www.czasopisma.pan.pl PAN POLSKA AKADEMIA NAUK www.journals.pan.pl

BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 66, No. 6, 2018
DOI: 10.24425/bpas.2018.125934

Detection of driver fatigue symptoms using transfer learning

J. CHMIELIŃSKA and J. JAKUBOWSKI*

Faculty of Electronics, Military University of Technology, 2 Witolda Urbanowicza St., 00-908 Warsaw, Poland

Abstract. This paper presents the results of the scientific investigations which aimed at developing the detectors of the selected driver fatigue symptoms based on face images. The presented approach assumed using convolutional neural networks and transfer learning technique. In the conducted research the pretrained model of AlexNet was used. The net underwent slight modification of the structure and then the fine-tuning procedure was applied with the use of an appropriate dataset. In this way all detectors of the selected fatigue symptoms were created. The results of conducted computations indicate that it is potentially possible to apply such an approach to the problem of fatigue symptom detection. The values of the overall misclassification rates for the most troublesome symptom are less than 5.5%, which seems to be a quite satisfactory result.

Key words: driver fatigue, convolutional neural networks, transfer learning, AlexNet.

1. Introduction

Driver fatigue is a relevant problem because it is one of the main causes of car accidents, which can have serious, or even fatal consequences. It is estimated that approximately 10 or even 20% of all car accidents happen with the participation of tired or drowsy drivers [1]. Therefore, development of driver monitoring systems is of the interest to many researchers all over the world. For more than two decades numerous algorithms of driver fatigue detection have been developed, which can be divided into three categories. First group consists of methods using biomedical signals such as EEG (electroencephalogram), ECG (electrocardiogram), EOG (electrooculogram), blood pressure or heart rate [2, 3]. These techniques give the most satisfactory results but require placing sensors on driver's body, which can be inconvenient and disturbing. Another category includes algorithms based on vehicle parameters such as steering wheel movements, speed changes or brake pedal pressure [4]. Such methods are absolutely nonintrusive but enable fatigue detection at the very late stage, while falling asleep. The last group of driver fatigue detection techniques uses cameras and various image processing and computer vision algorithms [5]. These methods probably have the biggest computing requirements but they are nonintrusive and enable drowsiness detection at its early stage.

2. Related work

The oldest systems of driver fatigue and drowsiness detection based on face image analysis use the so-called PERCLOS measure, which is defined as the percentage of eyelid closure over time [6]. It has been proved that while feeling sleepy, people blink more frequently and their eyes are closed longer than usually. Many publications deal with the problems of gaze direction detection and head pose recognition as the reliable indicators of driver drowsiness and inattention [7]. Another techniques apply image processing procedures to extract geometrical features, which can be indicators of tiredness or drowsiness. The estimation of the mouth openness, which enables yawning detection, is the most popular approach [8, 9]. Different methods use algorithms for automatic finding of fiducial point in face images to calculate geometric features, such as distances between certain points or angles created by lines connecting chosen points [10]. These features are used to describe changes in face expression while becoming tired or drowsy. They enable detection of some fatigue symptoms such as yawning, eye closing or eyebrow raising. Another techniques apply texture descriptors to chosen areas of driver's face in images to assess changes in face expression as the indicators of fatigue. Using LBP (local binary pattern) descriptor to this task is a typical approach [11]. Extracting characteristic features is also performed applying image filters whose parameters are established by an expert. In practice Gabor filters with different scales and orientations are often used [12].

Relatively new approaches use convolutional neural networks, whose application has become possible due to the development of deep learning algorithms. This methodology enables automatic extraction of characteristic features from images. Available scientific literature proves the effectiveness of convolutional neural networks in driver fatigue detection [13–18] and many other image recognition tasks. A great interest in this technique has led to the development of multilayer networks that are able to recognize between huge number of classes with outstanding results, for example AlexNet [19] or GoogleNet [20]. Such a good performance of these networks require a long-lasting training procedure. The newest image recognition techniques use pretrained convolutional neural networks able to perform a certain recognition task successfully and tune them up so that they can solve a different recognition problem. This methodology is called "transfer learning" and has been already used in various recognition task, e.g. medical image analysis [21], human action recognition [22], emotion recognition [23] and also driver monitoring systems.

*e-mail: jacek.jakubowski@wat.edu.pl

Manuscript submitted 2018-03-26, revised 2018-04-27, initially accepted for publication 2018-05-08, published in December 2018.

In [24] authors have used transfer learning to estimate driver gaze direction. They have compared the performance of two different pretrained convolutional neural network in classification of six various gaze zones and the state of closed eyes. Authors of [25] have developed driver drowsiness detection system integrating three pretrained networks fine-tuned to recognize four classes: lack of drowsiness and three drowsy states resulting in eye blinking, head nodding or yawning. Delivered results indicate that such a system performs better than each separate network.

3. Proposed approach

In this paper, transfer learning technique is used to detect selected but the most common driver fatigue symptoms. However, the assumed methodology differs from that described above. In presented approach each single symptom of tiredness or drowsiness is detected by a separate and dedicated classifier, which has been created by fine-tuning of a pretrained convolutional neural network. It means that the number of classifiers is the same as the number of detecting symptoms. Each classifier recognizes only two classes: the occurrence of a given symptom or its absence. Such an approach seems to simplify the classification task and results in high recognition accuracy. Besides, it enables detecting a few symptoms when they appear simultaneously. General idea of such an approach is depicted in Fig. 1 where the general view of a real-time application using the network discussed in this paper is presented.

In the conducted research the network called AlexNet was used as the starting net for the tuning process [19]. It has to be highlighted that presented methodology does not include assessment of the real level of fatigue, which should be performed on the basis of detection frequency of each fatigue symptom and also its duration.

4. AlexNet

AlexNet is the convolutional neural network that won the ImageNet Large Scale Visual Recognition Challenge, which is an annual competition in object recognition, in 2012. It was a spectacular success, because AlexNet achieved results about 10% better than other submitted approaches. This network was able to recognize images belonging to 1000 object categories and it was trained on a subset of ImageNet database consisting of more than million images [19]. The pretrained AlexNet model is delivered by MathWorks with Matlab environment. This implementation has been used by the authors of this paper. The exact architecture of the network is presented in Table 1.

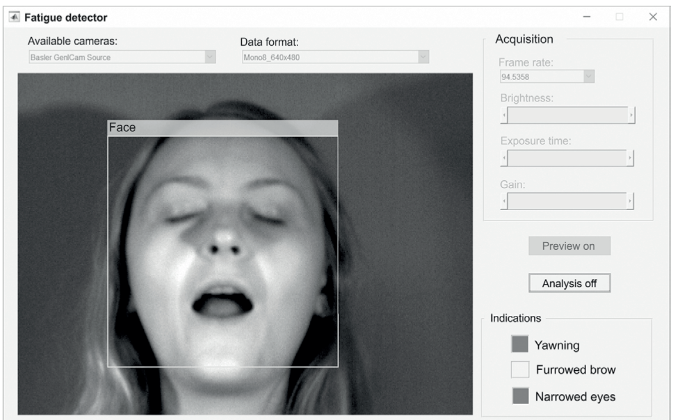

Fig. 1. General view of the real-time application developed with the transfer learning for the detection of selected fatigue symptoms

Table 1
The structure of AlexNet availabe in Matlab [26]

No.	Layer type	Description
1	Image Input	227×227×3 images with ‘zerocenter’ normalization
2	Convolution	96 11x11x3 convolutions with stride [4 4] and padding [0 0]
3	ReLU	–
4	Cross Channel Normalization	cross channel normalization with 5 channels per element
5	Max Pooling	3×3 max pooling with stride [2 2] and padding [0 0]
6	Convolution	256 5×5×48 convolutions with stride [1 1] and padding [2 2]
7	ReLU	–
8	Cross Channel Normalization	cross channel normalization with 5 channels per element
9	Max Pooling	3×3 max pooling with stride [2 2] and padding [0 0]
10	Convolution	384 3×3×256 convolutions with stride [1 1] and padding [1 1]
11	ReLU	–
12	Convolution	384 3×3×192 convolutions with stride [1 1] and padding [1 1]
13	ReLU	–
14	Convolution	256 3×3×192 convolutions with stride [1 1] and padding [1 1]
15	ReLU	–
16	Max Pooling	3×3 max pooling with stride [2 2] and padding [0 0]
17	Fully Connected	4096 fully connected layer
18	ReLU	–
19	Dropout	50% dropout
20	Fully Connected	4096 fully connected layer
21	ReLU	–
22	Dropout	50% dropout
23	Fully Connected	1000 fully connected layer
24	Softmax	–
25	Output	1000 classes

AlexNet consists of five convolutional layers and three fully-connected layers, whose parameters are established during the training procedure. In total there are 60 million parameters in the network, which makes the optimization procedure extremally time and resource consuming. Therefore, the training procedure usually requires using graphics processing unit for computation, which speeds up this process significantly [19]. Each convolutional layer is followed by ReLU (rectified linear unit) layer, which is a nonlinear activation function. This layer performs threshold operation. Any input value of this layer less than zero is set to zero, whereas positive values stay unchanged. It is proved that the application of ReLU as an activation function instead of its standard equivalents reduces the time it takes to train the neural network [27].

AlexNet also includes three pooling layers, which summarize the outputs of neighboring groups of neurons. These layers also reduce dimensions of feature maps, which influences the duration of the network training procedure. To avoid network overfitting to training data, the authors implemented dropout technique and local response normalization. The number of outputs of the last fully connected layer is equal to the number of categories recognized by the network. The next layer, called softmax layer, calculates the probability of classifying the input image to the particular class. This probability is the basis for making the final decision about the classification result [19]. The pretrained model of AlexNet available in Matlab environment requires input RGB images and their size has to be 227 by 227 pixels.

5. Methodology of the research

As it has been already mentioned, the aim of the conducted research was to develop the detectors of the chosen symptoms of fatigue, which are:
- yawning – a sign of drowsiness,
- furrowing the brow – a sign of a long-lasting need to stay focused,
- and narrowing or even closing the eyes – a sign of falling asleep.

Each detector is based on the pretrained AlexNet model available in Matlab environment and recognizes only two classes. Because the network model has 1000 outputs, it has to be slightly modified to fit the new task. Therefore, last three layers of pretrained AlexNet has to be substituted with new ones. The new fully-connected layer should have only two outputs, because there are only two classes – the symptom and lack of the symptom. The parameters of this layer has to be established in the training procedure, which is usually called fine-tuning while using transfer learning technique. The next two layers should be substituted because of the reduced number of outputs in the preceding layer, but they do not have learnable parameters.

All detectors were fine-tuned and tested using database collected by the authors. The base consists of labeled grayscale images of 19 individuals presenting simulated symptoms of fatigue and neutral facial expressions. All images were acquired with Basler acA2000–165umNIR camera whose detector is enhanced for near infrared spectrum. The camera was equipped with dedicated lens and band-passing filter fitting the spectrum of the near infrared illuminator. Such a choice of an image acquisition device and its accessories resulted from the necessity of monitoring drivers also in the night, when there is not enough visible light for traditional cameras. Each acquired image underwent the procedure of face detection with the use of Viola-Jones algorithm. Moreover, input images of developed fatigue symptoms detectors were limited to the regions including useful information. Fig. 2 shows the way of cropping the region of interest from the whole face images for each considered symptom of fatigue. Figures 3–5 present a subsets of images used for learning and testing the developed detectors. Input images of each detector had to be resized to fit input layer of the AlexNet model.

Research methodology assumed training and testing each detector for each individual separately. Therefore, in each evaluating session all images of one individual made the testing set and images of all other subjects were used to train the detector. In that way, the training and testing procedures were performed with 19 sets of training and testing data for each detector. More-

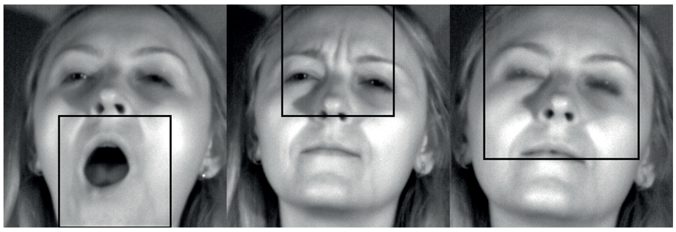

Fig. 2. Regions of interest for different symptoms (from the left: yawning, furrowing the brow and narrownig or closing the eyes)

Fig. 3. A subset of images of different individuals used to develop the detector of yawning

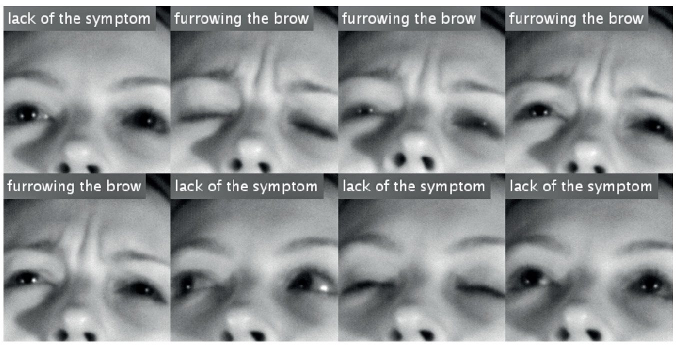

Fig. 4. A subset of images of a single individual used to develop the detector of furrowing the brow

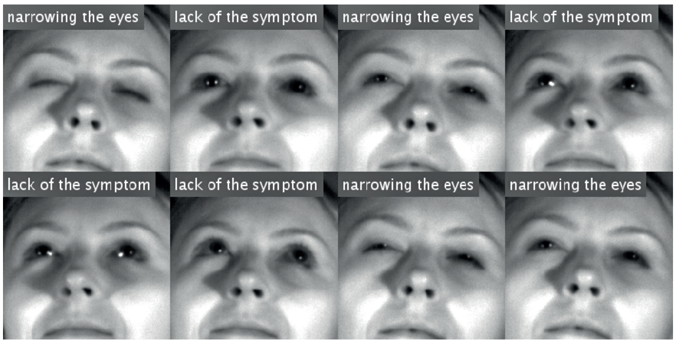

Fig. 5. A subset of images of a single individual used to develop the detector of narrowing or closing the eyes

over, for each individual the training and testing procedures were repeated a few times and the best obtained results have been presented in this paper. The highest testing accuracy was the measure of the result quality. It can be calculated as the quotient of correct detections and all analyzed cases. It can be expressed as a percentage or a fraction. Each training procedure was performed using 64 images in one iteration. To avoid overfitting, the training procedure was terminated if the mean accuracy in last 20 iterations was higher than 95%.

For the evaluation of the developed detectors, apart from the accuracy, the confusion matrix was used. It is a square matrix, whose rows refer to the exact classes of analyzed cases and columns present the decisions made by classifier. The number in a cell placed in the intersection of the i-th row and the j-th column means the number of cases belong to the i-th class and classified to the j-th class. This value can be also expressed as a relative value. It requires dividing the absolute value by the number of all cases of the i-th class. For binary classifier (with only two classes) the confusion matrix contains only four elements, which in this paper mean:

- absolute TN (true negative) or relative TNR (true negative rate) describing correct detections of a lack of a fatigue symptom,
- TP (true positive) or TPR (true positive rate) describing correct detections of a fatigue symptom,
- FN (false negative) or FNR (false negative rate) describing incorrect detections of a lack of a fatigue symptom
- and FP (false positive) or FPR (false positive rate) describing incorrect detections of a fatigue symptom.

6. Results

As it has been already mentioned, each detector was trained and tested using 19 sets of training and testing data. In this way, the results for each individual were obtained. It should be noticed that in all cases the training set did not contain images of a particular individual (all images of a given individual were used exclusively in testing). Such a methodology seems to be the most appropriate, because in practice a driver monitoring system should produce satisfactory results for individuals, whose images were not available while developing the detectors of the fatigue symptoms. Figures 6–8 illustrate the accuracy of training and testing obtained for each individual and all developed detectors. To simplify the comparison, the ranges of all vertical axes are the same. Besides, the minimum of each y-axis is not 0 but 85% to facilitate the analysis. It can be noticed that for each detector the values of training accuracy are similar for all individuals and are higher than 95%. On the other hand, the values of testing accuracy are in many cases slightly lower and depend on the fatigue symptoms and on the

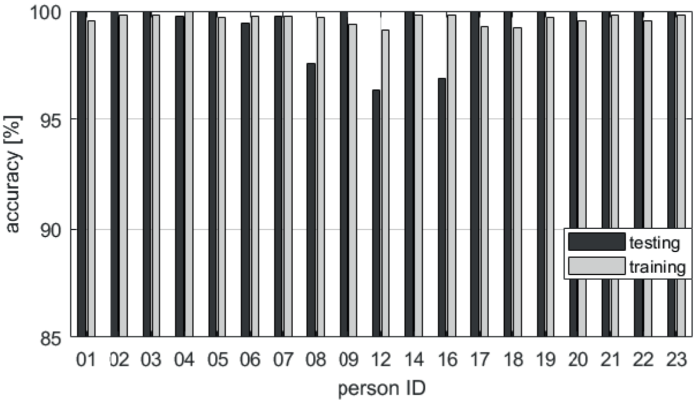

Fig. 6. Accuracy of training and testing obtained for each individual – detection of yawning

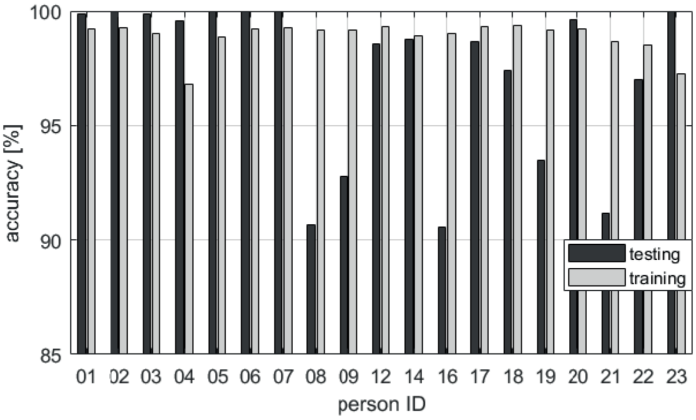

Fig. 7. Accuracy of training and testing obtained for each individual – detection of furrowing the brow

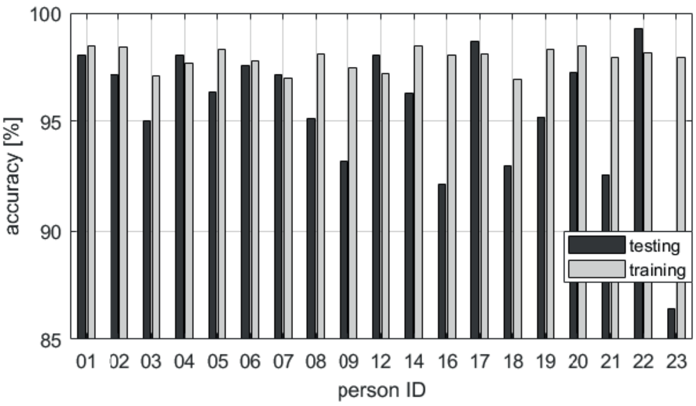

Fig. 8. Accuracy of training and testing obtained for each individual – detection of narrowing the eyes

individuals as well. Nevertheless, in all these cases except one, the accuracy of testing was over 90%.

However, accuracy is a global measure of the classifier quality and it does not show the distribution of the classification errors between the recognized classes. Therefore, the confusion matrices were also calculated in order to make a more precise and objective evaluation of the detectors. Table 2 presents the values of TPR and TNR of testing for each individual and for all developed detectors. It should be mentioned that three individuals were unable to simulate the symptom of furrowing the brow. In that cases the values of TPR (and also FNR) could not be determined, which is denoted as "–" in Table 2. A low value of TPR indicates that numerous instances of a particular fatigue symptom are undetected. Similarly, a low value of TNR denotes frequent incorrect detections of a particular symptom.

Table 2
The values of TPR and TNR of testing obtained for each individual

Person ID	Yawning		Furrowing the brow		Narrowing the eyes	
	***TPR* [%]**	***TNR* [%]**	***TPR* [%]**	***TNR* [%]**	***TPR* [%]**	***TNR* [%]**
1	100.0	100.0	100.0	99.8	96.2	99.5
2	100.0	100.0	100.0	100.0	94.6	99.7
3	100.0	100.0	99.7	100.0	96.1	94.3
4	99.4	100.0	–	99.6	97.1	99.1
5	100.0	100.0	–	100.0	93.2	100.0
6	100.0	99.4	100.0	100.0	96.2	99.7
7	99.1	100.0	100.0	100.0	95.8	98.7
8	96.3	97.7	80.3	100.0	93.7	95.9
9	100.0	100.0	61.2	100.0	90.5	94.5
12	85.2	100.0	97.4	99.8	94.7	98.6
14	100.0	100.0	98.1	99.5	99.8	92.3
16	94.2	100.0	100	86.2	89.3	100
17	100.0	100.0	97.3	100.0	99.8	97.6
18	100.0	100.0	94.4	99.8	99.8	85.1
19	100.0	100.0	87.6	100.0	97.2	92.2
20	100.0	100.0	99.2	100.0	91.5	99.3
21	100.0	100.0	85.5	97.3	91.6	94.2
22	100.0	100.0	100.0	94.4	100.0	98.5
23	100.0	100.0	–	100.0	71.4	99.5

On the basis of the absolute confusion matrices for all 19 sets of training and testing data, the overall confusion matrices were determined for each developed detector both for training data and for testing data. Each overall absolute confusion matrix was calculated as a sum of all obtained confusion matrices for a particular detector. Finally, for each detector two relative confusion matrices were computed, one for training data and the other for testing data. They are presented in Tables 3–5.

Table 3
The overall confusion matrices for training and testing data – yawning

Exact class	Result of recognition (testing)		Result of recognition (training)	
	Lack of the symptom	Yawning	Lack of the symptom	Yawning
Lack of the symptom	99.85%	0.15%	99.86%	0.14%
Yawning	1.72%	98.28%	0.57%	99.43%

Table 4
The overall confusion matrices for training and testing data – furrowing the brow

Exact class	Result of recognition (testing)		Result of recognition (training)	
	Lack of the symptom	Furrowing the brow	Lack of the symptom	Furrowing the brow
Lack of the symptom	99.30%	0.70%	99.44%	0.56%
Furrowing the brow	4.96%	95.04%	1.67%	98.33%

Table 5
The overall confusion matrices for training and testing data – narrowing the eyes

Exact class	Result of recognition (testing)		Result of recognition (training)	
	Lack of the symptom	Narrowing the eyes	Lack of the symptom	Narrowing the eyes
Lack of the symptom	96.65%	3.35%	97.93%	2.07%
Narrowing the eyes	5.41%	94.59%	2.10%	97.90%

7. Conclusions

The results presented in the previous section indicate that it is potentially possible to apply the proposed approach based on transfer learning to the problem of the fatigue symptom detection. The obtained overall statistics of misclassifications seem to be quite satisfactory, especially while taking into account the fact that all data of the individual whose images were used in testing were excluded from the training dataset. The best results were obtained for the developed detector of yawning. The level of errors for testing data in this case is less than 2%. The other two detectors produced worse results. The misclassification rates of testing for the detector of furrowing the brow and the detector of narrowing the eyes were estimated at about 5% and 5.5% respectively. It seems to be understandable as the yawning symptom is easily visible in images. Other symptoms are not so clear evidence of fatigue even for a human expert because there are people with natural face expressions containing slightly furrowed brow and narrowed eyes. The presented results also show that TNRs were higher than TPRs for all developed detectors. It means that the incorrect detection of a particular symptom, which is also called a false alarm, is less probable than the opposite misclassification.

It should be remembered that presented methodology does not include assessment of the real level of fatigue caused mostly by sleep deprivation that can seriously impair drivers. Finding the indications of drowsiness other that these presented in this paper is of paramount importance because of people that are getting drowsy without the visible symptoms discussed here. So this is natural step of the next research. It could be very difficult to generate features in above cases when even a human expert cannot recognize them but the convolutional networks with their automatic feature generation seem to be a very promising solution.

Acknowledgements. The work was co-financed by the Polish National Centre for Research and Development as a part of the project No. PBS3/B9/29/2015.

References

[1] P. Razin, M. Kruszewski, M. Niezgoda, and T. Kamiński, "Selected methods of driver fatigue detection", *Warsaw University of Technology Research Papers – Transport*, vol. 114, pp. 291–302 (2016) [in Polish].

[2] H.J. Eoh, M.K. Chung, and S.H. Kim, "Electroencephalographic study of drowsiness in simulated driving with sleep deprivation", *International Journal of Industrial Ergonomics* 35 (4), 307–320 (2005).

[3] X. Zhu, W.L. Zheng, B.L. Lu, X. Chen, S. Chen, and Ch. Wang, "EOG-based drowsiness detection using convolutional neural network", *International Joint Conference on Neural Networks*, Beijing, China, 2014.

[4] S.Y. Shi, W.Z. Tang, and Y.Y. Wang, "A review on fatigue driving detection", *ITM Web of Conferences*, vol. 12 (2017).

[5] M.H. Sigari, M.R. Pourshahabi, M. Soryani, and M. Fathy, "A review on driver face monitoring systems for fatigue and distraction detection", *International Journal of Advanced Science and Technology* 64, 73–100 (2014).

[6] R. Grace, V.E. Byrne, D.M. Bierman, J.-M. Legrand, D. Gricourt, B.K. Davis, J.J. Staszewski, and B. Carnahan, "A drowsy driver detection system for heavy vehicles", *Proceedings of the 17th Digital Avionics Systems Conference*, I36 1–8, 1998.

[7] I.H. Choi and Y.G. Kim, "Head pose and gaze direction tracking for detecting a drowsy driver", *Applied Mathematics and Information Sciences* 9 (2L), 505–512 (2015).

[8] Y. Lu and Z. Wang, "Detecting driver yawning in succesive images", *The 1st International Conference on Bioinformatics and Biomedical Engineering*, USA, 2007.

[9] N. Alioua, A. Amine, and M. Rziza, "Driver's fatigue detection based on yawning extraction", *International Journal of Vehicular Technology*, vol. 2014 (2014).

[10] T. Nakamura, A. Maejima, and S. Morishima, "Detection of driver's drowsy facial expresion", *Second IAPR Asian Conference on Pattern Recognition*, 2013.

[11] Y. Zhang and C. Hua, "Driver fatigue recognition based on facial expression analysis using local binary patterns", *Optik – international Journal for Light and Electron Optics* 126 (23), 4501–4505 (2015).

[12] X. Fun, B.C. Yin, and Y.F. Sun, "Dynamic human fatigue detection using feature-level fusion", in *Image and Signal Processing. ICISP 2008. Lecture Notes in Computer Science,* vol. 5099, eds. A. Elmoataz, O. Lezoray, F. Nouboud, D. Mammass, Springer, Berlin, Heidelberg, 2008.

[13] K. Yuen, S. Martin, and M.M. Trivedi, "On looking at faces in an automobile: issues, algorithms and evaluation on naturalistic driving dataset", *23rd International Conference on Pattern Recognition*, Mexico, 2016.

[14] S. Ribarić, J. Lovrenčić, and N. Pavešić, "A neural-network-based system for monitoring driver fatigue", *15th IEEE Mediterranean Electrotechnical Conference – MELECON 2010*, Malta, 2010.

[15] Ch. Yan, H. Jiang, B. Zhang, and F. Coenen, "Recognizing driver inattention by convolutional neural networks", *8th International Congress on Image and Signal Processing*, China, 2015.

[16] K. Dwivedi, K. Biswaranjan, and A. Sethi, Drowsy driver detection using representation learning, *IEEE International Advance Computing Conference*, India, 2014.

[17] X.P. Huynh, S.M. Park, and Y.G. Kim, "Detection of driver drowsiness using 3D deep neural network and semisupervised gradient boosting machine", in *Computer Vision – ACCV 2016 Workshops. ACCV 2016. Lecture Notes in Computer Science*, vol. 10118, eds. C.S. Chen, J. Lu, K.K. Ma, Springer, Cham, 2017.

[18] J. Chmielińska and J. Jakubowski, "Application of convolutional neural network to the problem of detecting selected symptoms of driver fatigue, *Electric Review* 93 (10), 6–10 (2017) [in Polish].

[19] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks", *Neural Information Processing Systems*, 2012.

[20] Ch. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions", *Computer Vision and Pattern Recognition* (2015).

[21] Ch.K. Shie, Ch.H. Chuang, Ch.N. Chou, M.H. Wu, and E.Y. Chang, "Transfer representation learning for medical image analysis", *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2015.

[22] A.B. Sargano, X. Wang, P. Angelov, and Z. Habib, "Human action recognition using transfer learning with deep representations", *International Joint Conference on Neural Networks*, 2017.

[23] H.W. Ng, V.D. Nguyen, V. Vonikakis, and S. Winkler, "Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning", *International Conference on Multimodal Interaction*, 2015.

[24] S. Vora, A. Rangesh, and M.M. Trivedi, "On generalizing driver gaze zone estimation using convolutional neural networks", *IEEE Intelligent Vehicles Symposium*, USA, 2017.

[25] S. Park, F. Pan, S. Kang, and C.D. Yoo, "Driver drowsiness detection system based on feature representation learning using various deep networks", in *Computer Vision – ACCV 2016 Workshops. ACCV 2016. Lecture Notes in Computer Science*, vol. 10118, eds. C.S. Chen, J. Lu, K.K. Ma,Springer, Cham, 2017.

[26] Matlab R2017a documentation – Neural Network Toolbox.

[27] X. Glorot, A. Bordes, Y. Bengio, "Deep Sparse Rectifer Neural Networks", *Proceedings of the 14th International Conference on Artifcial Intelligence and Statistics (AISTATS)*, Fort Lauderdale, FL, USA, pp. 315–323 (2011).