Applied Computer Systems
ISSN 2255-8691 (online)
ISSN 2255-8683 (print)
June 2022, vol. 27, no. 1, pp. 83–88
https://doi.org/10.2478/acss-2022-0009
https://content.sciendo.com

Detection of Driver Dynamics with VGG16 Model

Alper Aytekin[1*], Vasfiye Mençik[2],
[1]*Yıldız Technical University, İstanbul, Turkey*
[2]*Dicle University, Diyarbakır, Turkey*

***Abstract* – One of the most important factors triggering the occurrence of traffic accidents is that drivers continue to drive in a tired and drowsy state. It is a great opportunity to regularly control the dynamics of the driver with transfer learning methods while driving, and to warn the driver in case of possible drowsiness and to focus their attention in order to prevent traffic accidents due to drowsiness. A classification study was carried out with the aim of detecting the drowsiness of the driver by the position of the eyelids and the presence of yawning movement using the Convolutional Neural Network (CNN) architecture. The dataset used in the study includes the face shapes of drivers of different genders and different ages while driving. Accuracy and F1-score parameters were used for experimental studies. The results achieved are 91 % accuracy for the VGG16 model and an F1-score of over 90 % for each class.**

***Keywords* – Deep learning, drowsiness, transfer learning, VGG16.**

I. Introduction

Drowsiness and fatigue are one of the main factors that threaten road safety, as they cause vehicle drivers to be distracted quickly. Considering that the transition from drowsiness to sleep occurs unconsciously and suddenly, it can be understood how important it is to detect and prevent this situation. In a study conducted by the National Sleep Foundation of America, it was determined that almost 20 % of vehicle drivers feel drowsy while driving [1]. Other studies have revealed that 20 % of traffic accidents occurring worldwide and approximately 50 % of traffic accidents occurring on certain road types are associated with driver fatigue [2]. Considering all these facts, it is of great importance to develop sleep and drowsiness detection systems in vehicles for the establishment of traffic and community safety. With the development of technology in recent years, the interest in the use of different technological driving safety systems in vehicles has increased. Driver drowsiness and fatigue detection systems are one of the vehicle security systems that enable the drivers to focus when they are tired. There are various methods that can be used to measure the fatigue of vehicle drivers. Frequency of yawning, constriction of pupils, increase in the number of blinks, percentage of closure of the eyelids are among the important symptoms for the detection of drowsiness [3]. Drowsiness detection systems basically aim to analyse the images taken continuously or at regular intervals, depending on the system design, using various methods, by means of a camera placed on the windshield of the vehicles, and to regain the driver's attention by activating the warning systems in case the driver is determined to be tired as a result. Different methods can be tried to detect the sleep and drowsiness of vehicle drivers. The idea of using deep learning-based systems to detect sleep and drowsiness of vehicle drivers, which has attracted great interest in recent years, is promising for the future.

In the literature, there are various classification studies aiming to detect lethargy with transfer learning techniques. Dua et al. [4] obtained 85 % accuracy using ResNet and AlexNet networks. Dwivedi et al. [5] obtained 78 % accuracy rate by using SoftMax layer in order to perform classification in their studies aiming to detect lethargy using CNN architecture. Park et al. [6] achieved 73 % accuracy in their study where they proposed a deep learning-based network using Alex Net, VGG-FaceNet and FlowImageNet networks to find driver sleep from the given input videos.

In this study, the driver images obtained using the in-vehicle camera system were analysed using a CNN-based system and the VGG16 classification algorithm, and a classification process was carried out to determine whether the drivers' eyelids are open or closed and whether the drivers are yawning. Accuracy and F1-score parameters were used for experimental studies. Section 2 includes the materials and methods used; Section 3 presents results and discussion, and the last section draws up conclusions.

II. Material and Methods

A. Dataset

In this study, the YAWDD dataset [7] published by the Association for Computing Machinery was used for the classification process of the driver drowsiness. This data set, which consists of photographs of 322 drivers with various facial features in a real car recorded by an in-car camera, consists of four different classes: yawn, no-yawn, open eye and closed-eye.

Fig. 1 shows the image examples of the yawn, no-yawn, open eye and closed-eye classes in the data set. Of the 2900 samples in the data set, 726 are closed, 726 are open, 725 are no yawn and 723 are yawn classes. The data set was used as 80 % training and 20 % test set. The data used in the training phase were not used in the testing phase. The distribution of the

* Corresponding author's e-mail: alper.aytekin@std.yildiz.edu.tr

samples used in the training and testing phase is shown in Fig. 2.

Fig. 1. Image samples in the dataset.

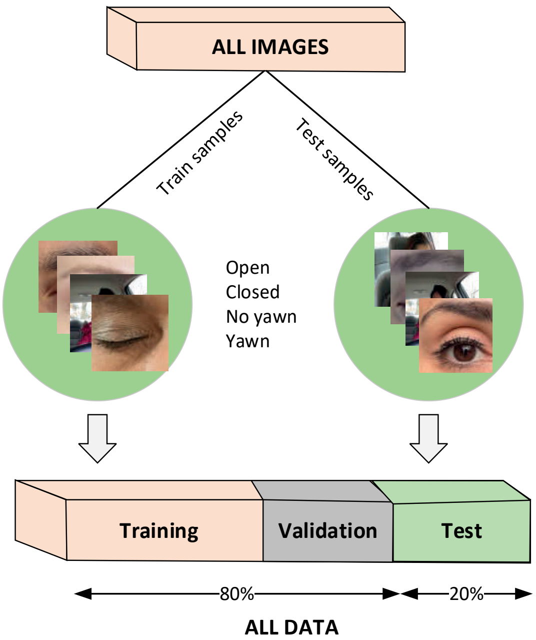

Fig. 2. Distribution of image samples used in training and testing phase.

While the validation set was used to have information about the evolution of the training process, the test set was used to show how the generated neural network has the ability to generalize in the correlation between the actual and predicted values. The training set is used to learn the weights in each layer.

B. Proposed Method

Accurately detecting the decrease in the alertness of the driver while driving is of great importance to prevent any accident that may occur. In this context, it is urgent to develop a method to warn the driver. In this study, driver images obtained using the in-vehicle camera system are analysed based on transfer learning. With the created system, it is determined whether the drivers' eyelids are open or closed and whether the drivers have yawning status. The generated transfer learning-based model provides a non-contact technique to assess various driver alertness levels and facilitates early detection of decline in alertness while driving. Fig. 3 shows the proposed method for detecting driver dynamics while driving based on transfer learning.

Small datasets can degrade CNN performance due to overfitting. Data augmentation is necessary for higher accuracy performance and to reduce the problem of overfitting [8]. In the proposed method, in order to detect the current dynamics of the drivers, data augmentation was performed by applying random rotation to the training images in order to greatly reduce the overfitting problem in the created neural network.

This greatly reduces the overfitting problem. Resizing was applied to the images to transform them into form suitable for the pre-trained models. Even if pre-trained CNN architecture is used in this study, fine-tuning strategy has been applied to create an effective deep learning model. In this context, the last five layers of the pre-trained deep network model are based on a fine-tuning strategy tailored to the target data. This process was performed using the framework of the VGG16 model with million parameters. By using such a large roof and fine-tuning, high quality code samples were created. The proposed approach was used to determine whether the drivers' eyelids were open or closed and whether the drivers were yawning.

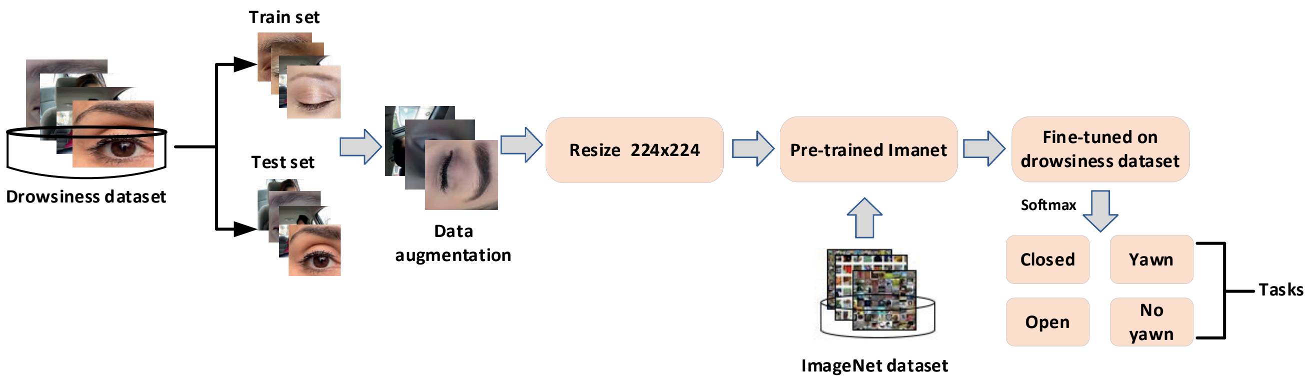

Fig. 3. Flow chart of the proposed method for transfer learning-based determination of driver dynamics while driving.

C. Convolutional Neural Network

Convolutional Neural Network (CNN), a model for processing data with grid-like structures such as images, has been proposed by Lecun et al. [9]. It can be stated that the most useful aspect of CNN, which is designed to automatically and adaptively learn spatial hierarchies of features from low-level models to high-level models, is that it leads to a reduction in the number of parameters. Another feature of CNN that makes it important is that abstract features can be obtained when the input is spread to deeper layers. For example, in image classification, edge features can be detected on the first layers, simpler shapes on the second layers, and higher-level features on the later layers. Fig. 4 shows the basic CNN architecture.

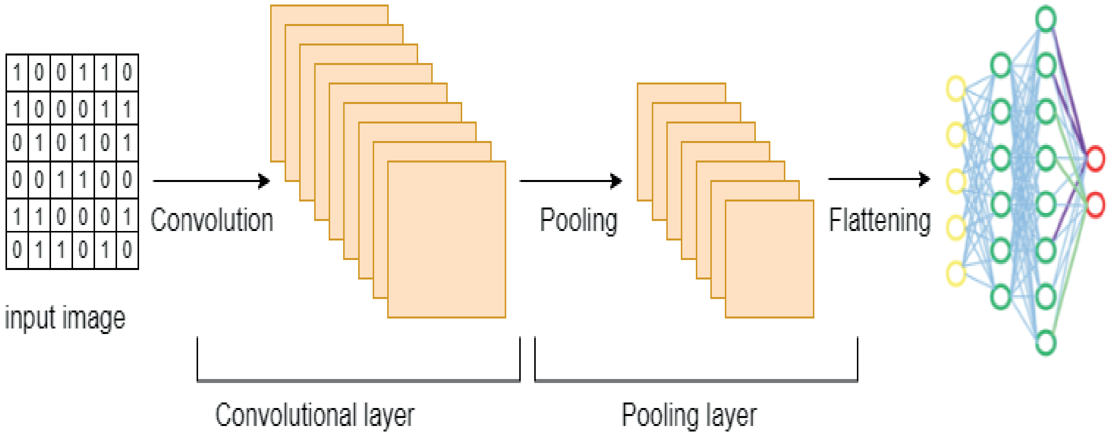

Fig. 4. General convolutional neural network architecture.

Here, the convolution layer and the pool layer extract deep features from the input data and pass them to the fully connected layer to ensure that the system produces accurate and fast results. The convolution process is performed as follows:

$$A_m^n = P_{x,y} * K_m^n \quad (1)$$

Here $P_{x,y}$ denotes input image. x, y represents the spatial location. K_m^n presents the m^{th} convolutional kernel of n^{th} layer. The pooling process, which is responsible for reducing the volume of the image, is carried out as follows:

$$\beta_n = X_c\left(A_{x,y}^n\right) \quad (2)$$

Here β_n denotes the n^{th} output feature map, while $A_{x,y}^n$ denotes n^{th} input feature map. X_c represents the type of pooling operation.

The fully connected layer transmits these properties to the output layer. The fully connected layer transmits these features to the output layer. Network training in CNN consists of two stages, forward propagation and backward propagation, and at the first stage, the input image is represented by the weight and bias at each layer and the forecast output is used to calculate the lost cost value. In the second step, the gradient value of each parameter is calculated backwards with the chain rule, and the gradient value of all parameters is updated and used for the next advanced step. This process ends after both stages are repeated enough and the network training is completed.

D. VGG16

VGG16 [10] is a simple network model developed for better results in the 2014 ILSVRC competition. The most important difference that distinguishes this model from previous models is that the dual or triple convolution layers are followed by the commoning layers. In previous models, the pooling and convolution layers are successive. There are a total of 21 main layers in the VGG16 model, which consists of pooling, convolution and fully connected layers.

Fig. 5 shows the general VGG16 architecture. In this architecture, the image input resolution is 224×224 pixels and the filter size in the convolutional layer is 3×3 pixels. The last layers in the architecture consist of fully connected layers used for feature extraction.

E. Performance Parameters

Accuracy, precision, recall and F1-score parameters were used to evaluate model performance. The formulas for these evaluation metrics are as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \quad (3)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (4)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (5)$$

$$\text{F1} - \text{score} = 2 \times \frac{\text{pecision} \times \text{recall}}{\text{precision} + \text{recall}}. \quad (6)$$

True Positive (TP) refers to correctly classified negative cells, while FN (False Negative) refers to incorrectly classified negative cells. Similarly, TN (True Negative) indicates correctly classified positive cells, while FP (False Positive) indicates incorrectly classified positive samples. Confusion matrix containing different combination of actual values and predicted values was used in this study.

The higher the accuracy value, the higher the classification performance. Precision and Recall are used to evaluate how accurate the classification is. A low Recall value indicates that the classifier suffers from a large number of FNs, and a low Precision value indicates that the classifier suffers from a large number of FPs.

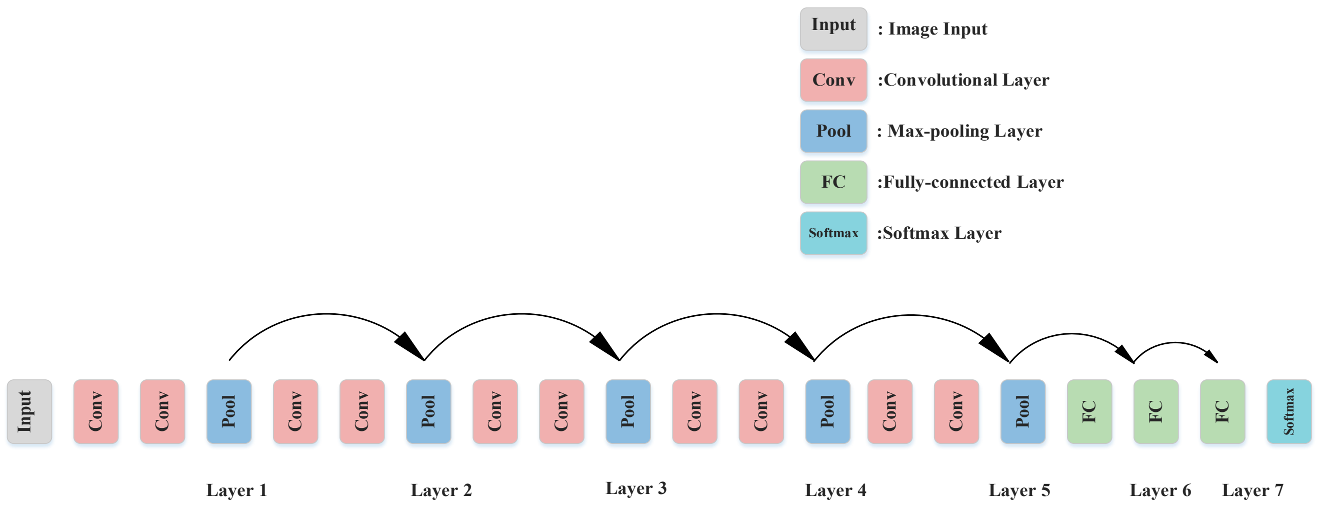

Fig. 5. VGG16 model architecture.

III. Experimental Results and Discussion

In this study, a neural network is trained to classify the data provided by the dataset to determine the drowsiness and sleepiness of vehicle drivers. Estimates were made with the test set consisting of 64 images. No overfitting problems were observed during the training phase. The confusion matrix obtained as a result of the training with the neural network created using the transfer learning strategy is shown in Fig. 6.

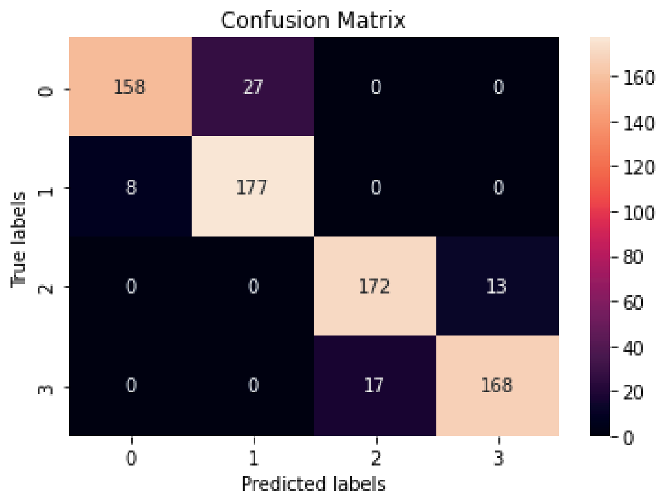

Fig. 6. Confusion matrix obtained in the test set of the created neural network.

When the confusion matrix obtained from the accuracy, precision, recall and F1-score parameters is examined, the first line represents the open eye, the second line represents the closed eye, next line represents no yawn and the last line represents the yawn class. When the Confusion matrix, which includes different combinations of real values and predicted values, is examined, it is seen that the model created has the ability to recognize each class. The data provided by the dataset were processed with the VGG16 classification algorithm in order to determine the drowsiness and sleep status of the vehicle drivers. The obtained classification accuracy is shown in Table I.

TABLE I

The Performance Measurement Values of VGG16 Model

CLASS	ACCURACY	PRECISION	RECALL	F1-SCORE
OPEN-EYE		0.95	0.85	0.90
CLOSED-EYE	0.91	0.87	0.96	0.91
NO YAWN		0.91	0.93	0.92
YAWN		0.93	0.91	0.92

When Table I is examined, it is seen that the average accuracy rate of 91 % was obtained in the classification process made in the test set, which was created using the VGG16 model and included four different classes.

Precision and Recall values should be 1 for good classification. In this study, precision results for open-eye, yawn, no yawn classes were at least 91 % and above. The F1-score value, which is the weighted average of Precision and Recall values, is at least 90 % and above. As seen in Table I, high F1-score values obtained indicate that the classifier has low FP and low FN values. All these results clearly show that transfer learning-based vehicle systems are effective for driving safety optimization. Deep learning-based methods can combine feature extraction and classification to create an end-to-end classification model, avoiding the complex and time-consuming design process of artificial features and classifiers. Table II shows a comparison of drowsiness detection methods in the literature in the context of Accuracy.

TABLE II
ACCURACY COMPARISON OF DROWSINESS DETECTION METHODS

AUTHORS	METHOD	RESULT
GWAK ET AL. [11]	ENSEMBLE MACHINE LEARNING	65.2 %
JABBAR ET AL. [12]	CNN	83.33 %
MEHTA ET AL. [13]	EYE ASPECT RATIO AND EYE CLOSURE RATIO	84 %
KEPESIOVA ET AL. [14]	CONV GRNN, CNN,	84.41 %
DUA ET AL. [4]	ALEXNET, VGG-FACENET, FLOWIMAGENET, RESNET	85 %
PROPOSED METHOD	TRANSFER LEARNING VGG16	91%

Gwak et al. in their study used the ensemble machine learning method to classify light sleepy states based on the perception of behavioural and physiological indicators of drivers [11].

Jabbar et al. used CNN-based neural network to detect microsleep and drowsiness. They used facial markings detected by the camera for drowsiness classification. The authors stated that the method with which they obtained an accuracy of 83.33 % could be used for the detection of sluggishness in embedded systems [12]. Mehta et al. used an android application for the detection of driver drowsiness and made real-time detection with image processing techniques. In their study, the authors calculated two different rates to detect drowsiness based on adaptive thresholding: Eye Aspect Ratio and Eye Closure Ratio. In the system created, videos were recorded and the driver's face was detected in each frame with image processing techniques. The authors achieved 84 % accuracy with the random forest classifier [13]. In their study, Kepesiova et al. determined whether the driver was awake or tired thanks to the neural network applied to the image captured with the help of the camera [14].

Dua et al. performed driver drowsiness detection with architecture consisting of four different deep learning models. The authors used RGB videos of the drivers as input in their work and considered four different features. Each of the deep learning models was used to detect different features, and the models divided the features into four classes such as sleeplessness, blinking and drowsiness, yawning and nodding. The authors achieved 85 % accuracy in their work [4].

In our proposed work, we applied the pre-trained VGG16 architecture using the concept of transfer learning to reduce the training time of our model and increase the performance of the classifier. Thus, it helps prevent automobile accidents caused by falling asleep at the wheel by monitoring the fatigue status of the drivers. The obtained results show that the proposed method is applicable.

IV. CONCLUSION

It is an important task for public health and road safety to detect the driver's sleepiness and drowsiness while driving before any accident occurs. In this study, a classification system based on transfer learning has been proposed using deep learning architectures. In the classification process, the data obtained from the data set have been classified in four classes: open-eye, closed-eye, no-yawn and yawn. With the VGG16 model used in experimental studies, it has been observed that 91% accuracy and an F1-score value of over 90 % have been obtained for each class. All these results clearly show that transfer learning-based tool systems are effective for security optimization.

In the future, we will investigate the interaction of heart rate measurements, low mental workload during driving simulation, and fatigue, as well as perform numerical analysis of tractor accidents using a driving simulator.

DECLARATION OF ETHICAL STANDARDS

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

CONFLICT OF 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] O. Ursulescu, B. Ilie, and G. Simion, "Driver drowsiness detection based on eye analysis", in *2018 International Symposium on Electronics and Telecommunications (ISETC)*, Timisoara, Romania, 2018, pp. 1–4. https://doi.org/10.1109/ISETC.2018.8583852

[2] R. O. Mbouna, S. G. Kong, and M. G. Chun, "Visual analysis of eye state and head pose for driver alertness monitoring", *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1462–1469, Sep. 2013. https://doi.org/10.1109/TITS.2013.2262098

[3] S. Chinara,"Automatic classification methods for detecting drowsiness using wavelet packet transform extracted time-domain features from single-channel EEG signal", *Journal of Neuroscience Methods*, vol. 347, 2021, Art no. 108927. https://doi.org/10.1016/j.jneumeth.2020.108927

[4] M. Dua, R. Singla, S. Raj, and A. Jangra, "Deep CNN models-based ensemble approach to driver drowsiness detection", *Neural Computing and Applications*, vol. 33, no. 8, pp. 3155–3168, Jul. 2021. https://doi.org/10.1007/s00521-020-05209-7

[5] K. Dwivedi, K. Biswaranjan, and A. Sethi, "Drowsy driver detection using representation learning", in *Advance Computing Conference (IACC)*, Gurgaon, India, Mar. 2014, pp. 995–999. https://doi.org/10.1109/IAdCC.2014.6779459

[6] S. Park, F. Pan, S. Kang, and C. D. Yoo, "Driver drowsiness detection system based on feature representation learning using various deep networks", in *Asian Conference on Computer Vision*, Taipei, Taiwan, Nov. 2016, pp. 154–164. https://doi.org/10.1007/978-3-319-54526-4_12

[7] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, "YawDD: A yawning detection dataset", in *Proceedings of the 5th ACM Multimedia Systems Conference*, Singapore, Mar. 2014, pp. 24–28. https://doi.org/10.1145/2557642.2563678

[8] D. Cireşan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification", arXiv: 1202.2745, Tech Rep. No. IDSIA-04-12, Feb. 2012. [Online]. Available: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://arxiv.org/pdf/1202.2745.pdf

[9] B. Boser, J. D. Y. Le Cun, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network", *Advances in Neural Information Processing*, vol. 2, 1989.

[10] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *arXiv*, preprint arXiv:1409.1556, 2014.

[11] J. Gwak, A. Hirao, and M. Shino, "An investigation of early detection of driver drowsiness using ensemble machine learning based on hybrid sensing", *Appl. Sci.*, vol. 10, no. 8, Apr. 2020, Art no. 2890. https://doi.org/10.3390/app10082890

[12] S. Mehta, S. Dadhich, S. Gumber, and A. J. Bhatt, "Real-time driver drowsiness detection system using eye aspect ratio and eye closure ratio", in *Proceedings of international conference on sustainable computing in science, technology and management (SUSCOM)*, Jaipur, India, Feb. 2019. https://doi.org/10.2139/ssrn.3356401

[13] Z. Kepesiova, J. Ciganek, and S. Kozak, "Driver drowsiness detection using convolutional neural networks", in *2020 Cybernetics & Informatics (K&I)*, Velke Karlovice, Czech Republic, Mar. 2020, pp. 1–6. https://doi.org/10.1109/KI48306.2020.9039851

[14] R. Jabbar, M. Shinoy, M. Kharbeche, K. Al-Khalifa, M. Krichen, and K. Barkaoui, "Driver drowsiness detection model using convolutional neural networks techniques for android application", in *Proceedings of the 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies*, Doha, Qatar, May 2020, pp. 2–5. https://doi.org/10.1109/ICIoT48696.2020.9089484

Alper Aytekin received the B. S. degree in electrical and electronics engineering from Erciyes University, Kayseri, Turkey, in 2019. He pursues Master degree in electronics and communication engineering at Yıldız Technical University, Istanbul, Turkey, in 2021. His current research interests include artificial intelligence, deep learning, communication, antenna design.
E-mail: alper.aytekin@std.yildiz.edu.tr
ORCID iD: https://orcid.org/0000-0002-1865-5983

Vasfiye Mençik received the B. S. degree in electrical and electronics engineering from Dicle University, Diyarbakır, Turkey, in 2016, the M. S. degree in electrical and electronics engineering from Dicle University, Diyarbakır, Turkey, in 2022. She was enrolled in PhD program in electrical and electronics engineering at Dicle University, Diyarbakır, Turkey, in 2022. Her current research interests include artificial intelligence, deep learning, machine learning, image processing, Graph Neural Network, and histopathological images.
E-mail: vasfiyemencik@gmail.com
ORCID iD: https://orcid.org/0000-0002-3769-0071