

METHODOLOGY FOR SELECTION AND DECISION MAKING OF GRADATION OF SLEEPINESS USING ALGORITHMS FOR FACE VIDEO IMAGE PROCESSING

Davis Muska, Matiss Erins

Riga Technical University, Latvia

davis-krisjanis.muska@rtu.lv, matiss.erins@rtu.lv

Abstract. The complexity of the problem to evaluate the human functional state and fatigue can be divided into the research of pathological and physiological forms of fatigue where current research is focused on the physiological form and characterized as proportional to the activity and acute as the functional state can be recovered by sleep and rest. Sleepiness or drowsiness is among the key indications of physiological fatigue and persists throughout the human daily cycle. The risks often associated with a certain gradation of sleepiness, classification of which is described in this article are mainly considered in the context of determining the degree of sleepiness of the driver or operation safety in the workplace. The main methods to be addressed in this paper are the selection of a set of sleepiness characteristics to determine the degree of sleepiness, the selection of the most informative traits for decision-making algorithms, and the resolution of problems associated with visual information extraction from the video. A methodology with multiple algorithms is proposed to obtain values for sleepiness characteristics from a video with human face regions. A total of 28 characteristics were obtained. The decision-making algorithms were trained by using a video database that provided sleep values for the existing video and validated with volunteer video dataset. The classification space three degrees of sleepiness was applied to label the videos, and two classes where low levels of alertness and severe sleepiness were combined into one class. For the prediction of the three degrees of sleepiness, the accuracy of decision algorithms ranged from 37 % to 53 %, which in all cases was better than the free guess limit of 33 %. For the prediction of two levels of sleepiness, the accuracy of decision algorithms ranged from 67 % to 73 %, which in all cases was better than the 50 % limit of free guess.

Keywords: drowsiness states, drowsiness feature selection, classification methods.

Introduction

Drowsiness can be described as a multidimensional and recoverable human functional state that is recovered after sleep or rest and does not involve pathological symptoms, as the authors' previous research was conducted towards normal functional human fatigue state [1]. Current research focuses mostly on driver and other operator sleepiness level detection. The application of dichotomy, in contrast to the majority of previously observed methods, allows to further extend the classification gradations; in this work it provides significant results with three output groups of sleepiness state. Based on authors' previous work on parameter selection [2], the sleepiness detection methods can be divided into four groups of observed data – physiological, activity and subjective features. The current paper focuses on the in-depth analysis and extraction of behavioural domain features to supplement the data acquisition in a decision support system.

Visual video-based sleepiness level prediction methods, which are the focus of the current article, consist of behavioural feature domain. Behavioural data-based studies attempt to extract features from the face region for sleepiness detection [3]. Mainly eyeblink analysis and eye state tracking image processing is applied. Physiological eye muscle signal measurements like electrooculography is a common approach, but visual methods offer a more convenient way to measure features, as there is no need for any attachment of electrodes and it is considered non-obstructive. So, there is a basis to explore visual methods and obtain evaluation alternatives to the ones based on subjective and objective feature domains. This paper focuses on feature extraction from video data and feature selection to build a classification model. Features are summarized from previous work in this field and realized in one framework.

Other work in this domain has the focus on a binary classification between alert and sleepy states, results show around 80 % accuracy to classify alert and sleepy state [4]. Results for eye movement analysis [5] on a binary classifier type show 72.2 % of classification accuracy. Other research to visually detect sleepiness uses eye blink features that are most widely used [4][6], head pose and gaze features [7], facial expression features [8].

This research uses all three feature groups and does feature ranking to find most relation to sleepiness, which are merged into a sleepiness level classification model. Research compares existing

machine learning algorithms to do multilevel and binary classification and to find the most suitable algorithm for sleepiness level classification. These algorithms are linear regression [9], logistic regression [10], Support Vector Machine [11], K-Nearest neighbours [12], Extra tree classifier [13], logistic regression with cross validation [14], decision trees [15].

Materials and methods

Overall methodology proposed in the current work to build a human sleepiness level classifier is composed from five steps (Fig. 1).

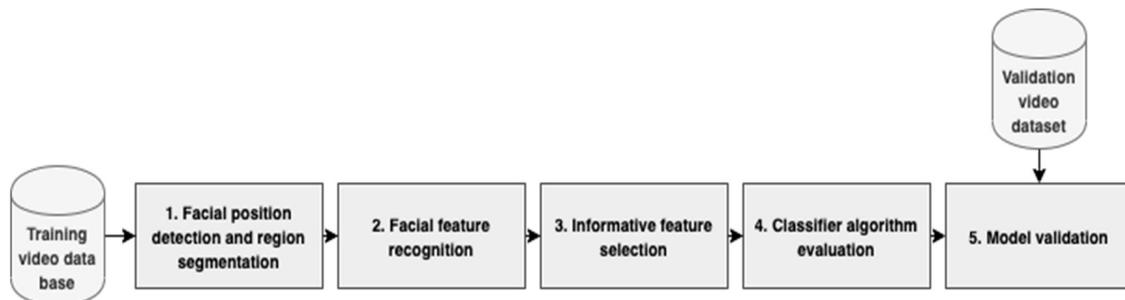


Fig. 1. Methodology for drowsiness classifier model

1. **Facial position and region segmentation** solve the problem to find the position of face in the coordinate space of the video and to extract different regions of face (eyes, mouth, etc.).
2. **Facial feature recognition** solves the problem to detect and quantify features of sleepiness. In this step a specific sub method is used to find feature values over time and merge them over a specific interval of time.
3. **Informative feature selection** solves the problem of informative feature extraction to build the classification model. All features are ranked by their importance weight and features with the weight higher than the importance threshold are selected.
4. **Classifier algorithm comparison** solves the problem to build different classification models using the same data and to compare accuracy of classification between them. At this step, all models are compared by accuracy of classification, dataset was divided into training and validation set, the model was trained against training set that was 70 % of all data and the model was validated against 30 % of all data. Multilevel and nested dichotomy are compared by accuracy of classification. For nested dichotomy there is a model built for each level of classification, there are two levels, as videos in the database [16] are labelled by three types of labels: alert, medium sleepiness and high sleepiness. The first level does classification between the alert and other state, the second level does classification between medium sleepiness and high sleepiness.
5. **Classifier model validation** experiment was conducted by the authors in order to obtain a test set for the model validation. The test set consisted of 10 subject video recordings for two significant sleepiness gradations – with and without sleep deprivation of > 4 hours of sleep previous day. The measurements were conducted during both conditions to obtain self-evaluation of sleepiness, as the first condition resulted in the classifier reporting all data under normal drowsiness class. The validation video data set was formed in the Riga Technical University from a set of healthy student volunteers and applied controlled measurement conditions, use of a measurement design and video acquisition devices.

Facial position detection and region segmentation

Two main problems to solve are the face position detection in video and facial region segmentation. Four popular face detection algorithms were chosen and compared by their usability in OpenCV software:

- Haar-like features;
- Deep Neural Network (DNN) algorithm;
- Histogram of Oriented Gradients (HoG) algorithm;
- Maximum-Margin Object Detector (MMOD) algorithm [17] and [18].

The algorithm usability was evaluated by using test software that converts video to frames and detects face in each frame. All experiments were done on a personal computer with CPU in same conditions, so the results may vary among different machines, but the trends should be the same. Usability was evaluated by three criteria:

- algorithm performance frames-per-second (Fig. 2);
- detected face boundary precision;
- false positive rate (Fig. 3).

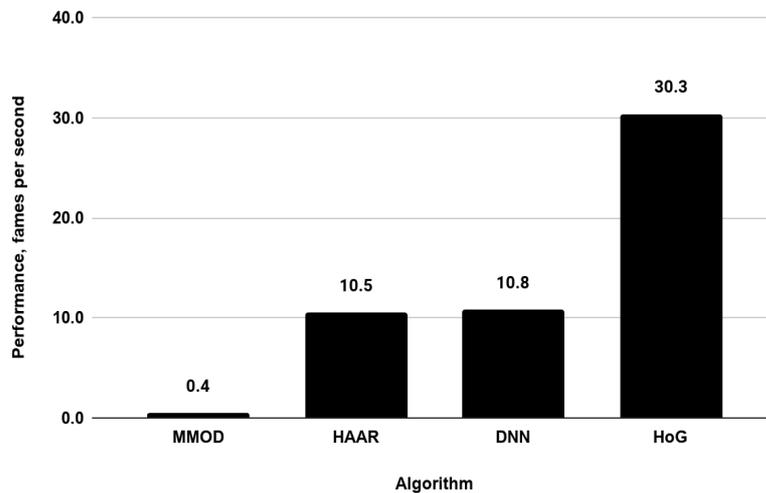


Fig. 2. Segmentation algorithm performance comparison

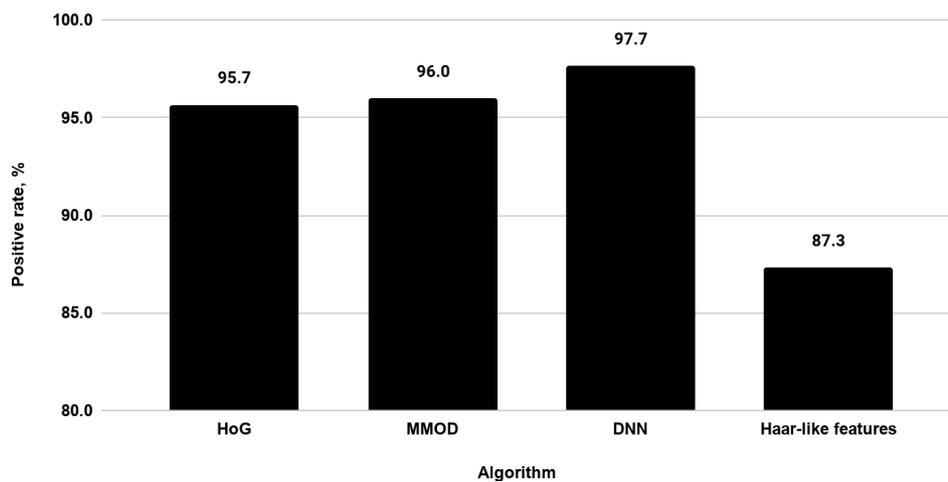


Fig. 3. Segmentation algorithm false positive rate comparison

There are four main methods to do face region segmentation:

- Artificial neural networks;
- Pattern matching;
- Facial landmark detection;
- Specific characteristics detection.

Facial landmark detection [19] was chosen as the most appropriate method, as it provides information about all face regions and contours. In other methods it is not possible to do face segmentation using one model. Facial landmark model consists of 68 points that can be grouped into six groups: jaw; eyes; eyebrows; chin; lips; nose.

There are some drawbacks of visual face detection and segmentation. Precision decreases when the face is partially covered (glasses, hair). Some details might be missing in low resolution videos, for example, a gap between eyelids might not be distinguishable. Another limitation is lighting conditions if no artificial lighting is used, as it might not be possible in some real-life scenarios.

Facial feature recognition

Facial features for sleepiness level prediction were selected from literature. At this step features were not sorted by their importance; all measurable features were selected for evaluation. All features can be sorted into three categories:

- Head movement features [20]:
 - head orientation;
 - swing frequency.
- Eye blink features [21]:
 - median eye-opening height;
 - median distance eyes closed;
 - median distance eyes open;
 - PERCLOS (percentage of eyes open below a certain threshold);
 - percentage of eyes open within a certain interval;
 - eyeblink frequency.
- Facial expression features [22]:
 - facial muscle skin fold level.

Informative features

Best features for classification algorithm training must be selected to ensure that the model is trained with features that contain information about the problem. Not all the features are informative and have information about the problem. There are many formal ways to find informative features in the data set. Random forest algorithm was selected in this study. Random forest algorithm is a decision tree-based algorithm that calculates the importance of features. Each feature gets a coefficient between 0 and 1 and the sum of all features is 1. If some feature has a higher value than others, then it is more important. Algorithm was realized using the Python Scikit-learn machine learning package [23].

Prior to inspecting the feature importance, it is important to check that the model predictive performance is high enough. Indeed, there would be little interest of inspecting the important features of a non-predictive model and the method reduces the computational cost through feature set minimization and reduces the chance of overfit the machine learning model. The results are shown in Table 1. There are 28 features, so the importance threshold is 0.036 (1/28). All features above the threshold are considered important and highlighted in Table 1.

Table 1

Drowsiness feature importance (Random forest MDI)

Feature	Importance	Feature	Importance
Median distance eyes closed	0.0496	PERCLOS 60 %	0.0342
Vertical swing frequency of head	0.0478	Eyes opened 30-40 %	0.0341
PERCLOS 30 %	0.0459	PERCLOS 80 %	0.0339
Eyes opened 10-20 %	0.0442	Eyes opened 60-70 %	0.0336
Eyes opened 20-30 %	0.0411	Eyes opened 90-100 %	0.0321
PERCLOS 20 %	0.0410	PERCLOS 90 %	0.0320
PERCLOS 10 %	0.0409	Eyes opened 70-80 %	0.0316
Median eye-opening height	0.0375	Skin fold level region 3	0.0311
PERCLOS 50 %	0.0363	Skin fold level region 1	0.0310
Eyes opened 30-40 %	0.0357	Skin fold level region 4	0.0303
Eyes opened 50-60 %	0.0355	Median distance eyes closed	0.0302
Eyes opened 80-90 %	0.0352	Skin fold level region 5	0.0298
PERCLOS 40 %	0.0351	Skin fold level region 2	0.0294
Eyeblink frequency	0.0348	PERCLOS 70 %	0.0260

Classifier algorithms

Nine important features were selected to train the sleepiness level prediction algorithm. Six classification algorithms were selected and one regression (linear regression) algorithm was selected.

For linear regression algorithm an extra step was done for conversion between continuous data to classes. Classification algorithms were [24]:

- Logistic regression (LR);
- Support vector machine (SVM);
- K-Nearest neighbours (KNN);
- Decision tree (DT);
- Extra tree classifier (EXTC);
- Logistic regression with crossover validation (LCV).

All classification algorithms were implemented by using Python Scikit-learn machine learning software [25]. The supervised algorithms were trained by using self-labelled sleepiness video database by the University of Texas at Arlington mentioned earlier. Database is composed of videos which contain faces of people. Each participant was filmed at three different degrees of sleepiness. There were 20 participants. Each video is 10 minutes long and is recorded using a general-purpose video recording device. Each participant evaluated his level of sleepiness using Karolinska sleepiness scale (KSS). Karolinska sleepiness scale has 10 degrees of sleepiness. Database authors reduced to three degrees of sleepiness- alert (KSS = [1; 3]), medium sleepiness (KSS = [6; 7]) and high level of sleepiness (KSS = [8; 9]). Each model was trained using 70 % of data in database and 30 % were used for model evaluation.

Results

Classifier algorithm evaluation. Figure 4 shows the achieved classification accuracy between different algorithms. All models are giving meaningful predictions as accuracy is above the guessing threshold. In the three class case this threshold is 0.33(1/3). The best performing classification model is decision trees. However, accuracy of all models is below 60 % of accuracy in the 3 class case.

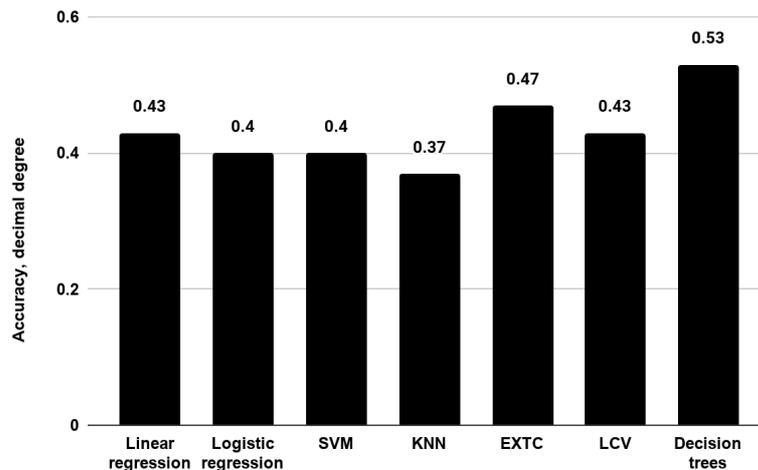


Fig. 4. Classification algorithm result 3 classes

When looking at the data, the biggest loss of accuracy results from classification between medium and high degree of sleepiness. An experiment was done to obtain the video samples where both medium and high degrees of sleepiness were merged into one class. Figure 5 shows classification accuracy at two class (alert (KSS = [1;5] and sleepy (KSS = [6;10]) case and it corresponds to the drowsiness detection task.

In the two class case classification accuracy is around 70 %. There is no noticeable difference between different algorithms. So, a conclusion can be made that the problem is not in the prediction algorithm, but in the features selected to predict the degree of sleepiness. To increase classification accuracy other features must be selected. In the two class case classification has reasonable accuracy. Previously the multiple class classification approach was used, one model does prediction in more than two class space.

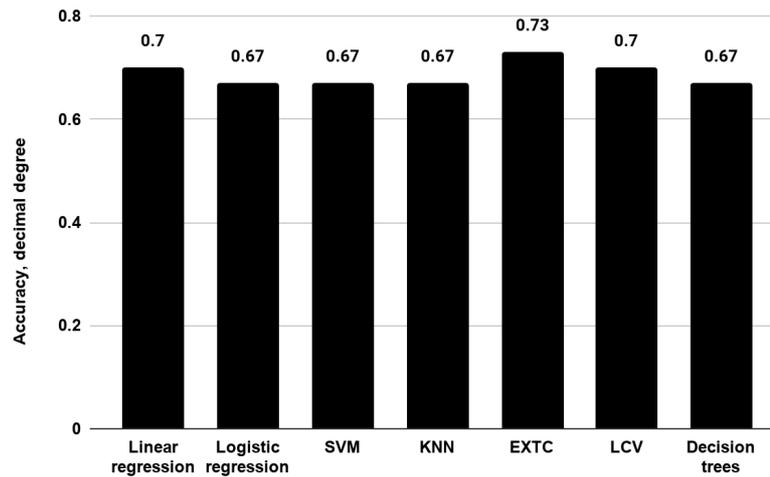


Fig. 5. Classification algorithm result 2 classes

Classifier model validation. Another approach to do multi class classification is to use a dichotomy method, where at each step decision is made between two possible outcomes. Two classification models were build- first to do classification between alert and sleepy state and second to do classification between medium and high degree of sleepiness. Figure 6 shows classification accuracy for the dichotomy method. Overall classification accuracy with the dichotomy approach is around 60 % that is higher compared to accuracy at the three class case that is around 40 %.

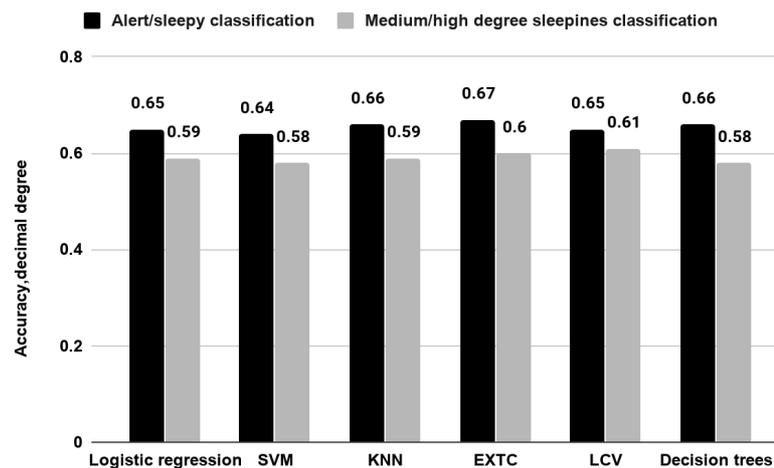


Fig. 6 Model validation result with dichotomy approach

Discussion

Current research goal is to provide a sleepiness classifier input for an expert system, which will perform decision-making in human fatigue and workability domains, where the proposed method extends the 2 class detection methods and provides 3 class output that is required in differentiation among the most common elevated state of fatigue more common within real working conditions and the high-risk drowsiness state that imposes the threat to individual or working environment. The parts of a hybrid-system use machine learning in the first phase to convert physiological and behavioural parameters into sub-decision drowsiness gradations as indices for an expert system. The decision is formulated in the expert system from fuzzy rules and the human knowledge applied on the drowsiness indices collected in the domain of human fatigue. The classifier results show enough processing time for applications in real-time video streaming online solutions which rely on approximate frame rates close to 30 fps, what is not achievable with methods that use 2-10 frames-per-second temporal decrease for classification response. The main downsides of the current approach are the static camera position and limitation to day-light environmental conditions. The authors' future work will focus on a wearable prototype with infra-red camera aimed toward improvement of the distance to eyes, closer

tracking of the eye from multiple head positioned cameras to eliminate the obstructions to the face object and limitations with distance and lightning conditions. The applications are aimed towards a stationary positioned person within office, equipment or vehicle operations environments.

Conclusions

1. This study demonstrates that the human sleepiness state can be detected with accuracy 70 % by using video sequence processing from the facial video. With the help of dichotomy, the classifier results show improved 3-class drowsiness classification.
2. Landmark detection of 68 feature points was selected as the segmentation method, as other methods do not allow facial segmentation by a single model, reaching performance evaluation of 30 frames per second in video processing and over 95 % precision rate for the HoG method.
3. A few drawbacks of the facial segmentation method including poor lighting, partial facial coverage and angle of orientation that decreases the overall recognition accuracy will be addressed by an improved wearable infra-red camera sensor design.
4. Decision tree classifier algorithm is suitable for non-linear feature dimensionality reduction and parameter selection tasks to avoid model overfit and create a supervised learning model for multi-class decision making with accuracy ranged from 67 % to 73 %.

Acknowledgements

EUREKA programme project “Non-intrusive human fatigue assessment” E! 11169 Fatigue.



References

- [1] Bjorness T., Greene R. Adenosine and sleep. *Curr Neuropharmacol.* 7(3), 2009, pp. 238-245.
- [2] Erins M., Minejeva O., Markovičs Z., Lauznis J., Kivlenieks R. Research of human fatigue and measurement parameters for workability assessment. *Environment. Technology. Resources. Proceedings of the International Scientific and Practical Conference.* 2, 2019, pp. 38.
- [3] Ngxande M., Tapamo J., Burke M. Driver drowsiness detection using behavioral measures and machine learning techniques: A review of state-of-art techniques. *Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech), Bloemfontein,* 2017, pp. 156-161.
- [4] Picot A., Charbonnier S., Caplier A. Drowsiness detection based on visual signs: blinking analysis based on high frame rate video. *IEEE Instrumentation & Measurement Technology Conference Proceedings, Austin, TX,* 2010, pp. 801-804.
- [5] Jin L. et al. Driver Sleepiness Detection System Based on Eye Movements Variables. *Advances in Mechanical Engineering,* Jan. 2013.
- [6] Danisman T., Bilasco I.M., Djeraba C. Drowsy driver detection system using eye blink patterns. *International Conference on Machine and Web Intelligence, Algiers,* 2010, pp. 230-233.
- [7] Choi In-Ho, Kim Yong-Guk. Head pose and gaze direction tracking for detecting a drowsy driver. *International Conference on Big Data and Smart Computing (BIGCOMP), Bangkok,* 2014, pp. 241-244.
- [8] Hachisuka S., Ishida K., Enya T., Kamijo M. Facial Expression Measurement for Detecting Driver Drowsiness. In: Harris D. (eds) *Engineering Psychology and Cognitive Ergonomics, EPCE,* 2011.
- [9] Schneider A., Hommel G., Blettner M. Linear regression analysis: part 14 of a series on evaluation of scientific publications. *Dtsch Arztebl Int.,* 107(44), 2010, pp. 776-782.
- [10] Peng J., Lee K., Ingersoll G. An Introduction to Logistic Regression Analysis and Reporting. *Journal of Educational Research - J EDUC RES.* 96, 2002, pp. 3-14.
- [11] Evgeniou T., Pontil M. Support Vector Machines: Theory and Applications, 2001, pp. 249-257.
- [12] Guo G., Wang H., Bell D.A., Bi Y., Greer K. KNN Model-Based Approach in Classification. *CoopIS/DOA/ODBASE,* 2003.
- [13] Geurts P., Ernst D., Wehenkel L. Extremely randomized trees. *Machine Learning,* 2006, pp. 3-42.
- [14] Berrar D. Cross-Validation. *Encyclopedia of Bioinformatics and Computational Biology,* 2018.

- [15] Song YY., Lu Y. Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry, 2015, pp.130-135.
- [16] Ghoddoosian R., Galib M., Athitsos V. A Realistic Dataset and Baseline Temporal Model for Early Drowsiness Detection, 2019. CVPR Workshops. [online] [30.03.2020]. Available at: <https://arxiv.org/abs/1904.07312v1>.
- [17] Dang K., Sharma S. Review and comparison of face detection algorithms. 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, Noida, 2017, pp. 629-633.
- [18] Gupta. Face Detection – OpenCV, 2018. [online] [30.03.2020]. Available at: <https://www.learnopencv.com/face-detection-opencv-dlib-and-deep-learning-c-python>.
- [19] Soukupová T., Cech J. Real-Time Eye Blink Detection using Facial Landmarks. International Symposium on Affective Science and Engineering. ISASE2018, 2018, pp. 1-6.
- [20] Garcia F., Diaz B., Jilliam M., Mirbach, B., Varanasi K., Stricker D. Combined Framework for Real-Time Head Pose Estimation using Facial Landmark Detection and Salient Feature Tracking, 2018.
- [21] Roshani T., Zoroofi R. Open/Closed Eye Analysis for Drowsiness Detection, 2008, pp. 1-7.
- [22] Hachisuka S., Kimura T., Ishida K., Nakatani H., Ozaki, N. Drowsiness Detection Using Facial Expression Features, DOI:10.4271/2010-01-0466, 2010.
- [23] Breiman L. Random Forests. Machine Learning, 45(1), 2001, pp. 5-32.
- [24] Hastie T., Tibshirani R., Friedman J. Elements of Statistical Learning, Springer, 2009.
- [25] Pedregosa et al. Scikit-learn: Machine Learning in Python, JMLR 12, 2011, pp. 2825-2830.