology Society, New York, NY, USA, 30 August–3 September 2006; pp. 2187–2190.
13. Bakheet, S.; Al-Hamadi, A. Chord-length shape features for license plate character recognition. *J. Russ. Laser Res.* **2020**, *41*, 156–170. [[CrossRef](#)]
14. Abdelwahab, H.T.; Abdel-Aty, M.A. Artificial neural networks and logit models for traffic safety analysis of toll plazas. *Transp. Res. Rec.* **2002**, *1784*, 115–125. [[CrossRef](#)]
15. Wijnands, J.; Thompson, J.; Aschwanden, G.; Stevenson, M. Identifying behavioural change among drivers using long short-term memory recurrent neural networks. *Transp. Res. Part F Traffic Psychol. Behav.* **2018**, *53*, 34–49. [[CrossRef](#)]
16. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *117*, 61–85. [[CrossRef](#)]

17. Park, S.; Pan, F.; Kang, S.; Yoo, C.D. Driver Drowsiness Detection System Based on Feature Representation Learning Using Various Deep Networks. In *Computer Vision—ACCV 2016 Workshops Part III*; Chen, C.S., Lu, J., Ma, K.K., Eds.; Springer: Taipei, Taiwan, 2017; pp. 154–164.
18. Huynh, X.; Park, S.; Kim, Y. Detection of driver drowsiness using 3D deep neural network and semi-supervised gradient boosting machine. In *Computer Vision—ACCV 2016 Workshops Part III*; Chen, C.S., Lu, J., Ma, K.K., Eds.; Springer: Taipei, Taiwan, 2017; pp. 134–145.
19. Jabbar, R.; Al-Khalifa, K.; Kharbeche, M.; Alhajjaseen, W.; Jafari, M.; Jiang, S. Real-time Driver Drowsiness Detection for Android Application Using Deep Neural Networks Techniques. *Procedia Comput. Sci.* **2018**, *130*, 400–407. [[CrossRef](#)]
20. Lenskiy, A.; Lee, J. Driver's eye blinking detection using novel color and texture segmentation algorithms. *Int. J. Control Autom. Syst.* **2012**, *10*, 317–327. [[CrossRef](#)]
21. Pauly, L.; Sankar, D. Detection of drowsiness based on HOG features and SVM classifiers. In Proceedings of the 2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, India, 20–22 November 2015; pp. 181–186.
22. Moujahid, A.; Dornaika, F.; Arganda-Carreras, I.; Reta, J. Efficient and compact face descriptor for driver drowsiness detection. *Expert Syst. Appl.* **2021**, *168*, 114334. [[CrossRef](#)]
23. Singh, A.; Chandewar, C.; Pattarkine, P. Driver Drowsiness Alert System with Effective Feature Extraction. *Int. J. Res. Emerg. Sci. Technol.* **2018**, *5*, 14–19.
24. Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple feature. In Proceedings of the 2001 IEEE Computer Society Conference on CVPR, Kauai, Hawaii, 8–14 December 2001; pp. 11–518.
25. Abdullah-Al-Wadud, M.; Kabir, M.H.; Dewan, M.; Chae, O. A dynamic histogram equalization for image contrast enhancement. *IEEE Trans. Consum. Electron.* **2007**, *53*, 593–600. [[CrossRef](#)]
26. Sadek, S.; Al-Hamadi, A.; Michaelis, B.; Sayed, U. Image Retrieval using Cubic spline Neural Networks. *Int. J. Video Image Process. Netw. Secur.* **2009**, *9*, 17–22.
27. Bakheet, S.; Al-Hamadi, A. Computer-Aided Diagnosis of Malignant Melanoma Using Gabor-Based Entropic Features and Multilevel Neural Networks. *Diagnostics* **2020**, *10*, 822. [[CrossRef](#)]
28. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IEEE Conference Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–26 June 2005; pp. 886–893.
29. Bakheet, S. An SVM Framework for Malignant Melanoma Detection Based on Optimized HOG Features. *Computation* **2017**, *5*, 4. [[CrossRef](#)]
30. Lewis, D.D. Naïve (Bayes) at forty: The independence assumption in information retrieval. In Proceedings of the 10th European Conference on Machine Learning (ECML-98), Chemnitz, Germany, 21–23 April 1998; pp. 4–15.
31. Bolstad, W.M. *Introduction to Bayesian Statistics*; Wiley & Sons: New York, NY, USA, 2004; pp. 55–105.
32. Khushaba, R.N.; Kodagoda, S.; Lal, S.; Dissanayake, G. Driver drowsiness classification using fuzzy wavelet-packet-based feature extraction algorithm. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 121–131. [[CrossRef](#)]
33. Takei, Y.; Furukawa, Y. Estimate of driver's fatigue through steering motion. In Proceedings of the 2005 IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, HI, USA, 10–12 October 2005; pp. 1765–1770.
34. Wakita, T.; Ozawa, K.; Miyajima, C.; Igarashi, K.; Itou, K.; Takeda, I.K.; Itakura, F. Driver identification using driving behavior signals. *IEICE Trans. Inf. Syst.* **2006**, *89*, 1188–1194. [[CrossRef](#)]
35. Ramzan, M.; Khan, H.U.; Awan, S.M.; Ismail, A.; Ilyas, M.; Mahmood, A. A Survey on State-of-the-Art Drowsiness Detection Techniques. *IEEE Access* **2019**, *7*, 61904–61919. [[CrossRef](#)]
36. Weng, C.H.; Lai, Y.H.; Lai, S.H. Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network. In Proceedings of the Asian Conference on Computer Vision Workshop on Driver Drowsiness Detection from Video, Taipei, Taiwan, 20–24 November 2016.
37. Celona, L.; Mammana, L.; Bianco, S.; Schettini, R. A Multi-Task CNN Framework for Driver Face Monitoring. In Proceedings of the 2018 IEEE 8th International Conference on Consumer Electronics-Berlin (ICCE-Berlin), Berlin, Germany, 2–5 September 2018; pp. 1–4.
38. Shih, T.H.; Hsu, C.T. MSTN: Multistage Spatial-Temporal Network for Driver Drowsiness Detection. In *Computer Vision—ACCV 2016 Workshops*; Springer International Publishing: Cham, Switzerland, 2017; pp. 146–153.
39. Yao, H.; Zhang, W.; Malhan, R.; Gryak, J.; Najarian, A.K. Filter-Pruned 3D Convolutional Neural Network for Drowsiness Detection. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 17–21 July 2018; pp. 1258–1262.
40. Yu, J.; Park, S.; Lee, S.; Jeon, M. Representation Learning, Scene Understanding, and Feature Fusion for Drowsiness Detection. In *Computer Vision—ACCV 2016 Workshops*; Springer International Publishing: Cham, Switzerland, 2017; pp. 165–177.
41. Chen, S.; Wang, Z.; Chen, W. Driver Drowsiness Estimation Based on Factorized Bilinear Feature Fusion and a Long-Short-Term Recurrent Convolutional Network. *Information* **2021**, *12*, 3. [[CrossRef](#)]
42. Lyu, J.; Zhang, H.; Yuan, Z. Joint Shape and Local Appearance Features for Real-Time Driver Drowsiness Detection. In Proceedings of the Asian Conference on Computer Vision, Taipei, Taiwan, 20–24 November 2017; pp. 178–194.