Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Driver drowsiness detection system using hybrid approach of convolutional neural network and bidirectional long short term memory (CNN_BILSTM)

Rajamohana S.P. *, Radhika E.G., Priya S., Sangeetha S.

Department of Information Technology, PSG College of Technology, Coimbatore 4, Tamil Nadu, India

ARTICLE INFO

Article history:
Received 21 October 2020
Received in revised form 18 November 2020
Accepted 24 November 2020
Available online xxxx

Keywords:
Eye dataset
Euclidean algorithm
CNN and BiLSTM

ABSTRACT

In today's world driver drowsiness is a major reason for fatal accidents of on road vehicles. Developing an automated, real-time drowsiness detection system is essential to provide accurate and timely alerts to the driver. In the proposed system, hybrid approach of CNN (Convolutional Neural Network) and BiLSTM (Bidirectional Long Term Dependencies) is used to detect the driver's drowsiness. Video camera is used to track the facial image and eye blinks of the driver. The proposed system works in three main phases: In the First phase, driver's face image is Identified and observed using a web camera. In the Second phase, the eye image features are extracted using the Euclidean algorithm. During the third phase, the eye blinks are continually monitored. The final stage decides whether the measure in eye square is closed state or open state. When a driver falls asleep, there will be a warning message to alert the driver to prevent road accidents.

Selection and peer-review under responsibility of the scientific committee of the International Conference on Advances in Materials Research 2019.

1. Introduction

The number of accidents is increasing rapidly in many countries and many people lose their life due to accidents. Due to driver's drowsiness, 100,000 vehicles per year collide, according to the National Highway Traffic Safety Administration (NHTSA). The German Road Safety Council also reports that 25% of highways deaths on motorways are caused by driver fatigue .There are many ways to identify driver drowsiness, e.g. Cardiac activity, facial expressions, flicker frequency and brain activity by electroencephalography (EEG) [5].

The human face is complex and exhibits a high degree of variation. Facial recognition is considered as a difficult problem in computer vision research. Human eyes play a significant role in the facial recognition and facial expression analysis. Human eyes are the stable feature than compared with the other facial features [6]. Therefore, in recognizing facial features, it is advantageous to recognize the eyes. Using the eye position we can estimate the location of certain facial features. Furthermore, only with the presence of both eyes the scale orientation and rotation of the face in the picture plane can be normalized. In this paper, CNN-BILSTM approach has been proposed to detect the driver drowsiness (Fig. 3.1).

2. Related works

In 2018, Hongzhe, B. et al. [25], proved the (Fig. 3.3) driver fatigue causes major road accidents. It was detected by drivers behavior, operation state of vehicle, facial expression. Here the methods are implemented by Image Preprocessing, Facial detection and localization, Positioning of Human Eye. The eye location and movement is tracked by Unscented Kalman filter(UKF).Data set was extracted from base vector to provide subspace .In this, fatigue training samples were included in the dataset [8]. After getting the image of the driver fatigue it is then compared with the images from dataset and provided a result. In 2016, Ding, W., Xu, et.al [4], shows video based sub-challenge with the help of Wild Challenge (EmotiW2016).The main level in this method is the facial and audio emotion recognition. A network pre-trained on image net training data for Recognition of face like feelings, emotions etc. A deep Convolutional Neural Network is used for training data for

* Corresponding author.
E-mail addresses: spr.it@psgtech.ac.in (S.P. Rajamohana), egr.it@psgtech.ac.in (E.G. Radhika), spa.it@psgtech.ac.in (S. Sangeetha).

https://doi.org/10.1016/j.matpr.2020.11.898
2214-7853/
Selection and peer-review under responsibility of the scientific committee of the International Conference on Advances in Materials Research 2019.

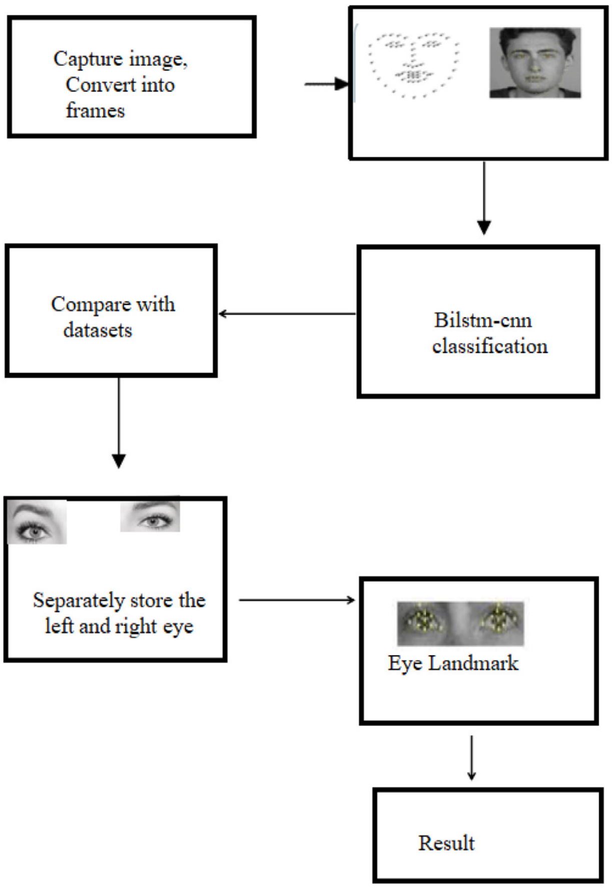

Fig 3.1. Workflow of the system.

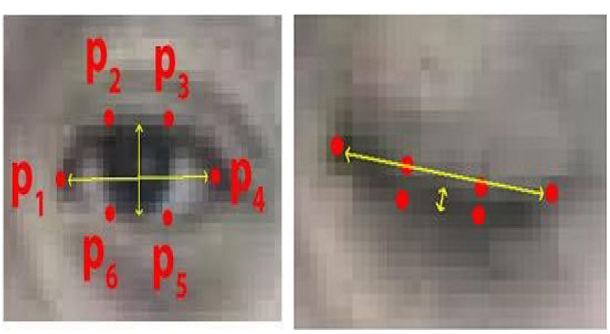

Fig 3.3. Eye Aspect Ratio Calculation.

extraction of the feature in a video-based approach that applies deep learning methods for small emotion recognition datasets and succeeds in a final precision.

In 2015, Lee, K. W. et.al [3], Yoon describes about the huge loss of life and property due to improper aggressive driving. To detect emotions and prevent violent driving using CNN (Convolutional Neural Network) based on input images of driver's face using thermal camera sensors. In 2010, Chai, T. Y., Woo et.al [1] defines about Electroencephalogram (EEG) and how it classifies human emotions. In this, EEG signal collects the data by its own for study of six human subjects, which was performed based on the effect of emotion stimuli. It varies based on the range of the emotions been reacted. The achieved classification rate is very high. In 2008, De Naurois, C. J. et.al [2] describes measuring behavioral and physiological activities such as respiration rate, heart rate variability eyelid and head movements (blink length, amplitude and PERCLOS (Percentage of eye closure) using Video-oculo-grapy and its recording as a driving action such as time-to lane crossing, speed, wheel angle, location on the road and time to lane crossing. Drowsiness is an interpersonal interaction between state of alertness and state of sleep. In this the participants were tested by the material process, the protocol, data analysis and modeling. It expects unintended carelessness to the driver [11].

3. Methodology

The proposed system utilizes human eye dataset and classification algorithms such as CNN and BiLSTM to classify the driver's drowsiness. For Eye data, Euclidean distance of the eye is calculated with CNN and BILSTM to classify the features accordingly. Driver drowsiness is detected by monitoring the eye blink rate and notification is sent.

3.1. Data collection

The eye dataset is gathered from the mrl.cs.vsb.cz / eye dataset web site. It requires 2208 images out of which 1104 images are related with open eye, with and without spectacles. The other set of images 1104 reflects closed eye, with and without spectacles.

3.2. Pre-processing

The content is received from the real time data of the facial image using shape predictor dlib. Eye images are captured using web camera and splits the eye region using distance calculation. Then the noisy data is removed. The left and right eye noisy data get converted to gray scale from the captured video and stored as frames in the separate folder for data analysis.

3.2.1. Facial landmarks with OpenCV and dlib

Within dlib, the facial landmark detector is implemented, and it produces sixty eight (x, y)- coordinates that maps different structures of face. These mappings were obtained by coaching a predictor in the form on the dataset labeled iBUG 300-W.The data was trained and stored in a .dat file format to utilize the facial part of the detected images and for real time capturing of videos. The facial land mark consists of numerical value assumption for understanding the value for all parts of the face. The right eye region values are started from the range 36 to 42.The left eye region values are started from the range 42 to 48. Using shape predictor, real time video eye images can be identified and tracked parallelly**.** Hence proposed system uses shape predictor for tracking the eye location thereby calculating the driver drowsiness level with a Euclidean Algorithm.

3.2.2. Euclidean algorithm with OpenCV

The technique which is based on computer vision are significantly successful. For detecting eye blinks, this method uses facial expression and visual bio-behavior, such as head optimistic gaze, eye opening and eyelid movements. The proposed algorithm is based on a computer vision system, which focuses mainly on blink detection by estimating the EAR (Eye aspect ratio).This is done by examining motive force of the eye motion in the video frames. A web camera is used to capture live footage of driver's facial expressions and eyes in all light conditions and frames**.** This will be periodically collected for video capture image processing scheme. EAR is defined as equation (3.1), in which p1, p2 p6 is a 2D facial landmark. The numerator (3.1) calculates the distance between the landmarks of vertical whereas the denominator calculates horizontal eye landmark distances. The eye ratio is remained to be constant when the focus is on the open eye, which during blink rapidly drops to zero. When the person is blinking, the aspect ratio of the eye significantly decreases, approaching zero. Eye ratio is constant, near to zero that increases more than once; indicating one blink has taken place. During capturing of an image if any facial-part is identified then calculate the eye region based on the ear aspect region. Any colored image should be converted into gray scale image before applying the algorithm. The colored image might take more time to identify and remove the noisy or unwanted data. The next step is to crop the image so that portion of eye has to be visible for quick identification of drowsiness condition of driver.

3.3. Feature extraction

The feature extraction steps deals with identifying the types of features used to detect drowsiness. The trained classified data weights are stored in the H5 file. A weight term is often used in information retrieval and drowsiness calculation.

3.3.1. Convolutional neural network (CNN)

Convolutional neural networks are a series of layers stacked sequentially which is composed of multiple layers - convolution, pooling and fully connected layers. The model parameters are trained using gradient descent.

The Convolutional layer contains a three dimensional image tensor, consisting of d feature map channels, size h / w pixels where w denotes the weight of the input image tensor and h its height

Through two dimensional kernel K convection and using input tensor I the convolutional layers will extract low level features. (3.2)

Where bm, n is a parameter of bias, and i, j denotes the coordinates of a pixel of the feature map. An activation function is applied after convolution to introduce non-linearity and to generate an output feature map a, which includes components, a(i,j) = f(c(i,j)).

$$f(c) = \left\{ \begin{matrix} c, ifc > 0, \\ 0, ifc \leqslant 0, \end{matrix} \right\} \tag{3.2}$$

The ReLU function is used as activation function to prevent deeper vanishing gradient problems in Convolutional Neural Networks. . After activations the pooling layers are also added. Here, down sampling is applied to reduce image dimensionality through a grouping process over activations in the input image in small spatial regions. A Convolutional layer with many filtering steps may also be used to achieve down sampling. Typically, the final layer is a fully connected output layer producing a network output

$$z = Wa + b \tag{3.3}$$

Where W is a matrix of weight, and where a is the previous layer entry. The activation function is the sigmoid function, usually used for the tasks of binary classification. Convolutional neural networks are also modeled to have parameters for reducing any loss using gradient descent. For categorical classification functions, such as identification of drowsiness, the categorical cross-entropy loss from

$$l(\bar{y}, y) = -\frac{1}{N}\sum_{i}^{N}[y_i \log \bar{y}_i + (1 - y_i)\log(1 - \hat{y})] \tag{3.4}$$

where ŷ value is assumed to be predicted. Categorical cross entropy will compare the prediction distribution with whenever the likelihood of the verity category is close to one and zero. In order to position it in an extremely different form, the concept of verity could be noted as a one-hot encoded vector. The closer the square measurement of the model outputs the lower the loss. Operating this loss helps to control parameters of model weight and bias, with the Adam optimizer.

3.3.2. Implementation of CNN

CNN extract the needed features from the data sets and splits the data set into 2 pieces. One for training sessions and another one for checking. Assuming a Ground-truth value as empty and Image-data as empty for storing the existing data set and real time data. The existing benchmark dataset was randomly divided and taken 70% for testing and 30% for training. In image-data the real time data randomly divides and take 70% for training and 30% for testing. InceptionV3, an inbuilt library function of keras. Using that function it will resize the images into same size (150, 150, 3) using max pooling function to get the maximum value of the input images . Consider the Dense value 1000 with Relu activation function. Given the same input image with dense value 2 with softmax activation function for prediction .Then assume a model with input image and prediction image. Compile the model with categorical cross_entropy with Adam optimizer and metrics as accuracy. Fit the Ground-truth trainable image and image-data trainable image run epochs 50 consider batch_size = 128 with verbose 1. The trainable model weight value is open with CNN.json and save the trainable weigh value into CNN.h5 for calculating the drowsiness detection of driver.

4. Classification

The results are shown based on accuracy calculated with combination of CNN-BILSTM.

4.1. Bidirectional LSTM (BILSTM)

CNN may provide better results for the classification and for the detection of drowsiness, but it raises problems like, though an eye is usually closed it is considered drowsy but however the driver's real world is not sleepy. The BiLSTM model can be used in conjunction with CNN to improve the accuracy. It is similar to long term memory in which neural network LSTM integrates this information. LSTM networks prevents gradient descent problem and manages dependencies efficiently. . The secret to neural LSTM networks is the cell state Ct, which is propagated over time. Ct is modified at every time stage as follows:

$$\begin{aligned} f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\ c_t &= \tanh(W_c[h_{t-1}, x_t] + b_c)^{\sim} \\ C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \end{aligned} \tag{4.1}$$

where $x_{t,}$ $h_{t-1,}$ f_t and i_t represents current input, previous state output, forget gate and input gates respectively. The weight matrices for gates represented as W_f, W_i and $W_{c.}$ with respective biases as b_f, $b_i$and $i_{c.}$ with sigmoid function as π.. The output of LSTM blocks is as follows:

$$\begin{aligned} \sigma_t &= \sigma(W_0[h_{t-1}, x_t] + b_0) \\ h_t &= o_t * \tanh(C_t) \end{aligned} \tag{4.2}$$

where o_t denotes gate of output, w_t and b_o are matrix of respective weight and bias So, h_t is a filtered cell-state version, governed by the output gate. Our classification model is shown. Dropout is added to the LSTM layer performance to make the model more robust. This is true for LSTM's forward approach and backward approach. It also helps you to define the merge mode which is to show, how to combine the forward and backward outputs before being moved on to the next layer. The possibilities are to:

'Sum': Additional unit of the outputs area together. 'Mul': The unit area of the outputs increased along.

'Concat': the unit area of the outputs concatenated along (the default) supplying twice the amount of output to the corresponding layer.

'Ave': is taken traditional of outputs.

To find the weights, the efficient ADAM optimization algorithm is used, and the accuracy metric is measured and published every epoch. The LSTM is to be educated in 1,000 epochs. Each epoch, a new random sequence of inputs will be generated for the network to operate on. This means that the model does not memorize a particular sequence but can generalize a solution to solve all possible sequences of random inputs for this problem. The network will be tested on yet another random sequence once it is educated. The predictions to drowsiness detection will then be correlated with the predicted performance sequence. Prints the log loss and the precision of the classification at random. It provides a good understanding of how well the model generalized a solution for the detection of drowsiness.

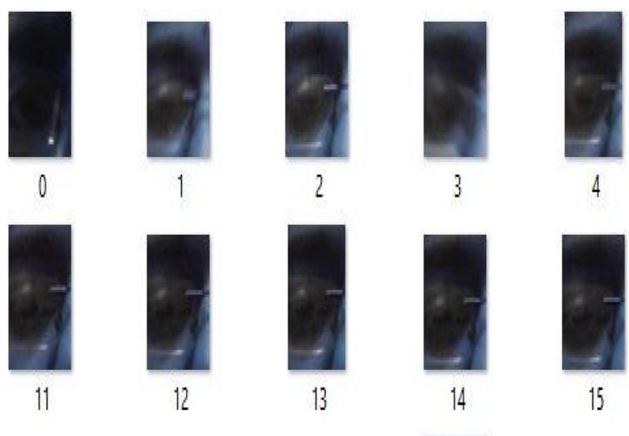

Fig 6.1. Pre-processed Result of Left Eye.

4.2. Implementation of BiLSTM

In BiLSTM deepfeat.npy file was downloaded with net connection before implementing. It contains the in-built library to fast and quickly identify the drowsiness state of a driver with that npy file. That npy file is used for deep feature selection of input images. It takes less time than CNN. Loading the deepfeat.npy into deep feature. in LSTM_model fit the Ground-truth value and deep feature run with Epoch 50 with batch_size 128 with verbose 1. open the LSTM_model with BiLSTM_model.json. That weight value was saving to BiLSTM_model.h5 file. Predict the LSTM model input with 3000 in single epoch and consider the maximum argument with axis value 1. Using Both CNN.h5 and BiLSTM_model.h5 file combine to detect the drowsiness of the driver. Both the combination gives approximate accuracy of detecting the condition of the driver condition and provides a proper notification.

5. Performance evaluation

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5.1}$$

Recall is the percentage of all properly listed categories. This evaluates the effectiveness of the specific framework that is used in classification. Recall is given in eqn (5.2).

$$Recall = \frac{TP}{TP + FN} \tag{5.2}$$

Accuracy is calculated as ratio of true positive and true negative to the overall measures. Equation (5.3) gives the accuracy formula.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.3}$$

F1 Score is calculated as the ratio of weighted average of precision and Recall to twice the combination of Precision and Recall. Equation (5.4) gives the F1 Score formula.

$$F1Score = \frac{2 * (Recall + \text{Precision})}{Recall + \text{Precision}} \tag{5.4}$$

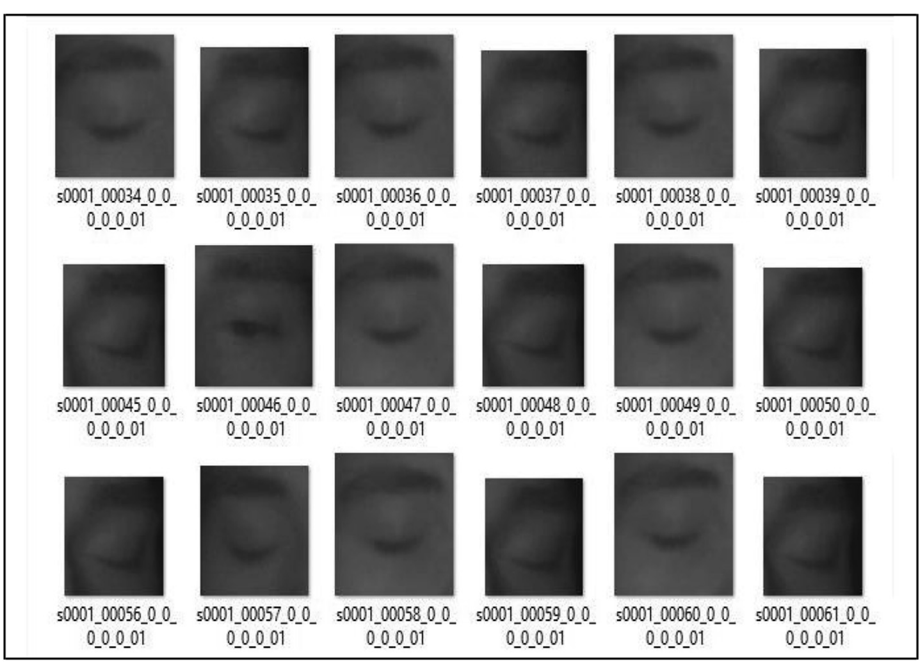

Fig 6.2. Pre-processed Result of Right Eye.

Table 6.1
Performance metrics of CNN and BILSTM.

Classification Algorithm	Accuracy (%)	Precision (%)	Recall (%)
CNN	85	81.25	83.7
CNN_BiLSTM	94	90.67	96

6. Experimental study

The proposed CNN and BILSTM algorithm is implemented in Windows 10 Qualified operating system environment using python with 2.66 GHz CPU, Intel P6 and 16 GB RAM. In this experiment, CNN and BILSTM feature selection and classification algorithm is used to detect the drowsiness stage of driver in real time with eye dataset. The phases of the results of the proposed methods are set out below.

6.1. Dataset description

The Eye dataset is extracted from the URL mrl.cs.vsb.cz/ eye dataset. It consists of 2208 images which includes 1104 closed left and right images with and without glass and 1104 open left and right images with and without glass.

Fig. 6.1 shows the pre-processed real time images of left eye which was converted into frames from video, for validating the eye with updating of CNN.h5 file and BiLSTM.h5 file of trained and tested datasets.

Fig. 6.2 shows the pre-processed real time images of right eye which was converted into frames from video for validating the eye with updating CNN.h5 file and BiLSTM.h5 file of trained and testing datasets.

The accuracy of classification algorithm is implemented with CNN _BiLSTM are as shown in Table 6.1.

7. Conclusion

The proposed work is to recognize facial landmarks from images are captured as the person is driving the vehicle, through a camera mounted onto the vehicle and provide the information acquired to the qualified model to identify the condition of the driver. If the data collected is identified to show signs of drowsiness of the driver would be notified to stop the vehicle to prevent accidents. The hybrid approach of CNN BiLSTM is proposed for open-eye detection of driver drowsiness. The performance of the proposed method is adequate and the use of a web camera during the night time could be improved. The approach to CNN BiLSTM has been shown to work effectively in low- resolution images for many eye postures. The data collection was split into 70 percent for preparation and 30% for testing, covering all stages of eye openings and eye-glass samples. Information about the drivers head and eye position are obtained through shape predictor with Euclidean algorithms. This method yields excellent results on images with varying backgrounds and lighting. The proposed system can therefore detect whether your eyes are opened or closed. If eyes are closed for too long, a notifying call may be sent to the driver. This system detects somnolence and warns the driver by notifying a call in real-time. This system senses drowsiness and in near real-time, it will warn the driver.

CRediT authorship contribution statement

Rajamohana S.P.: Writing - review & editing, Conceptualization, Methodology, Software. **Radhika E.G.:** Validation, Formal analysis, Investigation. **Priya S.:** Investigation, Resources, Data curation. **Sangeetha S.:** Writing - original draft, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. Zhang, H. Wang, R. Fu, Automated detection of driver fatigue based on entropy and complexity measures, IEEE Trans. Intell. Transport. Syst. 15 (1) (2014) 168–177, https://doi.org/10.1109/TITS.2013.2275192.
[2] N. Gurudath, H.B. Riley, Drowsy driving detection by EEG analysis using wavelet transform and K-means clustering, Procedia Computer Science 34 (2014) 400–409, https://doi.org/10.1016/j.procs.2014.07.045.
[3] B.N. Manu (2016, November). Facial features monitoring for real time drowsiness detection. In 2016 12th International Conference on Innovations in Information Technology (IIT) (pp. 1-4). IEEE.
[4] K. Fujiwara, E. Abe, K. Kamata, C. Nakayama, Y. Suzuki, T. Yamakawa, T. Hiraoka, M. Kano, Y. Sumi, F. Masuda, M. Matsuo, H. Kadotani, Heart rate variability-based driver drowsiness detection and its validation With EEG, IEEE Trans. Biomed. Eng. 66 (6) (2019) 1769–1778, https://doi.org/10.1109/TBME.2018.2879346.

Further Reading

[1] C. Jacobé de Naurois, C. Bourdin, A. Stratulat, E. Diaz, J.-L. Vercher, Detection and prediction of driver drowsiness using artificial neural network models, Accid. Anal. Prevent. 126 (2019) 95–104, https://doi.org/10.1016/j.aap.2017.11.038.
[2] K. Lee, H. Yoon, J. Song, K. Park, Convolutional Neural Network-Based Classification of Driver's Emotion during Aggressive and Smooth Driving Using Multi-Modal Camera Sensors, Sensors, 18 (4) 957.
[3] J. Izquierdo-Reyes, R.A. Ramirez-Mendoza, M.R. Bustamante-Bello, S. Navarro-Tuch, R. Avila-Vazquez, Advanced driver monitoring for assistance system (ADMAS): Based on emotions, Int. J. Interact. Des. Manuf. 12 (1) (2018) 187–197, https://doi.org/10.1007/s12008-016-0349-9.
[4] E.E. Galarza, F.D. Egas, F.M. Silva, P.M. Velasco, E.D. Galarza (2018, January). Real time driver drowsiness detection based on driver's face image behavior using a system of human computer interaction implemented in a smartphone. In: International Conference on Information Theoretic Security (pp. 563-572). Springer, Cham.
[5] A. Garcés Correa, L. Orosco, E. Laciar, Automatic detection of drowsiness in EEG records based on multimodal analysis, Med. Eng. Phys. 36 (2) (2014) 244–249, https://doi.org/10.1016/j.medengphy.2013.07.011.
[6] J. Jo, S.J. Lee, K.R. Park, I.-J. Kim, J. Kim, Detecting driver drowsiness using feature-level fusion and user-specific classification, Exp. Syst. Appl. 41 (4) (2014) 1139–1152, https://doi.org/10.1016/j.eswa.2013.07.108.
[7] H. Gao, A. Yüce, J.P. Thiran, Detecting emotional stress from facial expressions for driving safety. In: 2014 IEEE International Conference on Image Processing (ICIP) 2014, October, (pp. 5961-5965). IEEE.
[8] S. Bando, A. Nozawa, Detection of driver inattention from fluctuations in vehicle operating data, Artif. Life Robotics 20 (1) (2015) 28–33, https://doi.org/10.1007/s10015-014-0191-8.
[9] S.S. Bharambe, P.M. Mahajan, Implementation of real time driver drowsiness detection system, Int. J. Sci. Res. (IJSR) 4 (1) (2015) 2202–2206.
[10] S.J., Jung, H.S. Shin, W.Y. Chung, Driver fatigue and drowsiness monitoring system with embedded electrocardiogram sensor on steering wheel. IET Intelligent Transport Systems, 2014, 8(1), 43-50.
[11] R.R. Johnson, D.P. Popovic, R.E. Olmstead, M. Stikic, D.J. Levendowski, C. Berka, Drowsiness/alertness algorithm development and validation using synchronized EEG and cognitive performance to individualize a generalized model, Biol. Psychol. 87 (2) (2011) 241–250, https://doi.org/10.1016/j.biopsycho.2011.03.003.