Detecting Driver Drowsiness as an Anomaly Using LSTM Autoencoders

Gülin Tüfekci[1,2,★], Alper Kayabaşı[1,2,★], Erdem Akagündüz[3], and İlkay Ulusoy[2]

[1] Research Center, Aselsan Inc
[2] Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, Turkey
[3] Graduate School of Informatics, Middle East Technical University, Ankara, Turkey
{gulin.tufekci,alper.kayabasi,akaerdem,ilkay}@metu.edu.tr

Abstract. In this paper, an LSTM autoencoder-based architecture is utilized for drowsiness detection with ResNet-34 as feature extractor. The problem is considered as anomaly detection for a single subject; therefore, only the normal driving representations are learned and it is expected that drowsiness representations, yielding higher reconstruction losses, are to be distinguished according to the knowledge of the network. In our study, the confidence levels of normal and anomaly clips are investigated through the methodology of label assignment such that training performance of LSTM autoencoder and interpretation of anomalies encountered during testing are analyzed under varying confidence rates. Our method is experimented on NTHU-DDD and benchmarked with a state-of-the-art anomaly detection method for driver drowsiness. Results show that the proposed model achieves detection rate of 0.8740 area under curve (AUC) and is able to provide significant improvements on certain scenarios.

Keywords: Driver Drowsiness Detection, LSTM Autoencoder, Video Anomaly Detection

1 Introduction

Road accidents have unrecoverable outcomes affecting the lives of many human beings. According to studies, the primary cause in road accidents is indicated as the human factor [21]. Human factor is defined as inattention, cognitive distractions and improper lookout; which is the inadequate actions of the driver not paying attention to the traffic and stimuli in driving environment including the road and other vehicles, or paying attention for short periods of time. Researchers have been focusing on detecting driver drowsiness/distraction from various data sources. For example, placing sensors on the driver to gather physiological patterns [2], [14] such as brain activity, variability of heartbeat, thermal imaging, respiration pattern give clues about driver's attention; yet, may affect the driver

---

★ These authors contributed equally to this work

psychologically and mislead the results. Another data source is monitoring the vehicle state such as detecting lane changes or steering wheel dynamics [1], etc. One of the most preferred method is to monitor the driver's face, eyes, mouth, body, hands using a camera that may give substantial clues about driver's state [22].
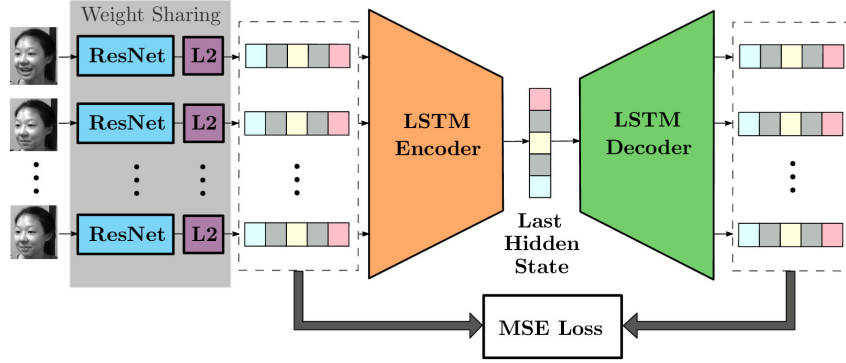

Fig. 1. Architecture of the proposed method

In this work, our aim is to detect driver drowsiness using a frontal camera. Detection of inattentive driver behavior is a time series problem, which is convenient for recurrent neural networks to extract hidden dependencies between time intervals. Since our approach is treating drowsy behavior as anomaly, we propose utilizing LSTM autoencoders [20] as LSTMs successfully catch the long-term dependencies while making use of short-term relations. To the best of our knowledge, there are no studies that adopt LSTM autoencoders to detect driver drowsiness by treating this problem as an anomaly. Because our approach is based on anomaly detection rather than binary classification; if attentive driving is considered as normality, drowsy driving should be qualified as an anomaly. Unlike classification, where both positive and negative samples are used to train a model, only the clips that are labeled as *normal* are presented to the network during training in order to learn a distribution of normal behavior and expected to produce *anomaly* output whenever out-of-distribution data is present. We believe that this problem definition is more relevant to real-life scenarios, mainly because the inattentive or drowsy behavior of a driver can be mostly idiosyncratic. Collecting a training set of *abnormal* behavior may never completely represent the universal set of inattentive or drowsy drivers.

In order to clarify the problem definition for drowsy and inattentive driver behavior using a frontal camera, we focus on the existing datasets and raise some important questions such as: *"how to define the abnormality degree throughout a time series data?"*, or - *"if it is adequate to label a clip as "normal" if majority of the labels belonging to the frames of the clip are normal?"* or *"how the*

network's performance is affected when half-confident data (having at least half of its frames labeled as normal) or full-confident data (having all of its frames labeled as normal) are used in training?". Because the network updates itself according to the normal data that it is encountered to, the confidence level of the training data has a crucial role on specifying the knowledge of the network. There are different approaches in the literature to label the clips such as using the label of the middle frame [10] or the most frequent one [24]. In most studies, the labeling method is not revealed. Therefore, we believe that it is necessary to observe the effect of confidence of the clip through labeling.

In our study, the experiments are conducted on NTHU Driver Drowsiness Dataset (NTHU-DDD) [23]. The proposed architecture, provided in Fig. 1, contains ResNet-34 [6] which is fine-tuned on NTHU-DDD as feature extractor and LSTM autoencoder to reconstruct the representations. The reconstruction loss is used for determining whether a clip is normal or drowsy according to a specified threshold. In order to enhance the quality of the frames, CLAHE [13] algorithm is applied. To find a place for the proposed method in the anomaly detection literature, a state-of-the-art method is trained on NTHU-DDD as a benchmark.

Contrast Limited Adaptive Histogram Equalization

The contributions of this work are twofold:

- LSTM autoencoders are experimented on NTHU-DDD, which is, to the best of our knowledge, the first time they are utilized for drowsiness detection using a frontal camera. State-of-the-art results are reached with learning only the *normal* representations in an anomaly detection task.
- Training strategies regarding the confidence levels of normal clips used in training are investigated and analyzed through recall and precision obtained by defining different normality and anomaly rates during test. Our experiments show that when the model is trained with low confident normal data, retrieval ability increases while reliability decreases; and vice versa for the model trained with high confident data.

2   Related Work

Prior to deep-learning era, hand-crafted machine learning methods were adopted to capture informative features about drowsiness such as eye blink rate [22], eye opening [5], head orientation angle [4]. With experience in deep neural networks, [15] fed the patches based on facial landmarks that are localized by MTCNN [25] to distinct networks to detect drowsiness on their own dataset. Shih et al. [18] followed a multi stage training approach for Modified VGG-16 [19] and LSTM [7] units on NTHU-DDD [23], taking 2nd place in the ACCV 2016 Workshop. The winner of ACCV 2016 competition [9] used CNN together with XGBoost [3] and semi supervised learning concept. [17] also utilized MTCNN [25] to obtain facial landmarks; creating a multi-feature architecture for frames and optical flows while adopting CLAHE [13] and Squeeze and Excitation [8] modules to enhance performance.

Anomaly detection in literature can be categorized into 3 groups; which are memory-based methods [12], one-class classification methods [16] and current

frame reconstruction [11]. Models in the first group learn prototypes that are proxies for different normal cases, which act as basis that spans normal space so anomaly samples tend to be orthogonal to such normal space. Second group aims to find the optimal hyperplane that separates normal samples from the abnormal ones. Third group is trained to reconstruct inputs from normal cases or predicts the next normal case given previous normal cases so they are expected to be in capable of reconstructing or predicting anomaly samples. Our approach is a member of last group and anomalies occurring for single subject in a steady environment are in scope of this work. The most similar approach to our method is the work of Kopuklu et al. [10] that proposed detecting drowsiness under the concept of anomaly detection, utilizing metric learning to get the representation for normal driving and making predictions by thresholding similarity between the concerned sample and prototypes for different modalities and views on their published Driver Anomaly Dataset.

3   The Dataset

NTHU-DDD [23] consists of videos belonging to 36 subjects from different ethnicities. The subjects are asked to play a driving game while pretending drowsy related actions, such as nodding, yawning, sleepiness, slow blinking, occasionally under different illumination and glasses scenarios. Drowsy frames are represented with binary labels where 1 corresponds to anomalous in our case. The videos are recorded using an infrared camera at 30 fps [23]. Multi Task Cascaded Convolutional Neural Network [25] is used for detecting faces in the dataset and cropping to eliminate redundant information.

In order to detect drowsiness on NTHU-DDD, normal and anomaly clips should be obtained. There are videos of 18 and 4 subjects in the training and validation sets, respectively. However, as the dataset was published for a competition, the test set is not available. The published training set of 18 subjects is divided into training and validation sets having 12 and 6 subjects respectively. The published validation set of 4 subjects is used as test set where there exists 5 videos belonging to different illumination and glasses conditions per subject. It is ensured that all sets have different subjects to prevent the network to memorize. All sets are divided into clips having 48 frames with a rate of 2 and window stride 23; so each clip is approximately 3 seconds long. By this way, around 7000 clips are obtained for the test set and around 52% of the clips represents the anomalous case, which creates a balanced set.

4   Methodology

ResNet-34 [6] which is pre-trained on ImageNet is fine-tuned on NTHU-DDD [23], which is a benchmark for driver drowsiness detection task, by replacing the FC layer with 2 concatenated FC layers having 128 and 2 nodes to classify the images under drowsiness label. Then, the FC layers are eliminated to obtain 512 dimensional extracted representations, which are L2 normalized

to create a regularized space. Let us denote this feature extraction process as $B(\cdot) : R^{224 \times 224 \times 3} \rightarrow R^{512}$ that maps $I(t)$ to $F(t)$. LSTM autoencoder reconstructs extracted features corresponding to each image, $I(t)$, in the normal clip consisting N frames, into $\hat{I}(t)$. The encoder of LSTM autoencoder, $LSTM_E$, consists of 2 layered LSTM that iteratively encodes the extracted the features to get holistic representation of spatio-temporal context of the clip at its last hidden state. Obtained context is constantly fed into $LSTM_D$, consisting 2 layered LSTM, that is responsible of decoding sequence in reverse order as [20] suggests. The reconstruction loss between the input and the output of the LSTM autoencoder is computed using Mean Squared Error loss for normal clips and weights of encoder and decoder are updated to minimize the clip loss as shown in (1) and (2).

$$L_{con} = \sum_{t=1}^{N} \|F(t) - \hat{I}(N - t + 1)\|^2 \tag{1}$$

$$\hat{I}(t) = LSTM_D \left(LSTM_E \left(F(t) \right) \right) \tag{2}$$

Since anomaly detection approach is adopted, only normal clips are used in training. To assign clip labels, normal confidence rate is defined as the controlling parameter. If the ratio of normal labeled frames over number of total frames in a clip is larger than or equal to the specified normal confidence rate, then the clip is labeled as normal. For example, with rate $1/2$, at least half of the frames should be normal; generating low confident clips. With rate 1, the clip is expected to have all of its frames labeled as normal. The clips span 3 seconds of data which is sufficient to observe drowsy behaviors through such intervals. Since drowsiness clues have lingering characteristics instead of fluctuations throughout frames, the lowest normal rate value is selected as $1/2$. It is reasonable to expect that the performance should increase when more confident clips are used in the training; yet, it may also mislead the network since the network learns the normal representation strictly. Hence, the performance is observed through training the model with normal rates $1/2$, $2/3$ and 1.

During testing, the network is subjected to anomalous clips as well as normal clips. Since the normal representations are learned throughout training, the network is expected to produce low reconstruction loss values for normal clips. However, the network is unaware of the anomaly behaviors and is expected not to successfully reconstruct the presented anomaly representations. Therefore, by applying a threshold which maximizes AUC of the corresponding receiver operating character (ROC), the test clip is predicted as normal or anomalous. In addition, anomaly confidence rate is defined so that the performance of the LSTM autoencoder can be observed under anomaly testing clips with different confidences. Anomaly rates are also tested with $1/2$, $2/3$ and 1.

Table 1. AUC for different CLAHE settings on NTHU-DDD

CLAHE Limit	-	10	5	5
CLAHE Grid Size	-	8	8	16
AUC	0.8240	0.8058	**0.8735**	0.8506

5   Experiment Details

During training, frames are resized to 224×224, randomly flipped horizontally, normalized by the mean and deviation for fine-tuning ResNet-34 and training the LSTM autoencoder. For fine-tuning ResNet-34 and training LSTM autoencoder; learning rates are 0.001 and 0.01, batch sizes are 128 and 4, optimizers are Adam with weight decay of 0.001 and SGD, respectively. Since NTHU-DDD consists of videos captured in night conditions, CLAHE [13] algorithm is used for enhancing the contrast of the frames, so the pre-trained backbone ResNet-34 can extract high quality features. CLAHE is a histogram equalization technique which applies the contrast enhancement within the specified grids locally instead of using the whole image. Therefore, overexposure or underexposure caused by equalization is prevented. In the experiments, the effect is observed by applying CLAHE with different limits (5, 10) and grid sizes (8, 16), as well as without the application of CLAHE. The results regarding various normal and anomaly rates are cross compared to interpret the detection performance of the network through AUC, accuracy, recall, precision and F1.

6   Results

For deciding CLAHE parameters, normal and anomaly rates are set to 1/2; which is equivalent to taking the most frequent label as clip label. The AUC results are provided in Table 1; which show that applying CLAHE is beneficial since it increases the visual perception of the network. CLAHE limit of 5 is a better contrast limit for NTHU-DDD for grid size 8. With CLAHE limit 5, the highest AUC is achieved with grid size 8.

The test results for varying normal and anomaly confidence rates are provided in Table 2. As we move towards more confident normal clips while keeping anomaly rate constant (moving towards right in rows); since representations are learnt in a relatively narrow space in a stricter sense during training, the perception ability of confident normal clips of the network increases. This effect is visible in precision; since precision defines how accurate the positive predictions are, requiring minimizing the false positives. However, when it comes to recall, the situation differs. The ability of detecting positive data (drowsy clips) of the network decreases towards right. We speculate that using low confident clips increases capacity of model to reconstruct positive examples at the expense of false positives. Therefore, in order to increase recall, using low confident clips in training is beneficial. Accuracy increases towards left; which provides the information

Table 2. Drowsiness detection results on NTHU-DDD

		AUC			Accuracy (%)			Recall (%)			Precision (%)			F1 (%)		
		Normal Rates			Normal Rates			Normal Rates			Normal Rates			Normal Rates		
		1/2	2/3	1	1/2	2/3	1	1/2	2/3	1	1/2	2/3	1	1/2	2/3	1
Anomaly Rates	1/2	0.8740	0.8758	0.8782	81.58	81.48	81.44	79.83	79.56	78.10	83.39	84.06	**85.41**	81.81	81.75	81.60
	2/3	0.8749	0.8768	0.8792	81.65	81.56	81.57	79.95	79.10	78.31	83.78	84.39	85.33	**81.82**	81.66	81.67
	1	0.8763	0.8782	**0.8806**	**81.79**	81.71	81.78	**80.18**	79.34	78.65	83.51	84.13	85.10	81.81	81.66	81.74

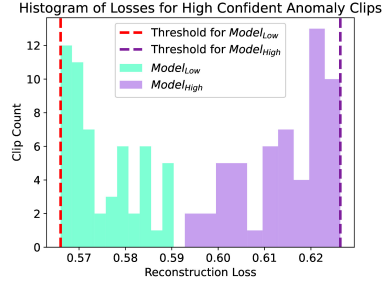(a)   (b)
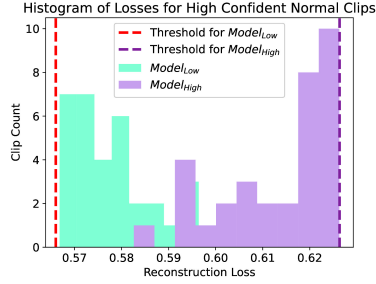

(c)   (d)

Fig. 2. Example frames from completely normal **(a)** and completely anomaly clip **(b)** Histogram of reconstruction losses for completely normal **(c)** and completely anomalous clips **(d)**. (Best viewed in zoom)

that the network predicts more accurate but at the same time not providing any information about false positives or false negatives. Accuracy values almost remain constant for varying confidence rates; therefore, it is not proper to make a conclusion using accuracy unlike AUC, which reflects performance independent of threshold value. AUC shows that with high confident clips, false positive rate decreases while true positive rate increases. It is expected since the predictions are made for more obvious clips. As we move towards more confident anomaly clips while keeping normal rate constant (moving towards bottom in columns); the general observation is reconstruction ability of the LSTM autoencoder gets better. Since more confident anomaly clips are easier for the network to interpret, the performance tends to increase.

Models trained with low and high normal confidence rates are denoted by $Model_{Low}$ and $Model_{High}$ respectively. Controlled experiments are performed for analyzing the predictions of $Model_{Low}$ and $Model_{High}$ on cases where completely normal and completely anomalous clips exist. To do so, reconstruction
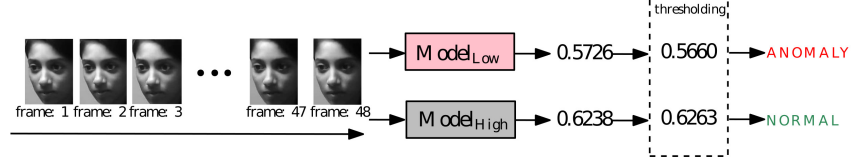
Fig. 3. Evaluation of high confident normal clip by $Model_{Low}$ and $Model_{High}$
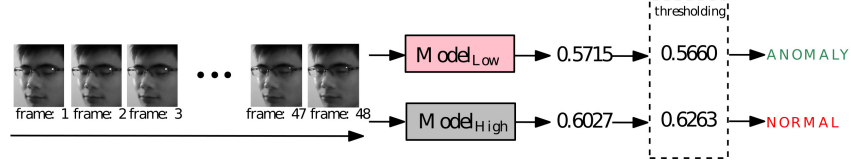
Fig. 4. Evaluation of high confident anomalous clip by $Model_{Low}$ and $Model_{High}$

losses are provided as histogram plots for two separate groups as shown in Fig. 2. The first group aims to reveal the reliability of predictions for $Model_{Low}$ and $Model_{High}$ with experiments on the same completely normal clips as shown in Fig. 2 (a). As the histogram suggests in Fig. 2 (c), $Model_{High}$ predicts the clips correctly while $Model_{Low}$ mistakes the same normal clips for anomaly. This outcome is consistent with our observation that with the model trained using low confident data, false positive rate increases even when it is subjected to completely normal data. This situation is present for 33 high confident normal clips and one example is interpreted in Fig. 3. The frames of a completely normal clip is provided in the left side of the figure. $Model_{Low}$ gives an anomaly score of 0.5726 while $Model_{High}$ gives 0.6238. When these outcomes are thresholded accordingly, $Model_{Low}$ predicts the clip as anomaly (false positive case) whereas $Model_{High}$ predicts normal (true negative case). This case shows that even though $Model_{Low}$ learns the representations in a broader sense when compared to $Model_{High}$, there are still some cases where it fails to detect such high confident normal clips.

The second group, provided in Fig. 2 (d), compares the retrieval ability of $Model_{Low}$ and $Model_{High}$ on entirely anomalous clips as shown in Fig. 2 (b). In second group, $Model_{High}$ misses the samples and results in false negative predictions while $Model_{Low}$ interprets the same clips successfully. There are 55 high confident anomaly clips in the test set and one example is interpreted in Fig. 4. The frames of a completely anomalous clip is provided in the left side of the figure. $Model_{Low}$ and $Model_{High}$ output scores of 0.5715 and 0.6027 respectively. When these values are thresholded accordingly, $Model_{Low}$ predicts the clip as anomaly (true positive case) whereas $Model_{High}$ predicts normal (false negative case). This case shows that although $Model_{High}$ is trained to recognize completely normal clips, it may still fail to reject such high confident anomalous clip.

Table 3. Comparison of [10] with our method on NTHU-DDD [23]

Method	AUC	Accuracy
[10] Before Post-Processing	0.8143	74.04
[10] After Post-Processing	0.8169	74.32
Ours	0.8740	81.79

Table 4. Comparison of [10] with our method on DAD dataset [10]

Method	AUC	Accuracy
[10] After Post-Processing	0.8737	-
Ours	0.8434	74.10

Throughout varying the confidence rates, F1 scores turn out to be almost same. Hence, as a training strategy, if the aim is not to miss anomalies without the concern of increase in the false positive rate, then it is appropriate to utilize low confident clips while leaning on the recall rate. If the application area is strict about lowering the false positive rate such that only the true positives should be predicted, then precision metric is more proper to make an evaluation while using high confident clips.

In our proposed method, only normal representation is learnt and drowsy clips are excluded from training data. This training strategy differs from classification. Hence, methods addressing anomaly detection are considered. For driver drowsiness detection, our method is compared with [10] due to the availability of their implementation. Their model is trained on NTHU-DDD with following changes: 32 frames with downsampling rate of 4 with front views are used to obtain clip length closed to ours and learning rate is decreased by 10-fold after 120 epochs. To make a fair comparison, our result for normal and anomaly rates of 1/2 is used. Table 3 and Table 4 provide results for experiments on NTHU-DDD and DAD dataset respectively. For the experiments on NTHU-DDD, results of [10] before and after post-processing, which corresponds to applying running average filter on the drowsiness scores, are provided and it is shown that our method is superior than [10] on NTHU-DDD. Lastly, our method is trained on DAD dataset, yielding an AUC score of 0.8434, which is comparable with the results reported in [10].

7   Conclusion

In this paper, we propose an anomaly detection-based method to detect drowsiness from camera; which consists of ResNet-34 as backbone and LSTM autoencoder that reconstructs the L2 normalized representations. The confidence levels of normal and anomalous clips are investigated to observe the effect on the knowledge of the network; which yields a training strategy depending on the application. The experiments conducted on NTHU-DDD provide an AUC value of

0.8740. As a future direction, in order to reveal the potential advantages of the proposed method, it can be applied to larger-scale data as well as multi-modal data such as depth image, data obtained from steering wheel and physiological sensors.

References

1. Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., Montanari, R.: Driver workload and eye blink duration. Transportation Research Part F: Traffic Psychology and Behaviour **14**(3), 199–208 (2011)
2. Brookhuis, K.A., De Waard, D.: Monitoring drivers' mental workload in driving simulators using physiological measures. Accident Analysis & Prevention **42**(3), 898–903 (2010)
3. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785–794 (2016)
4. Dong, B.T., Lin, H.Y.: An on-board monitoring system for driving fatigue and distraction detection. In: 2021 22nd IEEE International Conference on Industrial Technology (ICIT). vol. 1, pp. 850–855. IEEE (2021)
5. Garcia, I., Bronte, S., Bergasa, L.M., Almazán, J., Yebes, J.: Vision-based drowsiness detector for real driving conditions. In: 2012 IEEE Intelligent Vehicles Symposium. pp. 618–623. IEEE (2012)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
7. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997)
8. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7132–7141 (2018)
9. Huynh, X.P., Park, S.M., Kim, Y.G.: Detection of driver drowsiness using 3D deep neural network and semi-supervised gradient boosting machine. In: Asian Conference on Computer Vision. pp. 134–145. Springer (2016)
10. Kopuklu, O., Zheng, J., Xu, H., Rigoll, G.: Driver anomaly detection: A dataset and contrastive learning approach. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 91–100 (2021)
11. Luo, W., Liu, W., Gao, S.: Remembering history with convolutional LSTM for anomaly detection. In: 2017 IEEE International Conference on Multimedia and Expo (ICME). pp. 439–444. IEEE (2017)
12. Park, H., Noh, J., Ham, B.: Learning memory-guided normality for anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14372–14381 (2020)
13. Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J.B., Zuiderveld, K.: Adaptive histogram equalization and its variations. Computer Vision, Graphics, and Image Processing **39**(3), 355–368 (1987)
14. Poh, M.Z., McDuff, D.J., Picard, R.W.: Advancements in noncontact, multiparameter physiological measurements using a webcam. IEEE Transactions on Biomedical Engineering **58**(1), 7–11 (2010)

15. Reddy, B., Kim, Y.H., Yun, S., Seo, C., Jang, J.: Real-time driver drowsiness detection for embedded system using model compression of deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 121–128 (2017)
16. Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E., Kloft, M.: Deep one-class classification. In: International Conference on Machine Learning. pp. 4393–4402. PMLR (2018)
17. Shen, Q., Zhao, S., Zhang, R., Zhang, B.: Robust two-stream multi-features network for driver drowsiness detection. In: Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence. pp. 271–277 (2020)
18. Shih, T.H., Hsu, C.T.: MSTN: Multistage spatial-temporal network for driver drowsiness detection. In: Asian Conference on Computer Vision. pp. 146–153. Springer (2016)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
20. Srivastava, N., Mansimov, E., Salakhudinov, R.: Unsupervised learning of video representations using LSTMs. In: International Conference on Machine Learning. pp. 843–852. PMLR (2015)
21. Treat, J.R., Tumbas, N., McDonald, S.T., Shinar, D., Hume, R.D., Mayer, R., Stansifer, R., Castellan, N.: Tri-level study of the causes of traffic accidents: final report. executive summary. Tech. rep., Indiana University, Bloomington, Institute for Research in Public Safety (1979)
22. Vural, E., Cetin, M., Ercil, A., Littlewort, G., Bartlett, M., Movellan, J.: Drowsy driver detection through facial movement analysis. In: International Workshop on Human-Computer Interaction. pp. 6–18. Springer (2007)
23. Weng, C.H., Lai, Y.H., Lai, S.H.: Driver drowsiness detection via a hierarchical temporal deep belief network. In: Asian Conference on Computer Vision. pp. 117–133. Springer (2016)
24. Yu, J., Park, S., Lee, S., Jeon, M.: Representation learning, scene understanding, and feature fusion for drowsiness detection. In: Asian Conference on Computer Vision. pp. 165–177. Springer (2016)
25. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters **23**(10), 1499–1503 (2016)