

# State-of-the-Art Analysis of Modern Drowsiness Detection Algorithms Based on Computer Vision

Fudail Hasan  
ITMO University  
St. Petersburg, Russia  
fudail225@gmail.com

Alexey Kashevnik  
SPC RAS  
St. Petersburg, Russia  
alexey.kashevnik@iias.spb.su

**Abstract**—The paper presents a state-of-the-art analysis of modern drowsiness detection algorithms based on computer vision technologies as well as considers the problem of yawning detection for the vehicle driver. Based on the analysis of the literature we classify drowsiness detection techniques into three groups: the driving pattern of the vehicle; psychophysiological characteristics of drivers; and computer vision techniques for driver monitoring. The computer vision methods look most promising since they are non-intrusive for the driver. The importance of the driver drowsiness monitoring system rises from the number of drowsiness-related accidents. Yawning is an important identifier of drowsiness, even though it is not the most reliable drowsiness indicator. Some of the methods that are based on computer vision are presented and discussed in the paper. We developed and evaluated a yawning detection model. We analyzed available datasets for yawning detection and conclude that the existing datasets have to be enhanced by pictures taken in real driving conditions. We propose yawning detection dataset-preparation as well as detection model development and evaluation.

## I. INTRODUCTION

Road accidents are a worldwide problem that causes fatalities and economic loss. National Highway Traffic Safety Administration (NHTSA) in the USA estimated that in 2017, about 91,000 police-reported crashes involved drowsy drivers [1]. So, drowsiness monitoring is important to reduce the accident rate on public roads. The drowsy driver has different characteristics that appear on his body, behavior, and driving style. To identify the driver's alertness, various parameters are measured using different techniques. These techniques could be classified into 3 main categories [2]:

- Driving pattern of the vehicle;
- Psychophysiological characteristics of drivers;
- Computer vision techniques for driver monitoring.

The importance of driver monitoring and alerting system increases in long-time journeys. The probability of a drowsiness-related event (yawning, ...) occurring increases as the driving time increases [3]. In contrast, by increasing the number of stops during the journey, the probability of a drowsiness-related event occurring decreases. The number of fatigue/drowsiness events decreases significantly when drowsiness feedback is provided to the driver [4]. So, in this case, driver monitoring systems look like a useful solution to reduce the number of accidents on public roads. Such systems

should detect the drowsy driver as well as notify him about this situation.

Drive Safely system is aimed at dangerous states' detection based on the camera and the sensors of the smartphone [5], [6]. For this system, drowsiness is one of the dangerous states that should be detected in the vehicle cabin. Initially, we used the Dlib framework to determine facial landmarks. Based on landmark points we detected the opened or closed eyes and opened or closed mouth. Based on this information we detect drowsiness dangerous state. However, when we collected enough data of drivers' faces, we decided to train models that detect eye closeness as well as yawning detection instead of using Dlib framework. Such models are more reasonable than Dlib since it provides better accuracy as well as less computational time. The paper proposes a three-step method for yawning detection. The first step is to capture the frame from the camera, the second step is to use a face-detector with high accuracy to detect the face of the driver within the frame, and the third step is to classify the sub-image around the face (yawny face or not). The paper also contains a description of the available yawning detection datasets as well as our own dataset that we collected. Using both, online available dataset and our dataset, we trained the yawning detection model which is a modified version of the "MobileNetV2" classification model. The modified model has a lower accuracy than the original model but it is smaller and faster. In general, the new method increased the accuracy of drowsiness detection. So, the contribution of the paper is both:

- Exploring and discussing modern approaches to driver monitoring for drowsiness detection;
- Proposing a yawning detection method that consists of three steps;
- Yawning datasets analysis and preparing our dataset.
- Using a modified "MobileNetV2" model to classify the images of the faces.

The structure of the paper is as follows. In section II we present the drowsiness detection techniques and sources of parameters used for drowsiness detection. Approaches and methods for drowsiness detection are presented in section III. Section IV presents a comparison between these methods from an analytical view. Section V presents the method of yawning detection and the Deep Neural Network (DNN) model used for image classification. Evaluation of the model is shown in section VI. Section VII concludes the paper.

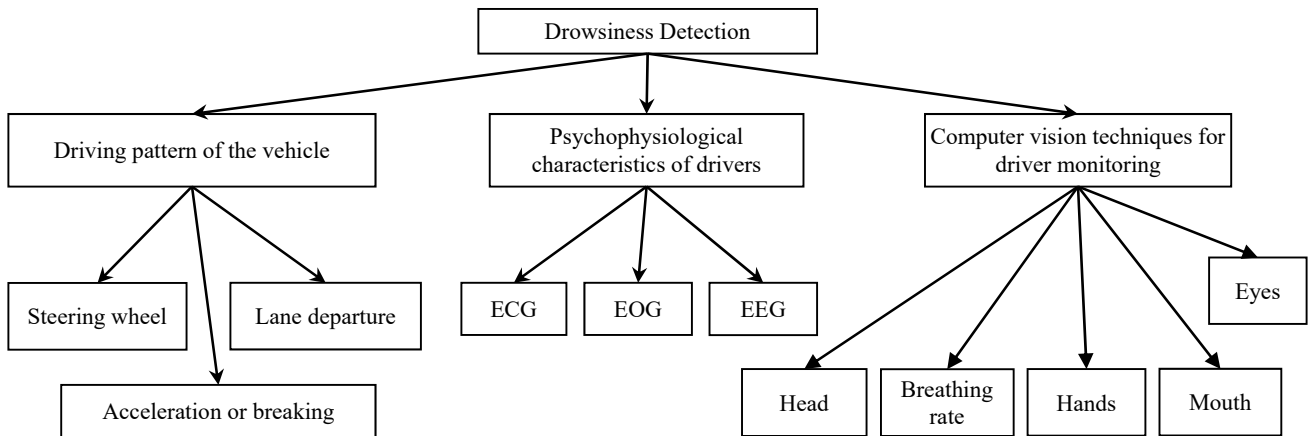


Fig. 1. Drowsiness detection techniques and parameter sources

II. DROWSINESS DETECTION TECHNIQUES

We propose a classification of drowsiness detection techniques that are applicable to detect such dangerous situation for a human driver in the vehicle cabin. Fig. 1 shows the general classification of such techniques. The first group of techniques uses vehicle-based measurements. Different movement parameters are used to identify drowsy driving. Previous studies used various parameters of the vehicle, such as Lane tracking, steering wheel angle, and velocity [7], [8], lateral acceleration, and yaw rate [7] for driver drowsiness detection. The parameters used in this group could be classified as: steering wheel, acceleration or braking, and lane departure. The advantage of the vehicle-based measures is that they can be easily acquired; however, these measures are confounded by the road state and surrounding traffic [9].

Psychophysiological-based measurements are based on the fact that physiological signals could allow for earlier drowsiness detection, as long as they change in the early stage of drowsiness [9]. One of the most reliable sources to detect drowsiness is the electroencephalogram (EEG) [10]. Also, electrocardiogram (ECG) and electrooculogram (EOG) are used as a source for earlier drowsiness detection [11, 12]. Psychophysiological-based techniques are the most accurate way for drowsiness detection. However, they are not used widely for driver monitoring since they are intrusive techniques and need special equipment to be worn on the body.

In the third group of techniques, which is computer vision techniques, the parameters that are used for driver drowsiness detection are more than the parameters in the above two techniques. Computer vision techniques can monitor driver state using eyes state, PERCLOS, blinking rate, blinking duration, blinking duration changes over time, yawning, face expressions, head angle, gaze orientation, and breathing rate. The most significant parameter source is the eyes because of the strong effect of drowsiness on the eyes. Parameter sources such as mouth, head, and breathing rate are less used. The parameters of hand such as position and movement speed could be used as indicators of alertness.

The computer vision-based techniques are more popular because they are non-intrusive and aren't confounded by the state of the road. Besides, data could be calculated easily using a camera pined in front of the driver. These techniques face some

difficulties such as light conditions and individual differences in face color and shape; however, these difficulties have been overcome, and most computer vision methods are robust to these difficulties.

III. COMPUTER VISION BASED METHODS

Let's consider a modern computer vision based method for drowsiness detection. A two-steps method is proposed in [13]: the first step is for eye detection by detecting the eye center, and the second step is for eye state estimation. The proposed method uses a specific filter in the Fourier plane of the Vander Lugt Correlator (VLC) for estimating eyes location and state.

Methods that use eye states and blinking to identify drowsy drivers are proposed in the papers [14-19]. A mobile-based system for drowsiness detection is proposed in [14]. The system extracts facial landmark points to detect closed eyes using eye aspect ratio (EAR). The EAR of the opened eye is calculated as the average of EAR measurements that are in the range 0.01 – 0.03. Drowsiness detection depends on a fact that eye closing takes 100 - 400ms in normal condition while it takes about 500ms in drowsy one [14].

The Haar cascade classifier is used for face detection, and face tracking is used to increase the accuracy of the classifier [15]. The authors used the template matching method to identify both left and right eyes, estimate their positions, and estimate their state (whether opened or closed). By estimating eyes positions they were able to calculate head tilting using geometrical equations. The authors proposed two ways to estimate the drowsiness level. The first way is by measuring blinking duration, and calculating blinking rate, taking into account the situation of eyes full closure (closing eyes more than two seconds). The second way is by using the head angle when the driver tilts his head, considering the driver drowsy if the head angle more than thirty degrees.

The use of EAR patterns for blinking and drowsiness detection is proposed in [16]. The authors used the Dlib library for facial landmarks estimation. Facial landmark points were used for EAR calculation. Using a machine learning model with an input of 15 consecutive EAR, the eye state is classified (open eye, short blink, or long blink). The model was trained for videos of different FPS (15 and 30 frames per second) to make the model robust to changes in FPS. The designed system works with an FPS of around 23 frames/second.

Another method is using blink patterns [17]. The authors used Haar cascade classifier [20] that is available in the OpenCV library for face detection, then they used an ML-based eye detector for locating the position of eye pupils, and created a region of interest (ROI) as a rectangle that contains the eye and the eye pupil as its center. After defining the ROI the authors used the horizontal symmetry property of the ROI for defining eye state (open or close). The driver whose eyes are closed for more than 800ms is defined as a sleepy driver, while the driver whose eyes are closed more than 400ms and less than 800ms is defined as a drowsy driver.

A method that uses blinking duration changes for estimating the drowsiness level is proposed in [18]. For calculating the blinking duration, the authors used the vertical distance between the eyelids. The authors took the measurements of blinking duration over distance and calculated the regression line for these measurements. Drowsiness level is estimated using regression line slope.

A method that uses adaptive boosting (AdaBoost) for face detection and an adaptive template matching method for face tracking is proposed in [19]. Within the area of the detected face, AdaBoost and blob detection are used to locate the regions of the eyes. The regions of the eyes are used for eye classification (open or closed) and estimating PERCLOS. The drowsiness is determined by a 2D Gaussian mixture model which takes PERCLOS and ECD (eye close duration) of the driver's normal state driving images as two inputs.

Machine learning model-based method for classifying driver state (drowsy or not) consists of two steps [21]. The first step is to extract facial landmark points using the Dlib library (68 points positions (x,y)). The second step is to use Multilayer Perceptron Classifier for the classification of the driver state. The estimated positions of the facial landmarks that are calculated in the first step, act as input for the classifier.

Another machine learning model-based method for classifying the driver's state is proposed in [2]. Although the two methods [21] and [2] depend on a machine learning model of two steps for drowsiness detection, the two methods are completely different. In the paper [2], the first step uses multi-task cascaded convolutional networks (MTCNN) for obtaining face boundary coordinates and five landmark points containing locations of left-eye, right-eye, nose, left-lip-end, and right-lip-end. Using the outputs of the first step, four sub-images containing face, mouth, left eye, and right eye are created. The second step uses the four sub-images as input for a neural network that classifies the driver state as normal, yawning, or drowsy.

The method proposed in [22] used image processing-based methods for eye detection and tracking. After locating eyes, PERCLOS and blinking duration are calculated. Also, a 3D facial orientation and position are estimated to detect head movement such as head tilts and to estimate gaze orientation. In addition to the previous parameters, some extra facial expressions around the eyes and mouth are detected. Using these parameters with the Bayesian Network model for fatigue, the authors were able to detect fatigue and drowsiness situations.

The method based on image processing detects skin profiles using skin color and texture [23]. After locating the face, the color difference between lips and face is used for mouth detection. Eyes are detected to verify that the mouth is detected correctly (using the geometrical relation between eyes and mouth in the human face). The yawn component is the largest non-skin area in the face (which is not the lips and in the mouth area). The area of the yawn component defines the driver state (yawning or not).

Another method that used yawning for drowsiness detection is proposed in [24]. The method is also used blinking duration as a factor for drowsiness detection. Blinking duration is calculated using eye state and yawning is detected using the mouth state. Initially, the face is detected and its position is estimated using the Viola-Jones face detector [25]. Within the region of the face, lips are searched for using spatial fuzzy C-means (s-FCM) clustering. In the upper part of the face, the pupils are detected. The driver state is classified using the information of eyes and mouth.

Yawning detection is also used for drowsiness classification in the paper [26]. The face region is detected using the Viola-Jones detector [25], while the mouth region detection is done within the face region using an image processing method. Yawning is defined as a continuous (more than 20 frames for the camera that captures 30 frames per second) large mouth-openness. The proposed method for calculating mouth openness degree is using the aspect ratio of the mouth bounding rectangle. The mouth-openness is defined as large when the ratio of mouth height over mouth width is bigger than 0.5. This method is robust to the head angle change and mouth scale difference in different frames.

Method for breathing rate estimation and using this estimation for drowsiness level estimation is proposed in [27]. Breathing monitoring in this method is done using a Kinect camera so it is a non-intrusive method. The system is designed to detect chest/abdomen movement by detecting their volume changes.

#### IV. COMPARISON

To compare the presented approaches, we propose to compare them through the following parameters (Table I):

- The calculated parameters and their sources: which represent which parts of the driver's body are monitored and what data is extracted from them;
- The technology used for data extraction: this allows us to know the accuracy, computation complexity, and restrictions of the method;
- Dataset used for data training or validating: the description of the dataset allows us to know the robustness of the method;
- Accuracy of the method: shows the method's success on the validating dataset;
- The restrictions of the method.

Most of the presented methods are using eye-related metrics for drowsiness detection because eye-related metrics are more reliable than yawning detection [35] and could be extracted using image processing methods.

TABLE I. DROWSINESS DETECTION APPROACHES

Approach	Source	Calculated Parameter	Technology	Dataset	Accuracy	Restriction
[13]	Eyes	Eyes state	Image processing	BioID [28], FEI [29], GI4E [30] and ICPR [31]	BioID: 99.90% FEI: 94.00% GI4E: 99.60% ICPR: 96.00%	Wearing sunglasses
[14]	Eyes	Blinking duration	Viola-Jones face detector	NO	92.85%	Wearing sunglasses
[15]	Eyes	Blinking duration, blinking rate, and head angle	Haar cascade classifier and template matching	NO	97.86%	Wearing sunglasses
[16]	Eyes	Blink patterns	Dlib library and machine learning	DROZY [32]	94.44%	Wearing sunglasses
[17]	Eyes	Eyes state	Haar cascade classifier and machine learning	JZU [33]	94%	Wearing sunglasses
[18]	Eyes	Blinking duration changes	Image processing	NO	-	Wearing sunglasses
[21]	Face	Face landmark points	Dlib library and machine learning	NTHU [34]	81%	-
[2]	Head, eyes, and mouth	Regions of face, eyes, and mouth	machine learning	-	94.8%	-
[19]	Eyes	PERCLOS and blinking duration	Template matching	NO	-	Wearing sunglasses
[22]	Head and eyes	PERCLOS, blinking duration, head tilt, and gaze orientation	Image processing	NO	-	-
[23]	Mouth	Yawning	Image processing	NO	-	Mouth covering, hand on mouth
[24]	Eyes and mouth	Blinking duration and yawning	Viola-Jones and image processing	NO	-	-
[26]	Mouth	Yawning	Viola-Jones and image processing	NO	-	-
[27]	Breathing	Breathing rate	Image processing	NO	-	Clothes

The reason of which yawning detection has less reliability is that gestures such as talking, singing, or laughing lead to wrong positives when using mouth openness as a yawning indicator. But with the development of machine learning science, more features could be used for these purposes.

However, taking eye state as the only parameter for drowsiness detection cannot cover all the situations of drowsiness [13]. In other words, it will detect only the situation of fully closing eyes, when the driver closes his/her eyes for a long period. This system is not useful for early drowsiness detection. One important fact when discussing the work in [13], is that the accuracy, presented in (Table I) is the accuracy of eye state estimation, not for drowsiness detection. The work is validated with 4 online-available datasets and 1 online-unavailable dataset, with an important fact, that all the datasets used for method validation do not contain images of people wearing sunglasses. This means that wearing sunglasses is one of the limitations of the work.

A robust real-time face detector called the "Viola-Jones" face detector is used for face detection in [14]. The accuracy of the Viola-Jones detector is up to 94.1%. One of the limitations of this detector is that the detector usually fails when the eyes are occluded. In other words, when the driver is wearing sunglasses, the system proposed in [14] is usually unable to detect driver drowsiness. The accuracy of the drowsiness detection system is 92.85% for testing data collected from 10 Indonesians.

The Haar cascade classifier [20], is used in [15] while did not use the Viola-Jones face detector that is proposed in 2004 [25]. Using four templates of eyes (left-opened, left-closed, right-opened, and right-closed), the template matching detects eyes' positions and their state. The calculation of the blinking duration is not accurate because it is calculated depending on the state of the eyes only. Besides that, blinking duration threshold estimation is based on 10 measurements on 10 young people from Thailand. The authors did not take into account that blinking duration may depend on age.

A machine learning-based model is built and trained in [16] to classify blinking patterns. The dataset used for model training is the "DROZY" dataset [32]. The input of the model is 15 consecutive EAR of the subject. The most important fact in blinking classification is the time factor, which is not entered as an input to the model. To solve this challenge, the authors trained the model for different FPS (15 and 30 FPS). The authors did not prove the robustness of the FPS changing as long as the model is trained and validated using the DROZY dataset.

Another work that uses the Haar cascade classifier is proposed in [17]. The differences between the work in [15] and the work in [17] are the methods of eye detection and eyes state estimation. While in the paper [15] an ML-based model is used for eye state estimation, in the paper [17] the horizontal symmetry property of the eye region is used to estimate its state. The accuracy of this method is 94% for the "DROZY"

dataset. One of the main factors that affect the performance of the method is the presence of glasses because it affects symmetry calculations [17].

The work proposed in [18] is not used widely. The authors used blinking duration changes as a factor for identifying the drowsiness level. Using the distance between eyelids is not robust to scale changing or to individual differences. The authors did not discuss two important factors that may affect their results: data collecting period and driver alerting. In case of a long-time journey, the data collected at the beginning of the journey will affect the drowsiness level estimation and in case the driver is alerted (by the system or by outer source) the authors did not discuss whether to start new data collection or how to reduce the effect of the old data.

In general, all methods that use eyes as the only parameter source have a common restriction, when the driver wears sunglasses, because eye features are covered. Even though the eye-related parameters are the most reliable parameters for drowsiness detection, it is not recommended to use only them to identify the drowsy driver.

The method proposed in [21] used the Dlib library for facial landmark points extraction and trained an ML-model for drowsiness detection using these points. The method has many strength points such as the usage of the “NTHU” dataset for model training which is consisting of videos for 22 subjects with different ethnicities. The accuracy achieved using this method is 81% with a model size of 100KB. The authors were limited by the model size because it is designed for the Android system, while the accuracy of such models could be increased by increasing model size.

ML-based models for face detection and drowsiness detection, are used in [2]. For face detection, MTCNN is used. The output of the MTCNN is the face boundaries and the location of the left eye, right eye, nose, left-lip-end, and right-lip-end. To reduce the effect of the MTCNN, the authors used the output of MTCNN and extracted four sub-images of face, mouth, and both eyes. Those sub-images are used to train the drowsiness detection model. Another strength point in this method is the dataset used for training. The dataset includes 33 subjects of different gender and ethnicities. 11 of these subjects are wearing glasses.

Two cameras with special infrared emission were used in the system proposed in [22]. The authors used image processing methods for eye detection. Then they depended on the geometrical calculations for estimating 3D face orientation and gaze orientation. The use of the Bayesian Network model allows the estimation of the fatigue probability with an explanation of the effect of each used parameter.

Unlike the method in [22] that used many factors for drowsiness detection, the method proposed in [23] used only yawning as a drowsiness identification. The method used only skin features for face and mouth (named as yawn component) detection. This technique has many restrictions that lead to false results such as mouth covering by non-skin material (mask, scarf ...) that leads to false positive, or putting a hand over the mouth while yawning leads to false negative.

Another method that used yawning for drowsiness detection is proposed in [24]. The method also used blinking duration as

a factor for drowsiness detection. The authors used Support Vector Machines (SVM) for analyzing yawning and eyes state. Even if the SVM decided that the driver is yawning, the authors decided that the eyes should be closed or half-opened to alarm the driver. The authors made this condition to save the system from generating a false alarm when the driver is talking [24]. This means that the SVM yawning classifier is not accurate.

In comparison with the method in [23], the method in [26] is using only yawning to detect the drowsy driver. However, the method in [26] is more robust to the situation when the mouth is covered by hand. The authors used two methods to determine mouth region boundaries, depending on the features of the lip region. The thresholds of the degree of the mouth-openness and yawning duration are chosen through experiments. The dataset used for system validation is not explained in [26], so it is not guaranteed that the threshold of the degree of the mouth-openness is suitable for different ethnicities.

An approach for drowsiness detection is discussed in [27]. The proposed method uses the involuntary movements of the driver provoked by respiration as an indicator of drowsiness. The authors used a Kinect camera for measuring the breathing rate. As long as, the measurements depend on the chest movements, the method has an advantage when the driver is using the seat belt. However, when the driver is not using a seat belt, the type of clothes will affect the accuracy of breathing rate measurements, especially with loose clothes.

## V. YAWNING DETECTION

Yawning recognition is one of the major parts of drowsiness detection. To recognize the yawning, it is necessary to recognize mouth openness and some extra facial features. For this purpose, we used Dlib framework to extract facial landmark points and used these points to calculate the mouth openness and estimate the yawning situation. We have noticed that the accuracy of Dlib is low with the bad lighting condition as well as for rotated faces. For this reason, we decided to use a method that uses a face-detector with high accuracy to extract the face from the image and then to use a deep neural network (DNN) that can extract and recognize facial features, and classify the state of the driver (yawning or not). Many factors affect the effectiveness of the classification of the DNN model. In this work, we focused on the dataset used for training the model and the structure of the model.

### A. The proposed yawning detection method:

In the Drive Safely system, we use a smartphone's camera and sensors for monitoring the driver during the car-trips. We proposed a method to detect driver's yawning depending on the frame captured from the smartphone's camera. Fig. 2 shows the process of the method which can be split into three steps:

1. Frame capture. In this step, we capture the frame from the smartphone's camera.
2. Face's image extraction. After we capture the frame, we use the “FaceBoxes” face detector to detect and locate the region of the driver's face in the frame.

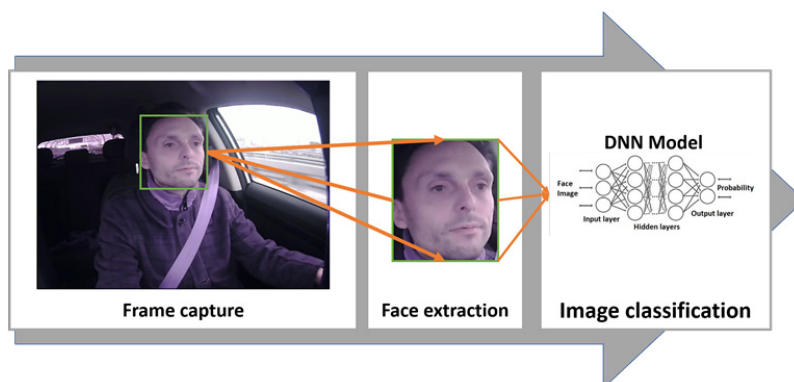


Fig. 2. Yawning detection method

We crop the region of the face from the frame and resize it to the size of (120\*120). At this point, the face image is ready.

3. Image classification. For yawning detection, we use a DNN model that classifies the face image into two classes (“Yawning” or “Not yawning”). The output of the model is two values which are the probability of belonging to each class. We classify the image to the class with the highest probability.

### B. Yawning Dataset:

For compatibility with our method, the dataset should contain faces’ images of drivers. These images should be prepared in the same way we prepare the face’s image during the yawning detection. To guarantee the effectiveness of the dataset, it is recommended that the images are for a non-small set of drivers, and are taken in different illumination conditions with real facial expressions.

There is an online-available dataset used for yawning detection. This dataset is called “YawDD” (Yawning Detection Dataset) [36]. YawDD is a dataset of videos, recorded by an in-car camera, of 107 drivers in a stopped car with various facial characteristics (male and female, with and without glasses/sunglasses, different ethnicities) while they are talking, singing, being silent, and yawning. The videos are taken in natural and varying illumination conditions during the daytime only. The videos are taken from two different camera positions (under the front mirror of the car and on the driver’s dash). Videos in the dataset are classified into three classes:

1. Videos that contain a normal situation.
2. Videos that contain talking/singing.
3. Videos that contain yawning.

One disadvantage of this dataset is that the videos that contain talking and yawning also contain a normal situation (yawning is only a part of the video). So to make a dataset of images an extra manual work is required; multiple frames could be taken out from these videos, but these frames should be manually classified to make a dataset that contains classified images. Other disadvantages of this dataset are: videos are recorded while the car is stopped, no videos are

recorded during nighttime, and the yawning is forced which means abnormal facial expressions.

The previously remarked disadvantages lead to the need for another dataset that fulfills the following conditions:

- Videos are recorded during day and night;
- Videos are recorded during real trips while the car is moving and the driver’s facial expressions are spontaneous.

We decided to prepare our yawning detection dataset. Dataset preparation is done in four steps: (1) video recording, (2) frame extraction, (3) face image extraction, and (4) manual classification.

1. Videos recording: The videos are recorded during real driving journeys by an infrared camera with a resolution (640\*480). The number of subjects is 10 (7 males and 3 females). In contrast to YawDD, the videos are recorded under different light conditions (day and night) and during real car trips.
2. Frame extraction: Frames are extracted from the recorded videos using python code that extracts 1 frame every 1 second.
3. Face image extraction: Using the “FaceBoxes” Face detector, faces are detected and the region of the driver’s face is located in the frame. The region of the face is cropped from the frame, resized to the size of (120\*120), and saved.
4. Manual classification: Images are classified manually into 2 classes: (1) “not yawning” (for normal situations including smoking and mouth opening), (2) “yawning”.

The main disadvantage of our dataset is the low number of subjects which means a low difference in facial characteristics between the subjects. So, we decided to use both the “YawDD” dataset and our dataset for training the DNN model. Fig. 3 shows an example of images from the dataset.

The dataset used to train the DNN model consists of:

- 4118 “not yawning” images. 706 images are from YawDD and 3412 images are from our dataset;
- 4068 “yawning” images. 706 images are from YawDD and 3362 images are from our dataset.



Fig. 3. Classified images from the dataset

C. Model's structure:

The developed model has the structure of DNN. Its structure is inspired by the structure of the "MobileNetV2" model [37]. In other words, our model is a modified "MobileNetV2" model. The "MobileNetV2" model is a general-purpose classification model. As long as our problem is a special classification model where the face fulfills the classified image, we decided to modify the structure of "MobileNetV2". The modification was to take the first 5 blocks of the original model and make a sequential model of these 5 blocks. Fig. 4 shows the structure of the model and the layers of each block.

The structure of our model has fewer layers (42 layers including input and output layers) than "MobileNetV2" to reduce the model size and execution time. The developed model consists of:

- Input layer. The shape of the input layer is (120\*120\*3); in other words, the input of the model is a colored image of size (120\*120).
- Hidden layers. The model contains 40 hidden layers.
- Output layer. The output layer contains two values which are the probability of belonging to each class.

VI. EXPERIMENTS AND RESULTS

To evaluate the results of our model we compared the results got with the results of other classification models that were published before and built in the Keras library (a part of the "TensorFlow" library).

Among the classification models we chose "MobileNet", "MobileNetV2", and "EfficientNetB0" models. In addition to our model, we trained those three models using the dataset described above and compare the results of the four models with three parameters:

- Model size. The model size affects the size of the system and the amount of RAM required for the system. The model size is measured in the format of ".keras" model for comparison between models;
- Executing time. The average executing time is measured on GPU "NVIDIA Tesla K80" (Google Colab GPU);
- Accuracy. The ratio of the correct classifications number to the number of images.

By comparing the results that are presented in (Table II), we noticed that the model of best accuracy is the "MobileNet" model with an accuracy of 96.8% while our model achieved an accuracy of 95.2%. On the other hand, our model has the fastest execution time of 10ms (at least two times faster than

the other models) and the smallest size of 772KB (at least 30 times smaller than the other models).

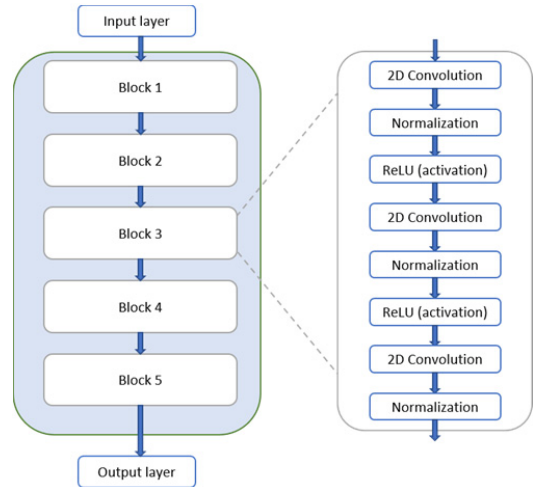


Fig. 4. DNN model's structure

TABLE II. EXPERIMENTAL RESULTS OF TRAINED MODELS.

Model	Model size	Execution time	Accuracy
MobileNet	37MB	22ms	96.80%
MobileNetV2	26MB	27ms	96.20%
EfficientNetB0	47MB	33ms	95.8%
Our model	772KB	10ms	95.20%

The difference in accuracy is very small and can be neglected in comparison with the difference in model size and execution time, which are important to achieve a real-time monitoring system.

VII. CONCLUSION

The paper presented a state-of-the-art analysis of the modern drowsiness detection approaches. We propose to classify drowsiness detection techniques into three groups:

- The driving pattern of the vehicle;
- Psychophysiological characteristics of drivers;
- Computer vision techniques for driver monitoring.

We focused on the methods based on computer vision because it is a non-intrusive technique for the driver. In the scope of the paper, we identify various parameters used (such as eyes state, blinking duration, blinking rate, head angle, gaze, facial expressions, yawning, and breathing rate) for drowsiness detection.

We consider the papers that discuss the drowsiness detection based on these parameters and compare them. The most reliable parameters among these are eye-related parameters. In general, it is not recommended to use only one of the parameter sources for drowsiness detection. In this paper, we started our work on drowsiness detection by detecting yawning. We proposed a three-step method that uses the "FaceBoxes" face detector to locate the face of the driver and uses a DNN model to classify the face and detect yawning. To train the DNN model we need a training dataset so we analyzed existing (at the moment) datasets for yawning detection. We conclude that we should enhance the existing

data with our data recorded in real driving conditions. We used our developed Drive Safely system to record such data of real drivers. We used this dataset to train our yawning detection model that achieved an accuracy of 95.2%. In the future, we are planning to develop our drowsiness detection model by measuring more parameters such as blinking duration and head angle, so we needed our model to be fast for future integration in the drowsiness detection model.

#### ACKNOWLEDGMENT

The research is funded by the Russian Foundation for Basic Research (project # 19-29-06099). Dataset accumulation (Section 5) has been supported by the grant of ITMO University (project № 620176). Drowsiness detection techniques analysis (Section 2) has been supported by Russian State Research #0073-2019-0005.

#### REFERENCES

- [1] N. Administration, "Drowsy Driving", NHTSA, 2019. Web: <https://www.nhtsa.gov/risky-driving/drowsy-driving>.
- [2] B. Reddy, Y. Kim, S. Yun, C. Seo and J. Jang, "Real-Time Driver Drowsiness Detection for Embedded System Using Model Compression of Deep Neural Networks", *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017.
- [3] S. Soares, S. Ferreira and A. Couto, "Drowsiness and distraction while driving: A study based on smartphone app data", *Journal of Safety Research*, vol. 72, pp. 279-285, 2020.
- [4] M. Fitzharris, S. Liu, A. Stephens and M. Lenné, "The relative importance of real-time in-cab and external feedback in managing fatigue in real-world commercial transport operations", *Traffic Injury Prevention*, vol. 18, no. 1, pp. S71-S78, 2017.
- [5] A. Kashevnik, I. Lashkov and A. Gurtov, "Methodology and Mobile Application for Driver Behavior Analysis and Accident Prevention", *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 6, pp. 2427-2436, 2020.
- [6] A. Kashevnik, I. Lashkov, A. Ponomarev, N. Teslya and A. Gurtov, "Cloud-Based Driver Monitoring System Using a Smartphone", *IEEE Sensors Journal*, vol. 20, no. 12, pp. 6701-6715, 2020.
- [7] S. Arefnezhad, S. Samice, A. Eichberger, M. Frühwirth, C. Kaufmann and E. Klotz, "Applying deep neural networks for multi-level classification of driver drowsiness using Vehicle-based measures", *Expert Systems with Applications*, vol. 162, p. 113778, 2020.
- [8] W. W. Wierwille, S. S. Wreggit, L. C. Kim, A. L. Ellsworth, J. R. Fairbanks, "Research on vehicle-based driver status/performance monitoring: development, validation, and refinement of algorithms for detection of driver drowsiness. final report", No. HS-808 247. 1994.
- [9] L. Oliveira, J. Cardoso, A. Lourenco and C. Ahlstrom, "Driver drowsiness detection: a comparison between intrusive and non-intrusive signal acquisition methods", *2018 7th European Workshop on Visual Information Processing (EUVIP)*, 2018.
- [10] M. Mikaili, Z. Mardi and S. Ashtiani, "EEG-based drowsiness detection for safe driving using chaotic features and statistical tests", *Journal of Medical Signals & Sensors*, vol. 1, no. 2, p. 130, 2011.
- [11] K. Chui, K. Tsang, H. Chi, B. Ling and C. Wu, "An Accurate ECG-Based Transportation Safety Drowsiness Detection Scheme", *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 1438-1452, 2016.
- [12] X. Zhu, W. Zheng, B. Lu, X. Chen, S. Chen and C. Wang, "EOG-based drowsiness detection using convolutional neural networks", *2014 International Joint Conference on Neural Networks (IJCNN)*, 2014.
- [13] E. Ouabida, A. Essadiki and A. Bouzid, "Optical correlator based algorithm for driver drowsiness detection", *Optik*, vol. 204, p. 164102, 2020.
- [14] C. Huda, H. Tolle and F. Utaminingrum, "Mobile-Based Driver Sleepiness Detection Using Facial Landmarks and Analysis of EAR Values", *International Journal of Interactive Mobile Technologies (IJIM)*, vol. 14, no. 14, p. 16, 2020.
- [15] O. Khunpisuth, T. Chotchinasri, V. Koschakosai and N. Hnoohom, "Driver Drowsiness Detection Using Eye-Closeness Detection", *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 2016.
- [16] C. Maior, M. Moura, J. Santana and I. Lins, "Real-time classification for autonomous drowsiness detection using eye aspect ratio", *Expert Systems with Applications*, vol. 158, p. 113505, 2020.
- [17] T. Danisman, I. Bilasco, C. Djeraba and N. Ihaddadene, "Drowsy driver detection system using eye blink patterns", *2010 International Conference on Machine and Web Intelligence*, 2010.
- [18] K. Sugiyama, T. Nakano, S. Yamamoto, T. Ishihara, H. Fujii and E. Akutsu, "Method of detecting drowsiness level by utilizing blinking duration", *JSAE Review*, vol. 17, no. 2, pp. 159-163, 1996.
- [19] J. Jo, S. Lee, K. Park, I. Kim and J. Kim, "Detecting driver drowsiness using feature-level fusion and user-specific classification", *Expert Systems with Applications*, vol. 41, no. 4, pp. 1139-1152, 2014.
- [20] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*.
- [21] R. Jabbar, K. Al-Khalifa, M. Kharbeche, W. Alhajyaseen, M. Jafari and S. Jiang, "Real-time Driver Drowsiness Detection for Android Application Using Deep Neural Networks Techniques", *Procedia Computer Science*, vol. 130, pp. 400-407, 2018.
- [22] Q. Ji, Z. Zhu and P. Lan, "Real-Time Nonintrusive Monitoring and Prediction of Driver Fatigue", *IEEE Transactions on Vehicular Technology*, vol. 53, no. 4, pp. 1052-1068, 2004.
- [23] S. Abtahi, B. Hariri and S. Shirmohammadi, "Driver drowsiness monitoring based on yawning detection", *2011 IEEE International Instrumentation and Measurement Technology Conference*, 2011.
- [24] T. Azim, M. Jaffar and A. Mirza, "Automatic Fatigue Detection of Drivers through Pupil Detection and Yawning Analysis", *2009 Fourth International Conference on Innovative Computing, Information and Control (ICICIC)*, 2009.
- [25] P. Viola and M. Jones, "Robust Real-Time Face Detection", *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, 2004.
- [26] Tiesheng Wang and Pengfei Shi, "Yawning detection for determining driver drowsiness", *Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology*, 2005.
- [27] J. Solaz et al., "Drowsiness Detection Based on the Analysis of Breathing Rate Obtained from Real-time Image Recognition", *Transportation Research Procedia*, vol. 14, pp. 3867-3876, 2016.
- [28] O. Jesorsky, K. Kirchner and R. Frischholz, "Robust Face Detection Using the Hausdorff Distance", *Lecture Notes in Computer Science*, pp. 90-95, 2001.
- [29] C. Thomaz and G. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis", *Image and Vision Computing*, vol. 28, no. 6, pp. 902-913, 2010.
- [30] A. Villanueva, V. Ponz, L. Sesma-Sanchez, M. Ariz, S. Porta and R. Cabeza, "Hybrid method based on topography for robust detection of iris center and eye corners", *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 9, no. 4, pp. 1-20, 2013.
- [31] Gourier, Nicolas, Daniela Hall, and James L. Crowley, "Estimating face orientation from robust detection of salient facial structures.", *FG Net workshop on visual observation of deictic gestures*. Vol. 6, No. 7, FGnet (IST-2000-26434) Cambridge, UK, 2004.
- [32] Q. Massoz, T. Langohr, C. Francois and J. Verly, "The ULg multimodality drowsiness database (called DROZY) and examples of use", *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016.
- [33] G. Pan, L. Sun, Z. Wu and S. Lao, "Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Webcam", *2007 IEEE 11th International Conference on Computer Vision*, Rio de Janeiro, Brazil, 2007, pp. 1-8.
- [34] C. Weng, Y. Lai and S. Lai, "Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network", *Computer Vision – ACCV 2016 Workshops*, pp. 117-133, 2017.
- [35] I. Bacivarov, M. Ionita and P. Corcoran, "Statistical models of appearance for eye tracking and eye-blink detection and



measurement”, *IEEE Transactions on Consumer Electronics*, vol. 54, no. 3, pp. 1312-1320, 2008.

- [36] “YawDD: Yawning Detection Dataset”, IEEE DataPort, 2021. Web: <https://iee-dataport.org/open-access/yawdd-yawning-detection-dataset>.
- [37] Howard, Andrew G., et al. “Mobilenets: Efficient convolutional neural networks for mobile vision applications”, 2017.