

Deep CNN: A Machine Learning Approach for Driver Drowsiness Detection Based on Eye State

Venkata Rami Reddy Chirra^{1,3*}, Srinivasulu Reddy Uyyala², Venkata Krishna Kishore Kolli³

¹ Department of Computer Applications, National Institute of Technology, Tiruchirappalli 620015, India

² Machine Learning & Data Analytics Lab, Department of Computer Applications, National Institute of Technology, Tiruchirappalli 620015, India

³ Department of Computer Science & Engineering, VFSTR, Guntur 522213, India

Corresponding Author Email: chvrr_cse@vignan.ac.in

<https://doi.org/10.18280/ria.330609>

Received: 24 September 2019

Accepted: 28 November 2019

Keywords:

viola-jones, stacked deep convolution neural network, SoftMax layer, CNN

ABSTRACT

Driver drowsiness is one of the reasons for large number of road accidents these days. With the advancement in Computer Vision technologies, smart/intelligent cameras are developed to identify drowsiness in drivers, thereby alerting drivers which in turn reduce accidents when they are in fatigue. In this work, a new framework is proposed using deep learning to detect driver drowsiness based on Eye state while driving the vehicle. To detect the face and extract the eye region from the face images, Viola-Jones face detection algorithm is used in this work. Stacked deep convolution neural network is developed to extract features from dynamically identified key frames from camera sequences and used for learning phase. A SoftMax layer in CNN classifier is used to classify the driver as sleep or non-sleep. This system alerts driver with an alarm when the driver is in sleepy mood. The proposed work is evaluated on a collected dataset and shows better accuracy with 96.42% when compared with traditional CNN. The limitation of traditional CNN such as pose accuracy in regression is overcome with the proposed Staked Deep CNN.

1. INTRODUCTION

In a car safety technology, driver drowsiness detection [1-3] is very essential to prevent road accidents. Now-a-days, many people using automobiles for daily commutation, higher living standards, comfortability, and timing constraints to reach destinations. This trend leads to high volumes of traffic in urban areas and highways. In turn, it will raise number of road accidents with several factors. Driver drowsiness could be the one reason for road accidents. One way to reduce number of accidents is early detection of driver drowsiness and alerting with an alarm. According to the NHTSA, every year around 1 lakh road accidents occurs because of driver drowsiness in the United States. NHTSA reported that 72,000 road accidents, 800 deaths and 44,000 injuries are occurred due to driver drowsiness. In 2017, around 1.47 lakh people are died due road accidents in India. Every year, over a Lakh people lost life due to road crashes and more than 4 times people get injured due to road accidents. In India average road accidents deaths are 1, 36,118 per year in last one decade. In 2016, 60% of people who lost their lives in road accidents were in age group of between in 18-35.

In India, since 2012 more than 500 people died due accidents on Yamuna express way and more than 100 people died due to vehicle crashes on Agra-Lucknow express way. Police officials and patrolling teams on these expressways revealed that most of the accidents are happened between 2 am and 5 am due to drivers drowsy-deprived. Drivers' sleep deprivation is major reason for accidents. So, technology for driver drowsiness detection system is required to reduce road accidents. The development of this technology is a big challenge for both an industrial and research community.

There are different signs of driver drowsiness can be observed while driving the vehicle such as in ability to keep eyes open, frequently yawning, moving the head forward etc. To determine the level of driver drowsiness various measures are used. These measures are Physiological Measures, Behavioral Measures and Vehicle-based Measures.

In physiological measures, Electrocardiography (ECG), Electroencephalography (EEG), and Electrooculogram (EOG) [4, 5] are used to access the driver's conditions. Even though these devices provide accurate results, because of their practical limitations, these are not widely accepted. In vehicle-based measures, drowsiness is analyzed based on steering wheel movements and braking patterns. These methods are reliant on the road nature and driving skills of the driver. Behavioral measures are based on person behavior instead of vehicle. Here smart camera is used to capture the driver information. Behavioral Measures are the best ways to detect driver drowsiness.

In our proposed method Behavioral Measures are used to detect driver drowsiness. Figure 1 shows the general architecture of driver drowsiness detection.

To identify the face regions from the input images, various face detection algorithms [6] have been used in the Face Detection phase. For human face detection task is easier, but this task is difficult in computer vision. Face detection techniques are classified into feature-based techniques and image-based techniques. Statistical, Neural Networks and Liner subspace methods have been used by the Image-based approaches for face detection. In the second step, different eye region detection algorithms were used to detect and extract the eye region from the face images. After locating facial regions, in the preprocessing stage normalization takes place to reduce

the effects of illumination. The contrast differences among face images can be adjusted by performing histogram equalization. In the third step, feature extraction was implemented on the input eye region images. There are two main methods for extracting features from images: appearance-based feature extraction and geometric-based extraction methods. The geometric extraction method extracts shape- and location-related metrics that are extracted from the eyes, eyebrows. Conversely, appearance-based feature extraction extracts skin appearance or facial features by implementing techniques such as PCA [7], Discrete Cosine Transform (DCT) [8] and Linear Discriminant Analysis (LDA). These methods can be applied on the entire face or particular regions for extracting the facial features of face. For extracting the local features of a face, Gabor wavelets can be used; however, the occurrence of high-dimensional feature vectors is the main problem with this method. In the fourth step of driver drowsiness detection is classification which uses a classifier to classify the sleeping and non-sleeping images based on the features extracted in the previous two steps. Staked Deep CNN is developed for classification of driver sleepy state.

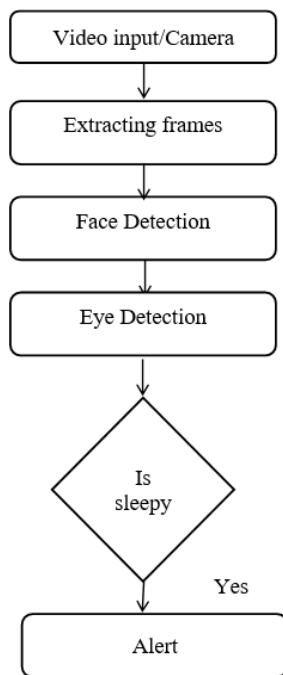


Figure 1. General architecture of drowsiness detection

2. RELATED WORK

In order to achieve better accuracy in detection of driver drowsiness system, many approaches were developed. Mardi et al., [4] have proposed a model to detect a drowsiness based on Electroencephalography (EEG) signals. Extracted chaotic features and logarithm of energy of signal are extracted for differentiating drowsiness and alertness. Artificial neural network was used for classification and yielded 83.3% accuracy. Noori et al., [5] proposed a model to find drowsiness based on fusion of Driving Quality Signals, EEG and Electrooculography. For selecting the best subset of features a class separability feature selection method was used. A self-organized map network was used for classification and achieved $76.51 \pm 3.43\%$ of accuracy. Krajewski et al., [9]

developed a model to find drowsiness based on steering patterns. In this model, to capture the steering patterns they generated three feature sets by using advanced signal processing methods. Performance is evaluated using five machine learning algorithms like SVM and K-Nearest Neighbor and achieved accuracy of 86% in detecting drowsiness.

Danisman et al. [10] developed a method to detect a drowsiness based on changes in eye blink rate. Here Viola Jones detection algorithm was used to detect face region from the images. Then neural network-based eye detector was used to find the location of the pupils. Calculated the no of blinks per minute if blinks increase indicated that driver becomes drowsy. Abtahi et al. [11] presented a method for drowsiness detection through yawning. In this method first face was detected then eye and mouth regions are detected. In mouth they calculated hole as result of wide mouth open. Face with largest hole indicates yawning mouth.

Dwivedi et al. [12] developed a model to find the drowsiness using CNNs. In this method convolutional neural network was used to capture the latent features then SoftMax layer was used for classification and yielded 78% of accuracy. Advanced Driver Assistance System was proposed by the Alshaqafi et al. [13] to minimize the road accidents due to driver drowsiness. Here an algorithm is proposed to locate, track and analyze face and eyes to measure PERCLOS for finding driver drowsiness. In this algorithm after detecting face, eye location detected then Hough transform for circles (HTC) method was used find eye state. If the state of the eye is closed more than 5 seconds, it is considered as drowsy. Park et al. [14] proposed a deep learning based network to find the driver drowsiness from the given input videos. Here three deep networks such as Alex Net, VGG-FaceNet and FlowImageNet are used for feature learning. In this paper experiments are performed on NTHU driver drowsiness video dataset and achieved around 73% accuracy.

Tadesse et al. [15] developed a method using Hidden Markov Model to detect the driver drowsiness. In this work, for extracting the face regions, Viola Jones algorithm was used. For extracting the features from the face regions Gabor wavelet decomposition was used. For selecting features Adaboost learning algorithm was used. HMM used for classification of drowsy or non-drowsy expression.

An Eye-tracking based driver drowsiness system was proposed by Said et al. [16]. In this work the system finds the driver's drowsiness and rings the alarm to alert to the driver. In this work, Viola Jones model was used to detect the face region and eye region. It has produced an accuracy of 82% in indoor tests and an accuracy of 72.8% for outdoor environment.

Picot et al. [17] used both visual activity and brain activity for detecting driver drowsiness. To monitor the brain activity a single channel EEG was used. Blinking and characterization are used to monitor the Visual activity. Blinking features are extracted using EOG. An EOG-based detector was created by merging these two features using fuzzy logic. This work was evaluated on dataset with twenty individual drivers and achieved an accuracy of 80.6%.

For bus driver monitoring, Mandal et al. [18] developed a vision-based fatigue detection system. In this work, A HOG and SVM are used for head-shoulder detection and driver detection respectively. Once driver is detected they used OpenCV face detector for face detection and OpenCV eye detector for eye detection. Spectral Regression Embedding

was used to learn the eye shape and a new method was used for eye openness estimation. Fusion was applied to fuse the features generated by two eye detectors I2R-ED and CV-ED. PERCLOS was calculated for drowsiness detection.

Jabbara et al. [19] introduced a model for detecting driver drowsiness based on deep learning for android applications. Here they designed a model which is based on the facial landmark point detection. Here first images are extracted from video frames then Dlib library was used to extract landmark coordinate points. These landmark coordinate points given as input to multilayer perceptron classifier. Classifier classify either drowsy or non-drowsy based these points. This method was evaluated on NTHU Drowsy Driver Detection Dataset and achieved accuracy of more than 80%.

3. PROPOSED DEEP CNN BASED DROWSINESS DETECTION SYSTEM

3.1 Proposed system algorithm

(1) Viola-jones face detection algorithm is used to detect the face the images and given as input to Viola-jones eye detection algorithm

(2) Once the face is detected, Viola-jones eye detection algorithm is used to extract the eye region from the facial images and given as input to CNN.

(3) CNN with four convolutional layers are used to extract the deep features and those features are passed to fully connected layer.

(4) Soft max layer in CNN classify the images in to sleepy or non-sleepy images.

The proposed architecture of Drowsiness detection system using Deep CNN is given in Figure 2. The proposed model has three phases 1. Preprocessing stage, 2. Feature extraction, 3. Deep CNN Classifier.

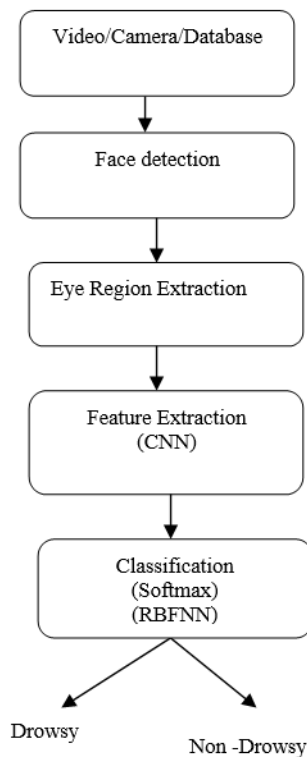


Figure 2. Proposed system architecture

3.2 Face detection and eye region extraction

Whole face region may not be required to detect the drowsiness but only eyes region is enough for detecting drowsiness. At first step by using the Viola-jones face detection algorithm face is detected from the images. Once the face is detected, Viola-jones eye detection algorithm is used to extract the eye region from the facial images. In 2001, P Viola and M Jones developed the Viola-Jones object detection algorithm [20, 21], it is the first algorithm used for face detection. For the face detection the Viola-Jones algorithm having three techniques those are Haar-like features, Ada boost and Cascade classifier. In this work, Viola-Jones object detection algorithm with Haar cascade classifier was used and implemented using OPEN CV with python. Haar cascade classifier uses Haar features for detecting the face from images. Figure 3 shows the Eye region images extracted from the face image.

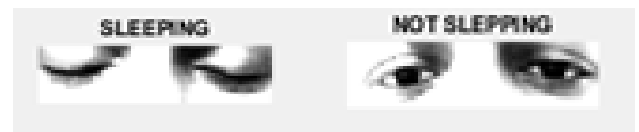


Figure 3. Eye region images

3.3 Feature extraction and classification

Feature extraction is one type of dimensionality reduction where useful parts of an image represented as a feature vector. In this paper features from the eye region images are extracted using a Convolutional Neural Network (CNN) [22-24].

3.3.1 Convolutional neural network

Convolutional neural network (CNN) is used in the proposed system for detection of driver drowsiness. Since a feature vector is needed for each drowsy image to compare with existing features in a database to detect either drowsy or not. Usually CNNs requires fixed size images as input so preprocessing is required. The preprocessing includes extracting the key frames from video based on temporal changes and store in database. From these stored images, feature vectors are generated in convolution layers of CNN. These feature vectors are then used for the detecting the driver drowsiness. CNN have layers like convolutional layers, pooling (max, min and average) layers, ReLU layer and fully-connected layer. Convolution layer is having kernels (filters) and each kernel having width, depth and height. This layer produces the feature maps as a result of calculating the scalar product between the kernels and local regions of image. CNN uses pooling layers (Max or Average) to minimize the size of the feature maps to speed up calculations. In this layer, input image is divided into different regions then operations are performed on each region. In Max Pooling, a maximum value is selected for each region and places it in the corresponding place in the output. **ReLU (Rectified Linear Units) is** a nonlinear layer. The ReLU layer applies the max function on all the values in the input data and changes all the negative values to zero. The following equation shows the ReLU activation function.

$$f(x)=\max(0,x) \quad (1)$$

The fully-connected layers used to produce class scores

from the activations which are used for classification.

3.3.2 Layers design of proposed deep CNN model

In our proposed work, a new Deep CNN model is designed

for detection of driver drowsiness using deep learning based on Eye State. Figure 4 shows the designed CNN model used in this work.

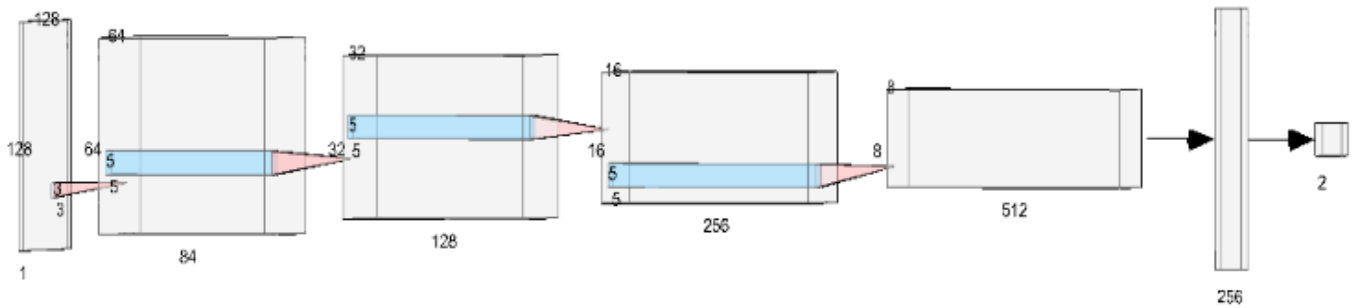


Figure 4. Proposed deep CNN model

In the proposed method, 4 convolutional layers and one fully connected layer are used. Extracted key images with size of 128 X 128 are passed as input to the convolution layer-1 (Conv2d_1). In Conv2d_1 input image is convolved with 84 filters of size 3x3. After convolution, batch Normalization, non-linear transformation ReLU, Max pooling over 2 × 2 cells are included in the architecture, which is followed by dropout with 0.25%. Conv2d_1 required 840 parameters. Batch_normalization_1 is done with 336 parameters. The output of convolution layer-1 is fed in to the convolution layer-2(Conv2d_2). In Conv2d_2, input is convolved with 128 filters with size 5x5 each. After convolution, batch Normalization, non-linear transformation ReLU, MaxPooling over 2 × 2 cells with stride 2 followed by dropout with 0.25% applied. Conv2d_2 required 268928 parameters. Batch_normalization_2 required 512 parameters. The output of convolution layer-2 is fed in to the convolution layer-3(Conv2d_3). In Conv2d_3, input is convolved with 256 filters with size 5x5 each. After convolution, Batch Normalization, non-linear transformation ReLU, MaxPooling over 2 × 2 cells with stride 2 followed by dropout with 0.25% applied, Conv2d_3 required 819456 parameters. Batch_normalization_3 required 1024 parameters.

The output of convolution layer-3 is fed in to the

convolution layer-4(Conv2d_4). In Conv2d_4 input is convolved with 512 filters with size 5x5 each. After convolution, Batch Normalization, non-linear transformation ReLU, Max Pooling over 2 × 2 cells with stride 2 followed by dropout with 0.25% applied. Conv2d_4 required 3277312 parameters. Batch_normalization_4 required 2048 parameters. Fully connected layer that is dense_1 required 8388864 parameters. Proposed CNN model required 12,757,874 trainable parameters. The output of classifier is two state, so output layer having only two outputs. Adam method is used for Optimization. Here softmax classifier is used for classification. In our proposed CNN framework, the 256 outputs of fully connected layer are the deep features retrieved from input eye images. The final 2 outputs can be the linear combinations of the deep features.

4. EXPERIMENTS AND ANALYSIS

Here we have performed two types of experiments. In First type, experiment is performed on collected dataset. In Second type, experiment is performed on video. To conduct first type of experiment, we have generated a dataset with 2850 images.



Figure 5. Non-drowsy and drowsy image samples from dataset

Table 1. Accuracy of the proposed model

Training Samples	Testing Samples	Validation Samples	Training Accuracy (%)	Validation Accuracy (%)	Testing Accuracy (%)
1200	1150	500	98	97	96.42

Few samples from the dataset are shown in Figure 5. Out of 2850 images, 1450 images are drowsy images and remaining are non-drowsy images. To conduct experiment, a total of 1200 images are used for training out of which 600 images are drowsy images and another 600 images are non-drowsy images. A total of 500 images are used for validation out of which 250 images are drowsy images and another 250 images are non-drowsy images. A total of 1150 images are used for testing out of which 550 images are drowsy images and another 600 images are non-drowsy images and proposed model has achieved an accuracy of 96.42% on test dataset. Table 1 shows accuracy of the proposed model after 50 epochs with batch size 4. The training loss and validation loss against number of epochs are shown in Figure 6. The training accuracy and validation accuracy against number of epochs are shown in Figure 7. Confusion matrix is shown in Figure 8.

In second type of experiment, first we have trained our model with 1200 samples.

During testing phase, we capture the video frames through camera and alert with an alarm when the model predicts drowsy output state continuously. Static images are used for training but during testing phase key frames are extracted from continuous video and tested against the trained static images. Experimental flow diagram is shown in Figure 9. Results in 2nd type of experiment are given in Figure 10.

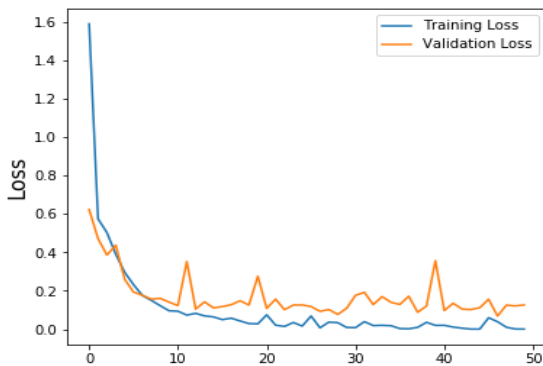


Figure 6. The training loss and validation loss against no of epochs



Figure 7. The training accuracy and validation accuracy against no of epochs

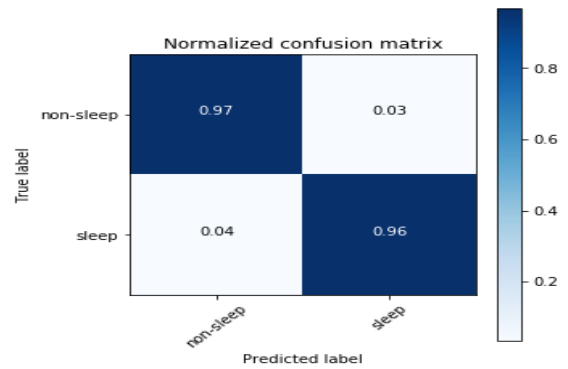


Figure 8. Confusion matrix

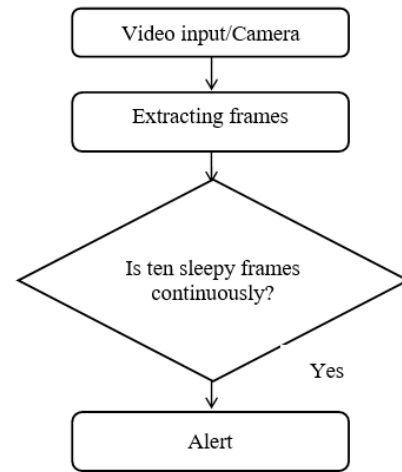


Figure 9. Experiment flow diagram

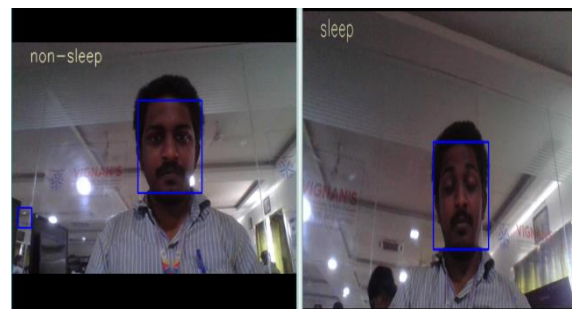


Figure 10. Results in 2nd type of experiment

5. CONCLUSIONS

In this proposed work a new method is proposed for driver drowsiness detection based on eye state. This determines the state of the eye that is drowsy or non-drowsy and alert with an alarm when state of the eye is drowsy. Face and eye region are detected using Viola-Jones detection algorithm. Stacked deep convolution neural network is developed to extract features and used for learning phase. A SoftMax layer in CNN classifier is used to classify the driver as sleep or non-sleep. Proposed system achieved 96.42% accuracy. Proposed system

effectively identifies the state of driver and alert with an alarm when the model predicts drowsy output state continuously. In future we will use transfer learning to improve the performance of the system.

REFERENCES

- [1] Amodio, A., Ermidoro, M., Maggi, D., Formentin, S., Savaresi, S.M. (2018). Automatic detection of driver impairment based on pupillary light reflex. *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-11. <https://doi.org/10.1109/tits.2018.2871262>
- [2] Yang, J.H., Mao, Z.H., Tijerina, L., Pilutti, T., Coughlin, J.F., Feron, E. (2009). Detection of driver fatigue caused by sleep deprivation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 39(4): 694-705. <https://doi.org/10.1109/tsmca.2009.2018634>
- [3] Hu, S., Zheng, G. (2009). Driver drowsiness detection with eyelid related parameters by support vector machine. *Expert Systems with Applications*, 36(4): 7651-7658. <https://doi.org/10.1016/j.eswa.2008.09.030>
- [4] Mardi, Z., Ashtiani, S.N., Mikaili, M. (2011). EEG-based drowsiness detection for safe driving using chaotic features and statistical tests. *Journal of Medical Signals And Sensors*, 1(2): 130-137. <https://doi.org/10.4103/2228-7477.95297>
- [5] Noori, S.M., Mikaeili, M. (2016). Driving drowsiness detection using fusion of electroencephalography, electrooculography, and driving quality signals. *J Med Signals Sens*, 6: 39-46. <https://doi.org/10.4103/2228-7477.175868>
- [6] Dang, K., Sharma, S. (2017). Review and comparison of face detection algorithms. *7th International Conference on Cloud Computing, Data Science & Engineering*, pp. 629-633. <https://doi.org/10.1109/confluence.2017.7943228>
- [7] VenkataRamiReddy, C., Kishore, K.K., Bhattacharyya, D., Kim, T.H. (2014). Multi-feature fusion based facial expression classification using DLBP and DCT. *Int J Softw Eng Appl*, 8(9): 55-68. <https://doi.org/10.14257/ijseia.2014.8.9.05>
- [8] Ramireddy, C.V., Kishore, K.V.K. (2013). Facial expression classification using Kernel based PCA with fused DCT and GWT features. *ICCIC, Enathi*, 1-6. <https://doi.org/10.1109/iccic.2013.6724211>
- [9] Krajewski, J., Sommer, D., Trutschel, U., Edwards, D., Golz, M. (2009). Steering wheel behavior based estimation of fatigue. *The Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, pp. 118-124. <https://doi.org/10.17077/drivingassessment.1311>
- [10] Danisman, T., Bilasco, I.M., Djeraba, C., Ihaddadene, N. (2014). Drowsy driver detection system using eye blink patterns. *2010 International Conference on Machine and Web Intelligence, Algiers, Algeria*, pp. 230-233. <https://doi.org/10.1109/icmwi.2010.5648121>
- [11] Abtahi, S., Hariri, B., Shirmohammadi, S. (2011). Driver drowsiness monitoring based on yawning detection. *2011 IEEE International Instrumentation and Measurement Technology Conference, Binjiang*, pp. 1-4. <https://doi.org/10.1109/imtc.2011.5944101>
- [12] Dwivedi, K., Biswaranjan, K., Sethi, A. (2014). Drowsy driver detection using representation learning. *Advance Computing Conference (IACC), IEEE*, pp. 995-999. <https://doi.org/10.1109/iadcc.2014.6779459>
- [13] Alshaquaqi, B., Baquhaizel, A.S., Amine Ouis, M.E., Boumehed, M., Ouamri, A., Keche, M. (2013). Driver drowsiness detection system. *8th International Workshop on Systems, Signal Processing and Their Applications (WoSSPA)*. <https://doi.org/10.1109/wosspa.2013.6602353>
- [14] Park, S., Pan, F., Kang, S., Yoo, C.D. (2017). Driver drowsiness detection system based on feature representation learning using various deep networks. In: Chen CS., Lu J., Ma KK. (eds) *Computer Vision – ACCV 2016 Workshops, Lecture Notes in Computer Science*, pp. 154-164. https://doi.org/10.1007/978-3-319-54526-4_12
- [15] Tadesse, E., Sheng, W., Liu, M. (2014). Driver drowsiness detection through HMM based dynamic modeling. *2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong*, pp. 4003-4008. <https://doi.org/10.1109/ICRA.2014.6907440>
- [16] Said, S., AlKork, S., Beyrouthy, T., Hassan, M., Abdellatif, O., Abdraboo, M.F. (2018). Real time eye tracking and detection- a driving assistance system. *Advances in Science, Technology and Engineering Systems Journal*, 3(6): 446-454. <https://doi.org/10.25046/aj030653>
- [17] Picot, A., Charbonnier, S., Caplier, A. (2012). On-Line Detection of drowsiness using brain and visual information. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 42(3): 764-775. <https://doi.org/10.1109/tsmca.2011.2164242>
- [18] Mandal, B., Li, L., Wang, G.S., Lin, J. (2017). Towards detection of bus driver fatigue based on robust visual analysis of eye state. *IEEE Transactions on Intelligent Transportation Systems*, 18(3): 545-557. <https://doi.org/10.1109/tits.2016.2582900>
- [19] Jabbar, R., Al-Khalifa, K., Kharbeche, M., Alhajyaseen, W., Jafari, M., Jiang, S. (2018). Real-time driver drowsiness detection for android application using deep neural networks techniques. *Procedia Computer Science*, 130: 400-407. <https://doi.org/10.1016/j.procs.2018.04.060>
- [20] Viola, P., Jones, M.J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2): 137-154. <https://doi.org/10.1023/b:visi.0000013087.49260.fb>
- [21] Jensen, O.H. (2008). Implementing the Viola-Jones face detection algorithm (Master's thesis, Technical University of Denmark, DTU, DK-2800 Kgs. Lyngby, Denmark).
- [22] O'Shea, K., Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*. <https://arxiv.org/abs/1511.08458v2>
- [23] Kim, K., Hong, H., Nam, G., Park, K. (2017). A study of deep CNN-based classification of open and closed eyes using a visible light camera sensor. *Sensors*, 17(7): 1534. <https://doi.org/10.3390/s17071534>
- [24] Lee, K., Yoon, H., Song, J., Park, K. (2018). Convolutional neural network-based classification of driver's emotion during aggressive and smooth driving using multi-modal camera sensors. *Sensors*, 18(4): 957. <https://doi.org/10.3390/s18040957>