Cascade width learning based fatigue detection method

Dong Wang*[a], Pianhui Wu[b]
[a]Big Data and Computing Institute, Guangdong Baiyun University, Guangzhou 510450, Guangdong, China; [b]School of Mechatornic Engineering, Guangdong Polytechnic Normal University, Guangzhou 510450, Guangdong, China

ABSTRACT

The fatigue detection method based on neural network has been widely applied in the field of land transportation. The accuracy varies by different methods and solutions. Disadvantages of current widely applied methods for fatigue detection have been exposed by rational review, e.g., cost-effective, potable implementation issues. Cascade width learning based detection method with outstanding features is a sound solution to overcome shortcomings of other traditional detection methods. Therefore, it is of great significance and value to research this method and present its academic value and applicable values. This paper introduces a cascade width learning based fatigue detection method. The paper looks forward to development of cascade width learning based fatigue detection method, both academic and applicable values of proposed method have been presented by rational evaluation based on rational discussion on experimental results, the results prove that proposed method has achieved 94.9% accuracy in 52.43ms, which suggests more accuracy and fast speed than other common methods including: combinations of LBP and SVM, ASM and Fuzzy, CNN and PERCLO.

Keywords: Neural network, fatigue detection, CNN, cascade width learning

1. INTRODUCTION

Road accidents persist as a significant public health concern, prompting ongoing efforts to enhance vehicle capabilities in recognizing and analyzing road conditions to prevent mishaps and safeguard passengers. Consequently, understanding driver behavior, particularly regarding fatigue—a leading cause of accidents and fatalities—has emerged as a prominent research focus. This paper proposes a method for analyzing and predicting driver fatigue through a review of pertinent research studies.

Fatigue encompasses visual, mental, and physical aspects, gradually impairing reaction speed, concentration, and alertness during prolonged or intense activity. Notably, fatigue exacerbates visual and mental strain and induces physical exhaustion. Muscle fatigue, occurring during sustained contraction or exertion, is detectable through changes in electromyographic signals and other physiological markers, reflecting its severity[1].

Recent advancements in neural network-based fatigue detection have yielded algorithms comparable in accuracy to physiological signal-based methods. Leveraging neural networks' proficiency in facial feature extraction and key point localization, these methods offer enhanced speed, accuracy, and robustness. As sample sizes grow and hardware capabilities improve, neural network-based fatigue detection is poised to become even more efficient and precise. This paper will introduce such a method and its industry applications.

2. FATIGUE DETECTION METHOD BASED ON NEURAL NETWORK

2.1 Visual fatigue detection

Visual fatigue stems from various factors such as eye physiology, environmental conditions, and mental state. Common eye issues like astigmatism, myopia, and hyperopia, coupled with insufficient ambient light and contrast, contribute to visual fatigue.

Detection of human eye state is crucial for assessing fatigue levels, especially for vehicle drivers[2]. Face detection forms the basis for this assessment, yet its accuracy is affected by factors like facial expressions, movements, lighting conditions, and occlusions[3]. Algorithms like the Active Shape Model (AAM) and Cascaded Pose Regression (CPR) have been

*dongwang@baiyunu.edu.cn

International Conference on Optics, Electronics, and Communication Engineering (OECE 2024), edited by Yang Yue, Proc. of SPIE Vol. 13395, 1339520 · © 2024 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.3048270

developed to enhance face detection accuracy[4].

The perclos criteria, introduced in the 1970s, utilize eye angle, blinking frequency, and pupil occlusion rate to gauge driver fatigue[5]. This method, endorsed by experts from the U.S. Federal Highway Administration, provides a non-contact, real-time fatigue assessment and is widely adopted in fatigue detection[5].

Based on the degree of eye closure over time, perclos criteria define three standards—EM, p70, and p80—indicating varying levels of eye closure corresponding to fatigue. Despite differing judgment standards, the underlying principle remains consistent: the duration of eye closure per unit of time determines the fatigue level[5].

As shown in Figure 1, we take p80 criterion as an example, during T4-T1, the pupil of the eye is considered closed when the closure is more than 80%, in the meanwhile, perclos ration in the period of time of T3-T2 is calculated as:

$$perclos = \frac{\mathrm{T3}-\mathrm{T2}}{\mathrm{T4}-\mathrm{T1}} \tag{1}$$

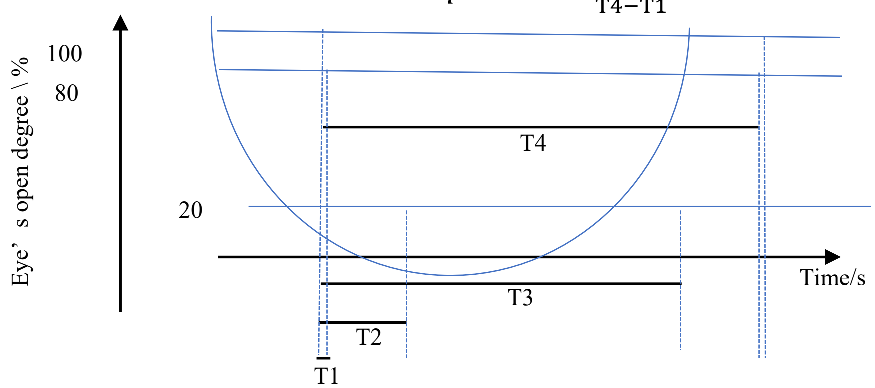

Figure 1. Degree of open and closure of human eye pupils[4].

To compute the perclos value using machine vision, it's essential to convert the time ratio to a frame ratio, representing the number of eye closure frames relative to the total frames per unit of time[5]. This calculation is done using equation (2):

$$perclos = \frac{\mathrm{Nc}}{\mathrm{N}} \tag{2}$$

In which, Nc denotes the number of closed-eye frames and N denotes the total number of frames per unit of time. A number of previous researches have revealed that p80 criterion is the best way to accurately reflect the level of driver fatigue[5].

2.2 Brain fatigue detection

Brain fatigue manifests as declines in attention, memory, and alertness, posing challenges for detection[6]. Studies indicate that intense brain activity results in fatigue-related changes in brain waves, including α, β, and theta waves[6]. Russell et al. developed a convolutional neural network optimization method for monitoring pilot fatigue, while Li et al. constructed a driver fatigue recognition model using deep sparse self-coding networks, showing superior classification performance[7,8]. Luo et al. proposed a fatigue classification algorithm based on Gamma deep belief networks, achieving notable recognition accuracy and stability[9].

The accuracy of EEG signal analysis depends on electrode placement methods, categorized as entry, semi-entry, and non-entry types[10]. Entry and semi-entry methods are more effective but require specialized expertise and support. Consequently, fatigue detection primarily utilizes EEG signals from the cerebral cortex, although this approach is hindered by cost and inconvenience[10].

2.3 Feature fusion detection

Features fusion detection of fatigue is capable of improving accuracy of fatigue detection. For example, by positioning face feature points, fatigue features can be recognized using Eye Aspect Ratio (EAR) and Lip Aspect Ratio (Mouth Aspect Ratio, MAR)[10], and the recognition points are shown in Figures 2 and 3, and the EAR and MAR calculation equations are respectively:

$$EAR = \frac{(P2-P6)+(P3-P5)}{2(P1-P4)} \tag{3}$$

$$MAR = \frac{(P2-P8)+(P3-P7)+(P4-P6)}{3(P1-P5)} \tag{4}$$

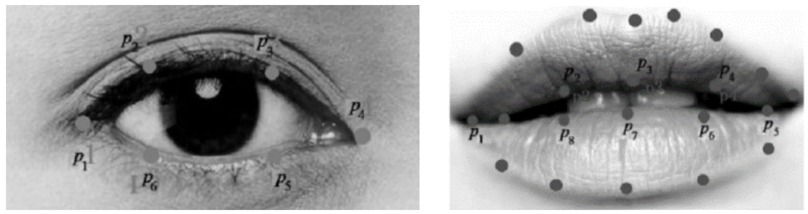

Figure 2. Recognition makers on the eye and mouth[10].

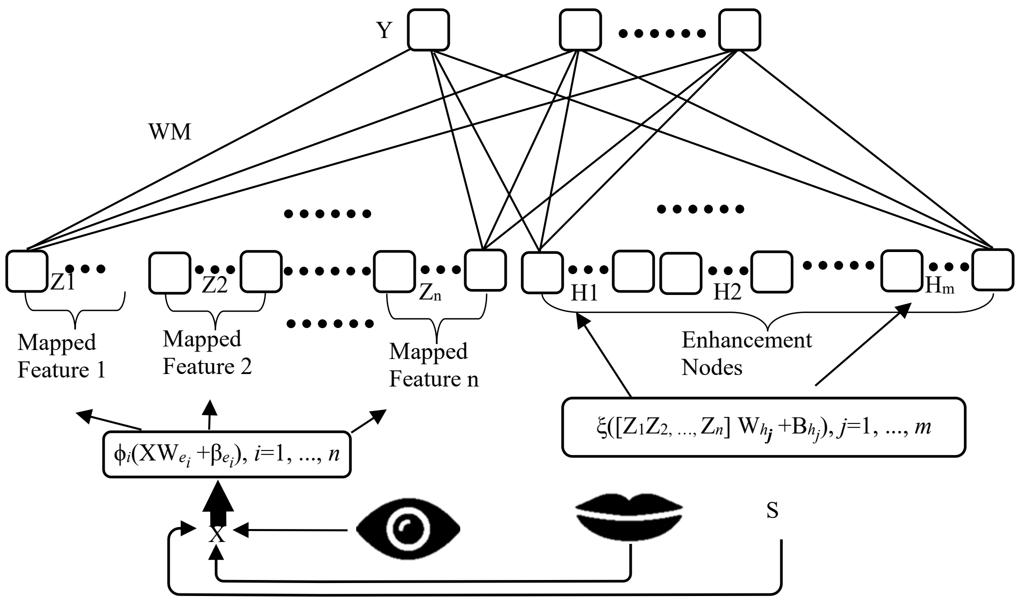

Figure 3. The structure of width learning system[16].

2.4 Weaknesses of current fatigue detection method

Visual and brain fatigue detection methods based on neural networks have shown promise, but they also exhibit shortcomings. Visual fatigue detection, while fast and convenient, is susceptible to accuracy issues when head movements or light blockages occur[11]. Brain fatigue detection via EEG signals requires filtering and frequency domain analysis, making signal extraction complex and costly[10]. Multi-feature fusion detection enhances accuracy but slows detection speed due to increased task complexity[12].

The rapid development of neural network models is facilitating the creation of portable and accurate detection systems[13]. In visual fatigue detection, neural networks are utilized for face recognition and eye localization, such as the MTCNN network improving accuracy through additional network layers[14].

Current fatigue detection methods for drivers rely on large, accurate sensors, but lack real-time, non-invasive functionality[3]. Future trends should focus on real-time detection, multi-feature fusion, and non-invasiveness. Lightweight neural networks can enhance real-time detection speed, while multi-task convolutional neural networks can fuse features for improved accuracy[3]. Non-invasive approaches, such as machine vision or wearable micro-detection devices, are crucial for unobtrusive fatigue assessment[3]. Thus, Cascade Width Learning-based methods could offer a viable solution for fatigue detection.

3. CASCADE WIDTH LEARNING BASED DETECTION METHOD

The cascade width learning-based detection method emerges as a prominent technological trend for the future[15]. Unlike traditional machine learning algorithms, deep networks significantly enhance classification and regression performance, yet their complex structures with numerous hyperparameters often lead to prolonged convergence times. To address this, Chen et al. propose a width learning system aimed at reducing model training time and enhancing task efficiency[16]. This system, depicted in Figure 3 involves generating mapping features from original data, enhancing them with randomized weights, and then utilizing both sets of features as inputs to a single-layer perceptron.

$$Z_i = \phi_i(XW_{e_i} + \beta_{e_i}),\ i = 1, 2, ..., n \tag{5}$$

$$Hj = \xi j(ZiWhj + \beta hj),\ j = 1, 2, \ldots, m \tag{6}$$

Equation (5) demonstrates the method for generating the i_{th} set of mapping features Z, utilizing original input data X, random weight matrix Wei with Gaussian distribution, bias constant β_{e_i}, and sparsification and regularization function ϕ_i. Equation (6) illustrates the generation of the j_{th} set of augmented features H_j, where Z_i represents overall mapping features, W_{h_j} is the random matrix post-orthogonal regularization, β_{h_j} is the bias constant, and ξ_j is the nonlinear function. Equation (7) outlines the expression of BLS, where Hj represents all augmented features, W^m denotes connection weights, and Y signifies the network output. Adjusting the values of i and j allows adaptation to tasks of varying complexity.

$$Y = [Z_1, ..., Z_i|\ \xi(Z^i W_{h_1} + \beta_{h_1})), \ldots, \xi(Z^i W_{h_j} + \beta_{h_j})]\ W^m = [Z_1, ..., Z_n \mid H_1, ..., H_j]\ W^m = [Z^i \mid H^j]W^m \tag{7}$$

BLS applies Gaussian filtering and sparse representation to original data during mapping node generation to reduce noise and linear correlation[15]. For augmented node generation, BLS maps data to a high-dimensional subspace using orthogonal matrices to enhance linear differentiability, then employs nonlinear functions to increase network model nonlinearity[16].

The image input to the network is represented by equation (8), where $I(n, n)$ denotes the pixel value at position (n, n) and X represents the network input. Output labels Y classify the state of each eye and mouth image, with open eyes or mouth labeled as 0 and closed eyes or mouth labeled as 1.

$$X = [I(1, 1), \ldots, I(1, n), I(2, 1), \ldots, I(n, n)] \tag{8}$$

Let $X_{zh} = [Z_1, Z_2, ..., Zn, H_1, H_2, ..., H_m]$, the network connection weights are solved as shown in equation (9).

$$W^m = (X_{ZH}{}^T X_{ZH} + \alpha I)^{-1} X_{ZH}{}^T Y \tag{9}$$

Compared to backpropagation in multi-layer neural networks, BLS employs the pseudo-inverse method, significantly reducing network training time[15]. BLS also supports three incremental learning methods—augmentation node increment, feature node increment, and input data increment—facilitating parameter updates through the calculation of pseudo-inverses for newly added nodes, further reducing training time[16].

As shown in Figure 4, Driver head attitude encompasses pitch, yaw, and roll. In a fatigued state, the head typically rotates downward. Small rotations around the y-axis may indicate mirror checking, while excessive rotation suggests a violation of path viewing.

Figure 4. Coordinate system for head positioning[15].

The face pose model is established using the nonlinear least squares method, as shown in equation (10). The basic head pose angle, represented by (α, β, γ), is calculated using POSIT. Here, n denotes the number of face feature points used to construct the model, q_i represents the facial feature points in the image, p_i denotes the feature points of the 3D generalized standard model, R is the rotation matrix, t is the spatial offset, and C is the scaling factor.

$$f(\alpha, \beta, \gamma) = \min \{ {}_{i=1}^{n} ||q_i - C \cdot [R(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) \cdot p_i + t]||\ ^2\} \tag{10}$$

When the driver lowers his head more than 30 degrees, which suggests pitch<-30°, or turns his head left or right more than 30 degrees which suggests |Yaw| > 30°, the head state S_h is set to 1, otherwise it is set to 0, as shown in equation (11).

$$S_h = f(x) = \begin{cases} 1, if (a < -30^\circ \; or \; |\beta| > 30^\circ) \\ 0, otherwise \end{cases} \tag{11}$$

Fatigued drivers exhibit increased eye closure time, yawning, and a larger head pitch angle. There exists a contextual correlation between changes in eye state, mouth state, and head pitch over time. Equation (12) constructs a facial state time series to characterize these changes, where S_{e_n}, S_{m_n}, S_{h_n} represent the state of the eyes, mouth, and head at frame *n*, respectively.

$$S = [S_{e_1}, S_{e_2}, \dots, S_{e_n}, \dots, S_{m_1}, S_{m_2}, \dots, S_{m_n}, \dots, S_{h_1}, S_{h_2}, \dots, S_{h_n}, \dots] \tag{12}$$

Facial state time series S is labeled as 1 for fatigue and 0 for normal states. A BLS fatigue detection network is trained using S and its corresponding labels. The fatigue detection process is illustrated in Figure 5.

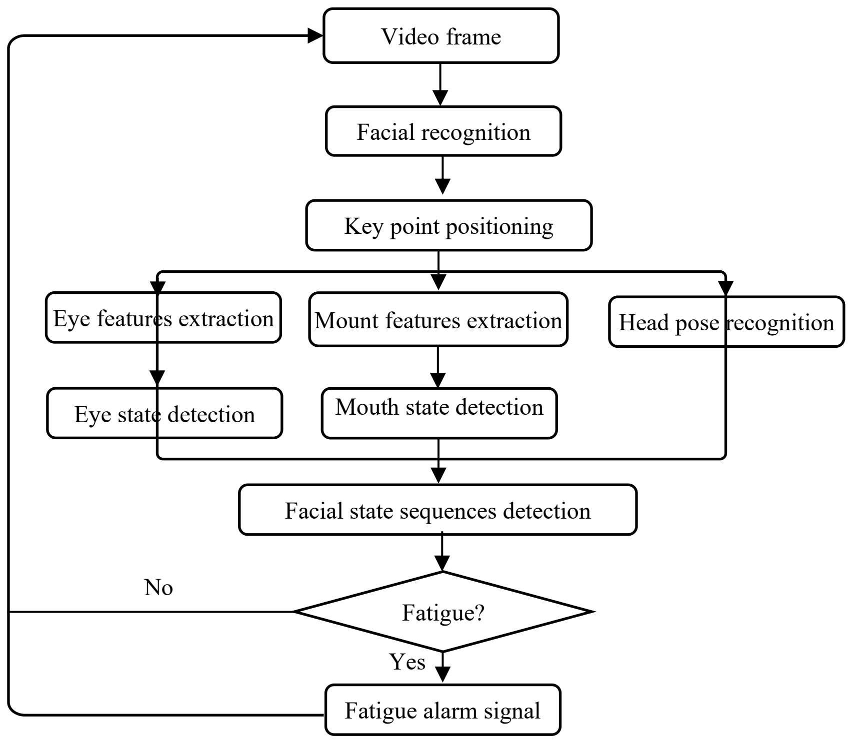

Figure 5. The workflow of cascade width learning based fatigue detection system[16].

4. RESULTS AND DISCUSSION

The experimental platform features an Intel(R) Core(TM) i7-10700k processor with a 3.8GHz main frequency and 16GB of RAM. The Kaggle face key points dataset comprises 7000 96×96 grayscale face images labeled with 15 key points. After expanding the dataset through rotation, 9000 images are obtained, with 6750 for training and 2250 for testing. The test set's mean square error is 1.95, indicating an average offset error value of 1.4 for each key point.

For fatigue detection, a facial fatigue video dataset (MT-F) is utilized, consisting of videos collected from 9 volunteers under various conditions. From these videos, 2000 open and closed eye images and open and closed mouth images are extracted for training and testing. A set of 60 consecutive frames is selected from each video type at a frame rate of 20fps to create facial temporal state sequences S for training and testing the fatigue detection network.

The width learning network adjusts its structure by altering the number of nodes. The restructuring process of the eye state recognition network is depicted in Table 1, wherein the number of enhancement nodes is gradually increased to enhance recognition accuracy. The mouth state recognition network comprises 100 mapping nodes and 2500 enhancement nodes, while the fatigue detection network features 50 mapping nodes and 600 augmented nodes. The average training time for the BLS eye state recognition network is 1.6 s.

Table 1. BLS eye state recognition network tuning process.

Mapping node	Enhancement node	Accuracy %
100	1000	94.6
100	2000	95.4
100	3000	96.7
100	3500	95.9
200	3000	96.2
400	3000	95.5

The DROWSY database comprises 3 videos per tester at various fatigue levels, excluding sunglasses. Table 2 presents the results of human eye state recognition on both the MT-F and DROWSY databases. Recognition accuracy on MT-F is lower than DROWSY due to complex lighting and sunglasses occlusion. The mouth state recognition accuracy is 98.3%.

Table 2. Eye and mouth condition recognition results.

Facial area	Dataset	Data volume	Accuracy %
Mouth	MT-F	1000	98.3
Eye	MT-F	1500	96.5
	DROWSY	1500	99.2

The algorithm is tested with videos consisting of 60 frames per segment, with results presented in Table 3. To validate the effectiveness of multiparameter fusion, videos are detected using a fatigue detection network trained with eye state sequences and another trained with facial state sequences. Experimental findings indicate that the multi-parameter fusion network achieves higher detection accuracy.

Table 3. Fatigue detection results.

Manner	Dataset	Mode	Data volume	Accuracy %
Single	MT-F	No glasses	400	94.5
		Lens	400	92.8
		Sunglasses	200	90.5
	DROWSY		500	94.8
Multi	MT-F	No glasses	400	96.3
		Lens	400	94.5
		Sunglasses	200	93.5
	DROWSY		500	96.5

To validate the proposed algorithm’s superiority, it is compared with other fatigue detection algorithms in terms of detection accuracy and time, as depicted in Table 4. The proposed algorithm exhibits higher accuracy than CNN+PER CLOS due to improved detection accuracy through multiparameter fusion. While the BLS and CNN-based methods have longer detection times compared to traditional methods, attributed to increased network parameters, the proposed algorithm achieves a 6.18ms reduction in detection time despite using multiple parameters.

Table 4. Algorithm performance comparison.

Algorithm	Accuracy %	Time /ms
LBP + SVM	89.7	44.32
ASM + Fuzzy	91.0	45.84
CNN + PERCLO	93.1	58.61
Proposed method	94.9	52.43

5. CONCLUSION

this paper outlines classical driving fatigue detection methods, discussing their advantages and disadvantages through rational comparison. various applications of neural networks in fatigue detection are analyzed and compared. to validate the proposed method's performance and efficiency, a verification experiment is conducted using an infrared camera to mitigate the impact of light changes and glasses on detection results. A CNN regression network is employed to detect facial key points based on a width learning system. the driver's fatigue state is inferred through the sequence of eye, mouth, and head states using a second-level width learning network. results indicate a detection accuracy of 94.9% and a single-frame detection time of 52.43 ms, demonstrating superior real-time accuracy compared to traditional methods. the experimental outcomes underscore the academic and practical significance of the proposed method.

ACKNOWLEDGMENT

This work has been supported by the Guangdong Province Department of Education on the study of the Intelligent Fatigue Test Based on Eye Aspect Ratio (Type: Characteristic innovation project (Natural Science), ID: 2022KTSCX157)).

REFERENCE

[1] Abbas, Q. and Abdullah, A., "Driver fatigue detection systems using multi-sensors, smartphone, and cloud-based computing platforms: a comparative analysis," Sensors, 21(1), 56 (2020).
[2] Sikander, G. and Shahzad, A., "Driver fatigue detection systems: A review," IEEE Transactions on Intelligent Transportation Systems, 20(6), 2339-2352 (2018).
[3] Bjørheim, F., Siriwardane, S. C. and Pavlou, D., "A review of fatigue damage detection and measurement techniques," International Journal of Fatigue, 154, 106556 (2022).
[4] Sunandha Shri, S. and Ezhilarasan, M., "An efficient classification algorithm for employee well-being prediction using deep learning," IoT Based Control Networks and Intelligent Systems: Proceedings of 3rd ICICNIS 2022. Singapore: Springer Nature Singapore, 467-481 (2022).
[5] Zhuang, Q. Y., et al., "Driver fatigue detection method based on eye states with pupil and iris segmentation," IEEE Access, 8, 173440-173449 (2020).
[6] Deng, P. Y., et al., "Detecting fatigue status of pilots based on deep learning network using EEG signals," IEEE Transactions on Cognitive and Developmental Systems, 13(3), 575-585 (2020).
[7] Russell, B., et al., "Predicting fatigue in long duration mountain events with a single sensor and deep learning model," Sensors, 21(16), 5442 (2021).
[8] Li, J. J. and Qiang, W., "Multi-modal bioelectrical signal fusion analysis based on different acquisition devices and scene settings: Overview, challenges, and novel orientation," Information Fusion, 79, 229-247 (2022).
[9] Luo, Y. X., et al., "Pilots' brain fatigue state inference based on gamma deep belief network," ACTA Electronica Sinica, 48.6, 1062(2020).
[10] Geetavani, B., et al., "Drivers fatigue detection and alarming using ear and MOR values," 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), (2021).
[11] Shalash, W. M., "Driver fatigue detection with single EEG channel using transfer learning," 2019 IEEE International Conference on Imaging Systems and Techniques (IST), (2019).
[12] Deng, W. H., et al., "Fatigue driving detection based on multi feature fusion," 2019 IEEE 4th International

Conference on Image, Vision and Computing (ICIVC), (2019).
[13] Parekh, V., Darshan, S. and Manan, S., “Fatigue detection using artificial intelligence framework,” Augmented Human Research, 5, 1-17 (2020).
[14] Ku, H. C. and Wei, D., “Face recognition based on MTCNN and convolutional neural network,” Frontiers in Signal Processing, 4(1), 37-42 (2020).
[15] Shi, X., Shan, S., Kan, M., et al., “Real-time rotation-invariant face detection with progressive calibration networks,” Computer Vision and Pattern Recognition, Salt Lake City, (2018).
[16] Zhang, Z., Ning, H. and Zhou, F., “A systematic survey of driving fatigue monitoring,” IEEE Transactions on Intelligent Transportation Systems, (2022).