Array 19 (2023) 100301

Contents lists available at ScienceDirect

Array

journal homepage: www.sciencedirect.com/journal/array

Optimizing student engagement in edge-based online learning with advanced analytics

Rasheed Abdulkader [a], Firas Tayseer Mohammad Ayasrah [b], Venkata Ramana Gupta Nallagattla [c], Kamal Kant Hiran [d], Pankaj Dadheech [e], Vivekanandam Balasubramaniam [f], Sudhakar Sengan [g,*]

[a] Electrical Engineering Department, Imam Mohammad Ibn Saud Islamic University (IMSIU), Saudi Arabia
[b] College of Education, Humanities and Science, Al Ain University Al Ain, United Arab Emirates
[c] Department of Computer Science and Engineering, Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, 520007, India
[d] Department of Computer Science and Engineering, Lincoln University College, Malaysia
[e] Department of Computer Science & Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan (SKIT), Jaipur, Rajasthan, 302017, India
[f] Faculty of Computer Science and Multimedia, Lincoln University College, Wisma Lincoln, No. 12-18, Jalan SS 6/12, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia
[g] Department of Computer Science and Engineering, PSN College of Engineering and Technology, Tirunelveli, 627152, Tamil Nadu, India

ARTICLE INFO

Keywords:
Deep learning
VGG-19
Online learning
Edge computing
Face detection
DLIP

ABSTRACT

Edge-Based Online Learning (EBOL), a technique that combines the practical, hands-on approach of EBOL with the convenience of Online Learning (OL), is growing in popularity. But accurately monitoring student engagement to enhance teaching methodologies and learning outcomes is one of the difficulties of OL. To determine this challenge, this paper has put forth an Edge-Based Student Attentiveness Analysis System (EBSAAS) method, which uses a Face Detection (FD) algorithm and a Deep Learning (DL) model known as DLIP to extract eye and mouth landmark features. Images of the eye and mouth are used to extract landmarks using DLIP or Deep Learning Image Processing. Landmark Localization pre-trained models for Facial Landmark Localization (FLL) are one well-liked DL model for facial landmark recognition. The Visual Geometry Group-19 (VGG-19) learning model then uses these features to classify the student's level of attentiveness as fatigued or focused. Compared to a server-based model, the proposed model is developed to execute on an Edge Device (ED), enabling a swift and more effective analysis. The EBOL achieves 95.29% accuracy and attains 2.11% higher than existing model 1 and 4.41% higher than existing model 2. The study's findings have shown how successful the proposed method is at assisting teachers in changing their teaching methodologies to engage students better and enhance learning outcomes.

1. Introduction

Numerous facets of daily life are transforming due to the Internet of Things (IoT). The ubiquitous nature of IoT technologies sets them apart from earlier advancements and encourages the development of intelligent and autonomous solutions [1]. A significant strategic technology trend is the development of the IoT [2]. The conceptual model in the novel approach to learning is predicated on having multiple sensors that connect the physical and virtual machines. Sensors are embedded into any object, which supports this change in basic assumptions. Next, embedded sensors use Machine-to-Machine (M2M) communication to connect billions of devices to the Internet [3]. Face Detection (FD) using Deep Learning is a popular Computer Vision (CV) task that involves identifying and locating human faces in images or video streams. DL techniques, specifically Convolutional Neural Networks (CNNs), have proven highly effective for FD tasks.

On the whole, the physical world is going online quickly. The IoT is causing worldwide interest and apprehension as it develops swiftly and

* Corresponding author. .
E-mail addresses: rmabdulkader@imamu.edu.sa (R. Abdulkader), firas.ayasrah@aau.ac.ae (F. Tayseer Mohammad Ayasrah), nallagattla@gmail.com (V.R.G. Nallagattla), pdf.drkkhiran@lincoln.edu.my (K. Kant Hiran), pankajdadheech777@gmail.com (P. Dadheech), vivekanandam@lincoln.edu.my (V. Balasubramaniam), sudhasengan@gmail.com (S. Sengan).

https://doi.org/10.1016/j.array.2023.100301
Received 21 April 2023; Received in revised form 19 June 2023; Accepted 24 June 2023
Available online 28 June 2023
2590-0056/

has become an ever-growing topic [4]. Several signs show that the IoT's potential can transform many industries, including Higher Education (HE) institutions, particularly universities [5].

As a result of the internet's rootedness in our schools, e-Learning is becoming a widely used practice in many school education systems. But the IoT has many educational uses, and this disruption has enormous ramifications. Schools may now increase the safety of their campuses, monitor essential resources, and improve information access in the learning platform due to the development of smartphone technology and the IoT [6]. Instead of the Traditional Lesson Plans (TLP) of the past, this technology for developing "Smart Lesson Plans (SLP)" by teachers [7]. From paper books, college students significantly progressively switch to handheld devices. Now, the learning students are done at their own pace and have a comparable learning experience in their homes and classrooms because they have access to all the required material at their fingertips.

Additionally, when this tendency increases students' accessibility, it helps professors teach more effectively. The proliferation of linked technology has eliminated the need for tutors to grade tests on paper manually or conduct other daily activities. Instead, tutors concentrate on customized learning to help their students the best. Professors can collect data on their students using cloud-connected devices and then decide the one that requires the most outstanding individualized care and attention. These data support tutors in enhancing student engagement and modifying their lecture plans for upcoming classes.

On the one hand, due to the limitations of various hardware facilities, today's online instructional System (OTS) cannot apply and manage the teaching resources. On the other hand, while Cloud Computing (CC) is becoming increasingly widespread, it typically uses centralized management, making it impossible to guarantee online classes' low delay and reliability and even posing some data security challenges [8]—Edge Computing (EC) technology to enhance the existing OTS. EC technology integrates a network, a computer system, storage spaces, and application capabilities on free-standing platforms. It starts on edge to provide fast network service responses and fulfill industry standards for real-time business, application intelligence, security, and privacy. Between the physical subject and industrial connection, or even above the physical subject, is where the edge computer is situated. Data processing and analysis with EC is faster than with CC [9].

Online Learning (OL) is being used more in schools for readings, articles, video lectures, virtual learning environments, and timed tasks [10], although there are still problems in the online setting in traditional classrooms. Some learning resources given to the students include timed tasks with long reading materials attached or 1-h video lectures; these harm some students' attention spans [11]. OL and their attention span impact the motivation of students. Because online classes are performed over the Internet, students cannot communicate in person with their classmates, and procrastination is inevitable due to the deadlines' flexibility [12]. Higher motivation levels may cause the student to persevere and seek out more challenging tasks even when challenges are performed, but lower motivation levels may cause the student to disengage from an activity [13]. Their research aims to determine how OL affects motivation and attention span. Earlier researchers suggested that several factors can affect a student's ability to pay attention in an online course. Eye tracking, reduced class feedback, and facial direction, when seen through a camera, can do it [14].

Researchers have been drawn to the study of OL attention, which has produced fruitful results. In real life, it might be challenging to gauge how attentive kids are in class. Accurately detecting each student's status of learning is a challenging task. Researchers have conducted studies on recognizing Learning Attention (LA) [15,16], producing potential outcomes [17]. examined the level of LA at home and abroad, split the emphasis of the research into two classes—Attention Recognition (AR) based on behavior and AR based on facial expression—and then discussed the AR's development phase. OL platforms and offline classroom settings are the main focus of research scenarios. Researchers have been working hard to understand AR in online classes for the convenience of voice and image acquisition. Using 636 students from management and economics disciplines at national universities as samples, a technique of developing a Triadic Theory of Learning based structural equation model that analyzed the effectiveness of OL was proposed by Ref. [18]. Deep Learning (DL) and Shallow Learning (SL) are the two types of teaching quality. This method contends that learning vigor, teaching methods, and teaching contents significantly impacted students' LA, which led to further optimization of the OTS design to increase the effectiveness of OL [19]. proposed a technique with three sub-modules. For head pose detection, the first module primarily identifies the angle of deviation of each student's head. By closing the mouth and eyes, the second module scores fatigue. The third module scores emotion by finding facial expressions. After combining the abovementioned data, a fuzzy comprehensive evaluation technique objectively measures their LA.

To check students' fatigue levels in real-time and with accuracy [20], developed a Fatigue Monitoring System (FMS) that assessed a student's fatigue state based on changes in the contour of their eyes. In order to realize the checking of quality of the class [21], developed a method to employ Facial Recognition (FR) technology for calculating the classroom learners' head-up rate by checking the rotation Angle of the head [22]. proposed a method to recognize learners' 3D head rotation angle, degree of eye closure, and facial expression to assess their facial concentration degree [23]. proposed an FR-based fuzzy comprehensive evaluation technique. The four detection parts—left (right) head-turning angle, head-lifting (low) head-turning angle, eye closure, mouth closure, and facial expression—were found by analyzing face images. The learning focus was processed using the quantitative scores of the four detection parts.

DL technology-based image feature classification has appeared as a renowned research direction while computer software and hardware development, Artificial Intelligence (AI) technologies, and digital image processing have continuously developed. As known earlier, DL research has set off in recent years, thus covering a wide range of topics, including text, pronunciation, and visual elements. DL has been successfully used in recognition and image classification and may efficiently circumvent the issues associated with Feature Extraction's (FE) artificial selection. Though the CNN-based student fatigue state detection has superior accuracy, it is challenging for real-time application on the terminal side. These are the two other key issues with the current research: (a) the eye's positioning is subject to the peripheral situation, and (b) despite the incredible accuracy of the CNN-based student fatigue state detection, it is challenging for real-time implementation of the terminal side.

Using a robust Face Detection (FD) algorithm and a sophisticated FE model, DLIP, which extracts eye and mouth landmark features, provides an edge-based student attentiveness analysis system in this study. The FE is fed into the Visual Geometry Group-19 (VGG-19) learning model to determine whether the student is focusing or fatigued on paying attention. The model's performance, as applied in Edge Devices (ED), is linked to server-based models. The findings that prove the viability of the proposed model are summarized.

2. Research contribution

(a) **Integration of FD and AR:** The paper proposes the integration of Face Detection (FD) and Attentiveness Recognition (AR) using the MTCNN algorithm for FD and DLIB for extracting Eye and Mouth Aspect Ratios (EAR and MAR). This combination enables the system to identify and analyze the attentiveness of students.

(b) **Deployment of VGG-19 Network:** This work proposes using the VGG-19 learning network for FE and analysis. By leveraging this DL model, the EBSAAS can effectively analyze the facial FE to determine student attentiveness.

(c) **Implementation of Face Recognition (FR) System: The** paper describes implementing an FR system within the EBSAAS framework. The FR system utilizes the ED (a smartphone) and the Edge Server (ES) for processing and recognition tasks, providing the ability to recognize individuals and assign names to their faces.

(d) **Training and Processing Pipeline:** The paper presents a training and processing pipeline where the EBSAAS model is initially trained on a face dataset within the ED. Once trained, the system can capture images of individuals or groups, perform FS, recognize faces, and analyze attentiveness by sending images to the ES for processing. This pipeline enables real-time analysis of student attentiveness.

(e) **Integration of Edge Server:** The paper highlights the inclusion of an ES in the EBSAAS architecture. The ES is responsible for processing the attentiveness analysis and training the recognition classifier. By offloading these tasks to the ES, the EBSAAS can leverage its computational capabilities and potentially enhance processing speed.

The article is systematized as follows: The overview and the research works are discoursed in Section 1, Section 2 has the proposed EBSAAS model, Section 3 has the investigational performance analysis with comparative illustration, and Section 4 is the paper's conclusion with recommendations for future work.

3. Proposed edge-based student attentiveness analysis system (EBSAAS)

FD and AR are the two significant aspects of EBSAAS. Using the Multi-task Cascaded Convolutional Networks (MTCNN) algorithm, a face's potential location is found inside an image during the FD phase. In order to find students who are LA, the attentiveness recognition phase uses DLIB for extracting the student's Eye and Mouth Aspect Ratios. The FE is then put into the VGG-19 learning network, which analyzes the student attentiveness. The EBSAAS is connected to the ED (a smartphone) and the ES to find out the FR system's processing time.

The learning model is first trained in the ED using a face dataset. The EBSAAS then uses its camera to snap an image of a person or group and starts the FD process. Once the Face is recognized, it is marked on the photo, and the person's name is assigned. In order to recognize, the photo is sent to the ES by the ED using the Internet rather than internally processing the attentiveness analysis. In the ES, the training of the recognition classifier takes place, which is very much like the ED. Following the FD and attentiveness recognition procedures, it passes the tagged and labelled photo back to the ED after obtaining the image from the same. The depiction of the ES-FR system through a block diagram is shown in Fig. 5. In Fig. 1, the framework outlined in this paper is depicted. The specifics of each part's implementation will be covered in detail.

3.1. Face Detection

In order to perform FD and face alignment rapidly and effectively for this work, this work employs MTCNN (Fig. 2), a DL-based FD and face alignment algorithm [24]. The left and right corners of the mouth, the nose, and the left and right eyes are the five facial features that MTCNN can recognize. The five significant points, however, are not enough to extract information about facial fatigue; hence this work employs MTCNN for FD. Proposal Network (*P*-Net), Refine Network (R-Net), and Output Network (ONet) are the three subnets that face MTCNN. They are arranged in cascades to form MTCNN [25].

In FD, the goal is to identify and locate the presence of faces within an image or video frame. The primary focus is detecting the overall face region rather than specific facial features. However, during the FD process, certain facial features can be used as cues to locate and validate the presence of a face. Some common facial features that can be applied are eye, nose, mouth and face contour. The facial features are retrieved using DLIB.

P-Net. This network's primary task is to decide the candidate window's regression box and BB. Highly overlapping windows are removed after calibrating the candidate window using non-maximum suppression. *P-Net* is a face region-specific proposal network. The network employs border regression and a locator of essential points in a face to generate a face area's primary proposal after using a face classifier to identify if a face is present. This section will produce many candidate windows, which R-Net will then use as input.

R-Net: This network's primary job is to weed out bogus samples while continuing to gather Bounding Boxes (BB) and Regression Vectors (RV). *R-Net* has a connecting layer that is more conclusive than the preceding network. This network's primary job is to weed out false samples while continuing to gather BB and RV. *R-Net* has a connecting layer that is further complete than the preceding network. Many candidate windows are obtained once the test sample passes on the *P-Net* layer. The network filters out a high number of candidate windows. The prediction outcomes are further improved on the selected candidate boxes using Non-Maximum Suppression (NMS) and BB regression.

O-Net. The third network is more complex than the first two. There are 256 Fully Connected (FC) layers in *O-Net*. Also, this network layer will calculate the facial feature points' position after further filtering the *R-Net* candidate window. Additionally, this procedure can eliminate various impediments like hats, sunglasses, and ordinary glasses.

3.2. Detection method based on eye and mouth feature

3.2.1. Face localization

The system localizes the FD after successfully detecting it. Face localization includes Facial Landmarks Detection (FLD). The purpose of FLL is to precisely identify and locate specific key points or landmarks on a human face. These landmarks represent important facial features such as the eyes, eyebrows, nose, mouth, and jawline. By accurately localizing these landmarks, performing several tasks related to facial analysis, FR, tracking, and FD becomes possible. FLD is the image shape predictors' subset. The following is a list of the facial landmarks found in

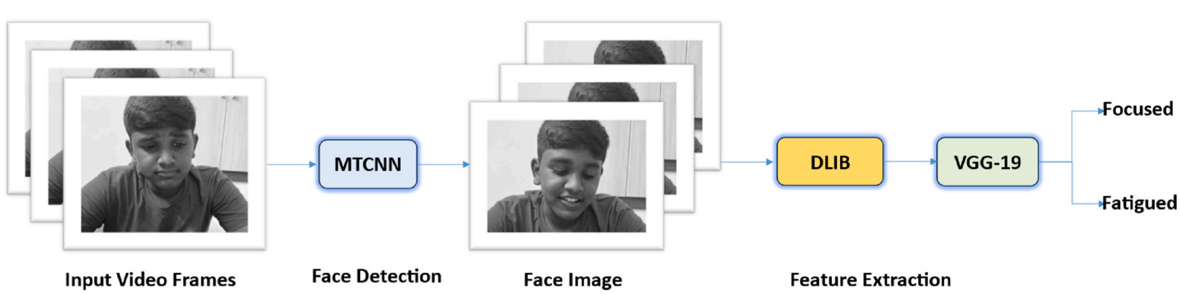


Fig. 1. Framework of fatigue detection.

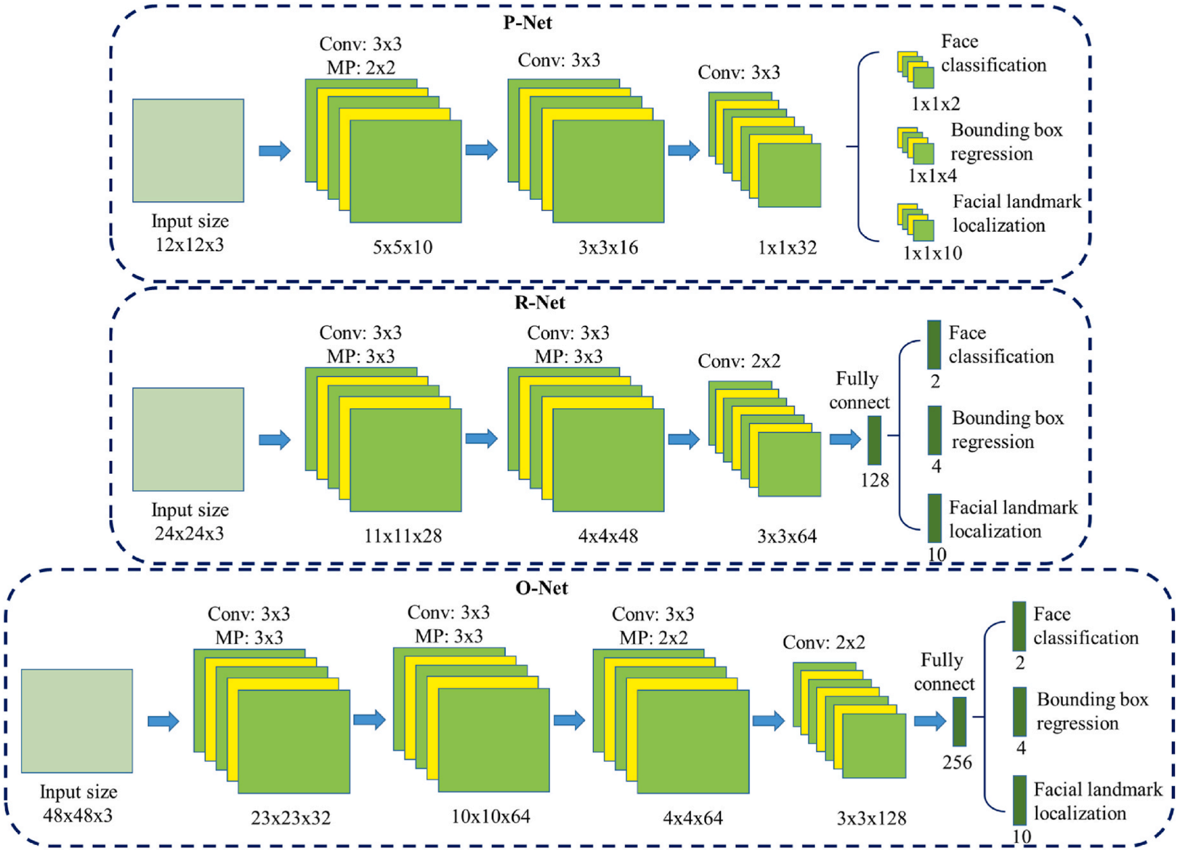


Fig. 2. MTCNN architecture.

the face region.

Eyes
Eyebrows
Lips
Nose
Jawline

Using the shape predictor method, one may localize these landmarks on the Face by quickly extracting the eye region. These landmarks are found around the eyes. The DLIB library used *C++* programming to help the Machine Learning (ML) algorithm. Using the function named *"dlib. shape predictor"* ("shape predictor 68 landmarks. dat"), 68 unique feature points are returned by the algorithm in the given frames for each person [26]. This article uses the *"dlib.get_frontal_face_detector"* function to obtain the frames. A semi-automatic method for the landmark points was introduced by Ref. [27], and the outcomes of 300 different faces in "The Wild Challenge (300-W)" was reported in that article. It was the prime FLD test where earlier techniques were compared. The 68 individual landmark points that were detected are listed with precise numerals in Fig. 3. This study determined the lip distance and EAR using these landmarks.

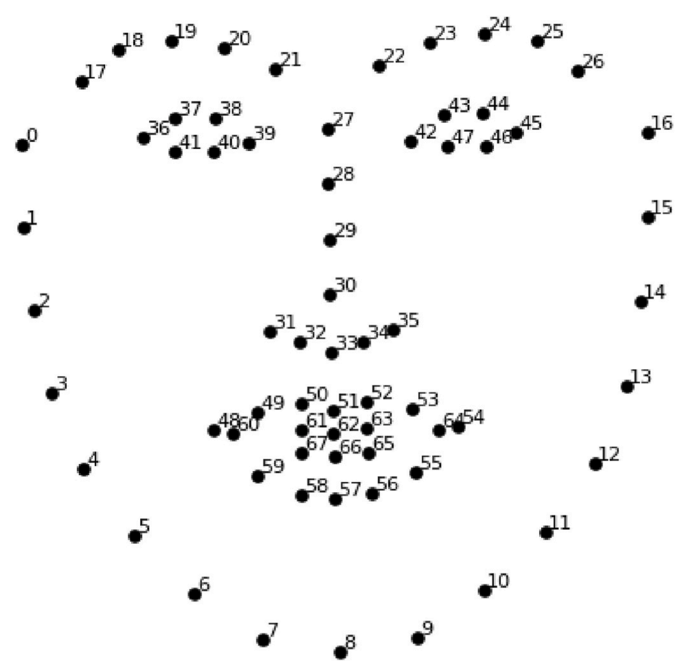


Fig. 3. The 68 FLD coordinates from the iBUG-300 W dataset.

3.2.2. Eye aspect ratio (EAR)

The EAR, a crucial part of this algorithm, can be used to assess whether or not someone is closing their eyes in the provided video frame.

Fig. 4 shows the eye that a set of six facial points with labels and specific coordinates. The distance between points p_1 and p_4 (the width of an eye) is along a horizontal line, whereas the distance along a vertical line is between the middle of p_2 and p_3 and p_6 and p_5 (the height of an eye). While the vertical line's length will vary when the eye opens and closes, the horizontal line's length will always remain constant. This work detects blinking by assessing the ratio between the lengths of these two lines. While the eye is open, this ratio will remain constant, but it will swiftly go to '0' when the eye blinks.

The aspect ratio will increase and remain constant over time, as seen

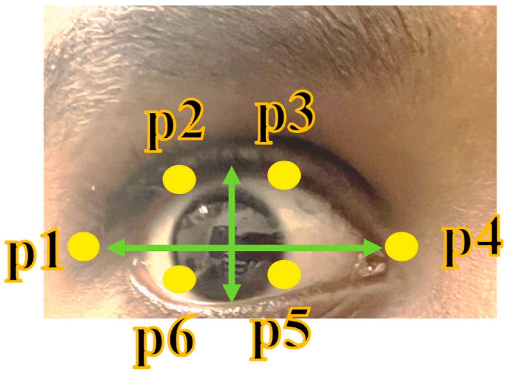


Fig. 4. Eye key points.

in Fig. 5(a). Fig. 5 (b), on the other hand, shows that the EAR will be equal to '0', showing that the person is blinking at that moment. Equ (1) below can be used to calculate the EAR.

For the active, fatigued, and sleep stages, the EAR values range from 0.38 to 0.30, 0.255 to 0.18, and 0.155 to 0.03. The experiment created a non-overlapping graph that helps calculate the classification boundary for the three states together. The threshold range (THrange) was measured using this graph. The range is,

$$TH_{\text{range}} = \begin{cases} EAR \geq 0.28\ ;\ \text{Active} \\ 0.17 < EAR \leq 0.27\ ;\ \text{Fatigue} \\ EAR \leq 0.17\ ;\ \text{Sleep} \end{cases} \tag{2}$$

In this study, the average eye blinks were prevented from interfering with fatigue detection based on EAR by choosing a time rate of 1000 ms, or 1 *Sec*. Drowsiness was detected by the proposed system when the EAR value, as EQU (2), goes beyond the threshold for a continuous time of 1000 ms.

3.2.3. Mouth aspect ratio (MAR)

Another symptom of fatigue, the yawn, is classified by lip distance. "Lip landmark points," or exciting points for calculating lip distance. The distance is calculated by deducting the upper lip weight's meaning from the lower lip weight's meaning. Fig. 6 depicts the distinctive top and lower lip points. Notably, each of the top and lower lips have eight distinct points. The weights of the points are added to determine the meaning of each lip.

The exciting point weights of the top lip are added together in EQU (3).

$$Top_{lipW} = \sum_{i=50}^{53} W_{P_i} + \sum_{i=61}^{64} W_{P_i} \tag{3}$$

To add up every lower lip's point weights, EQU (4) is used.

$$\text{Lower}_{lipW} = \sum_{i=56}^{59} W_{P_i} + \sum_{i=66}^{68} W_{P_i} \tag{4}$$

Then, as in the following, the mean of the top and lower lips is determined: The lip distance is determined as EQU (5) and EQU (6)

$$Top_{lip\_mean} = \frac{\text{Top}_{lipW}}{8} \tag{5}$$

$$Lower_{lip\_mean} = \frac{\text{Lower}_{lipW}}{8} \tag{6}$$

Top_{lip_mean} and $Lower_{lip\_mean}$ are derived in EQU (7).

$$\text{Lip}_{distance} = \left\| Top_{lip\_mean} - Lower_{lip\_mean} \right\| \tag{7}$$

3.3. VGGNet architecture

Authors [28] from the University of Oxford devised the CNN model as VGGNet. This study's primary aim is to examine the impact of CNN depth on accuracy. A 224 × 224 RGB image serves as the input to the VGG-based convNet. The preprocessing layer subtracts RGB images with pixel values between 0 and 255 from the ImageNet training set's mean image values.

Weight and CL transform input images into training imagesVGG16 using 13 CL and 3 FC layers, with smaller (3 × 3) large filters instead of huge ones. With one 7 × 7 Convolutional Layer (CL), it now has the same effective receptive field. VGGNet is the version used in this study, which includes 19 wt, 16 CL, 3 FC, and 5 Pooling Layers (PL) (Fig. 7). The VGGNet uses two FC layers with 4096 channels each, followed by a further FC layer with 1000 channels and a SoftMax layer to classify data.

3.3.1. Details of architecture

- The CL holds the first two layers with 3 × 3 filters, and since 64 filters are used in the first two layers, the volume is 224 × 224 × 64 due to the deployment of the same convolutions. The filters always have a 3 × 3 stride of 1.
- The volume was reduced from 224 × 224 × 64 to 112 × 112 × 64 using a PL with a 2 × 2 size and stride 2.
- Two CLs have 128 filters, resulting in a new dimension of 112 × 112 × 128.
- The PL reduces the volume to 56 × 56 × 128.
- The size of the Down-sampling is 28 × 28 × 256, followed by two CLs with 256 filters each.
- A max-PL separates two stacks with three CLs.
- After the last PL, volume 7 × 7 × 512 is flattened into an FC layer with 4096 channels and 1000 SoftMax classes.

4. Experiment

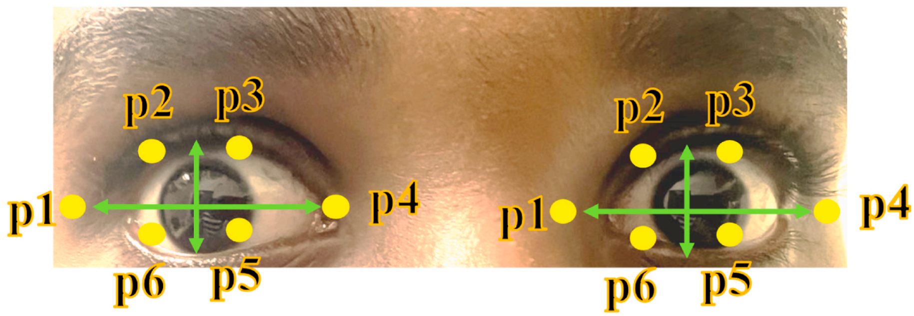
4.1. Real-life drowsiness dataset (RLDD)

RLDD includes 30 h of RGB videos of 60 people in good health. 180 videos—one for alertness, low vigilance, and drowsiness—were given to each tester. Staff and students who volunteered or received extra credit participated. Everyone who took part was older than 18.51 men and 9 women aged 20–59, with a mean age of 25 and standard deviation of 6. From 180 videos, frames were extracted for training and testing. The training uses 40% of the frames, and 60% is achieved using 40% of testing.

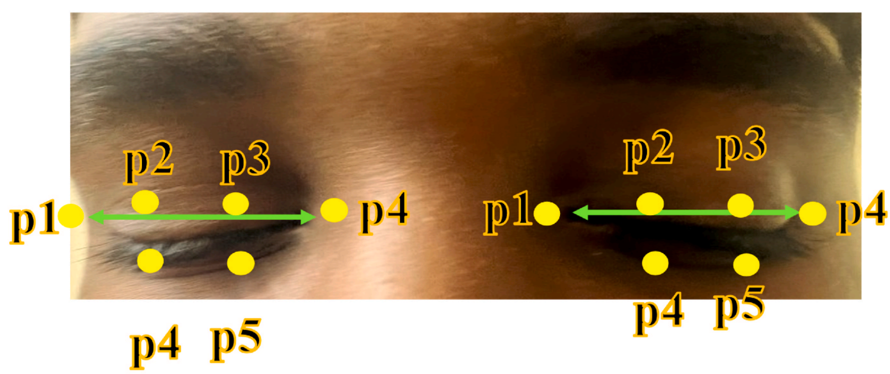
10 Caucasians, 30 Indo-Aryans and Dravidians, 8 Middle Easterners, and 7 East Asians. In 72 of the 180 videos, the subjects had much facial hair, and 21 of the 180 videos had the subjects' wearing glasses. Videos were shot at different angles against unusual circumstances from real life. The students input each video themselves using a webcam (or) smartphone. The video frame rate was consistently below 30 *fps*, typical of client cameras. Fig. 8 indicates RLDD images.

4.2. Experimental results and analysis

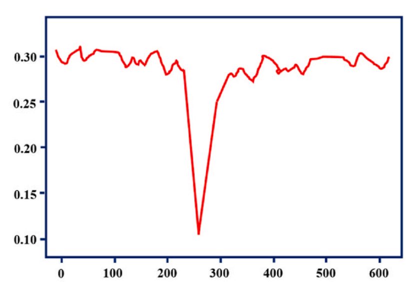
The Operating System (OS) environment used for this experiment is Linux (Ubuntu14.04). The computer's configuration details are Windows 10 OS, Intel (R) Xeon (R) CPU E5-2630 v3, 8 GB of RAM, and GTX1080 graphics card. Application based on smartphone technology for fatigue detection is created using Android Studio, and the ES program is created using Java and Eclipse IDE. Using MTCNN for detecting and cropping the images of faces, this work divided the video dataset into images for the experiment. The DLIB library labels the basic facial features after cropping the face image to determine the state value of the eye and mouth. This paper detects close fatigue by deciding the aspect ratio of the eyes and mouth. Ten videos from the RLDD test set were



(a) EAR for an Opened eye will be Grater and Constant



(b) EAR for a closed eye will be equal (or) Closer to ‘0’ and Constant



(c)

Fig. 5. **(A)** Open eyes; **(b)** Closed eyes; **(c)** with automatically recognized landmarks with EAR drop graph.

$$E\_A\_R = \frac{\| X_2 - X_6 \| + \| X_3 - X_5 \|}{2 \| X_1 - X_4 \|} \tag{1}$$

chosen randomly for this study, and the findings are displayed in Table 1. In the 10 selected videos at random, the detection accuracy of fatigued driving is 95%. The results of sample detection are shown in Fig. 9.

Our algorithm proves that the device can properly detect the driver’s fatigue state and function at high speeds while working in many conditions. The class labels assigned in the research work are fatigue and focused, where the outcome is depicted in Fig. 9. This proposed approach increases the fatigue driving detection algorithm’s accuracy compared to FaceNet + KNN [29] and MTCNN + LSTM [30] algorithms. Additionally, it performs better in real-time, which satisfies the needs of the attentiveness analysis system [31,32]. In Table 1, the comparative result is displayed.

The detection of student reaction has to be quick to predict the nature of listening, and it is necessary to analyze the performance speed for the existing model with the proposed model. A comparatively proposed model identifies the outcome quickly, which is minimal than existing models.

4.3. Edge model analysis

Combining 10 random images into 1 image, this work adjusted the dataset to replicate a classroom environment to evaluate the ES and ED performance. It allowed us to analyze the performance associated with the proposed model. The processing time for FD by the ES and ED is shown in Fig. 10. As the number of faces rises in the ED, there is a drastic increase in the detection time. On the other hand, the ES experiences a modest increase in FD time for a similar instance.

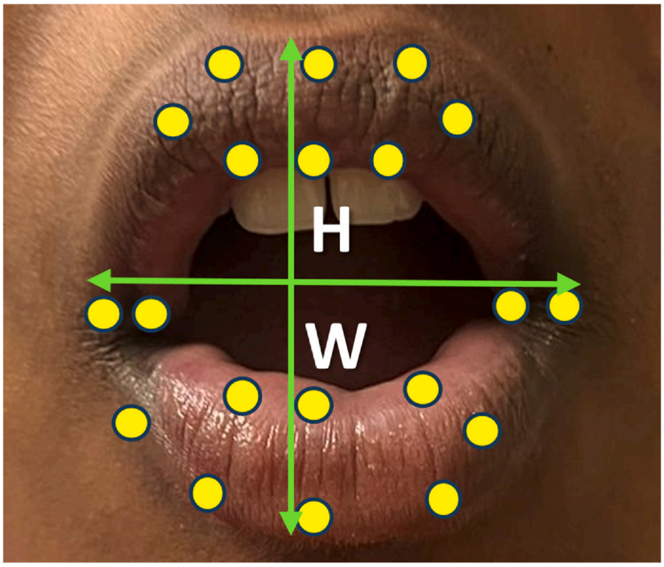


Fig. 6. Top and lower lips landmark points.

The image resolution influences the FD time rather than the image file size. For different image file sizes (image resolution 446 × 614 pixels), the FD time difference is displayed in Fig. 11 for the ES and ED.

The processing time for attentiveness recognition in the ED and ES is displayed in Fig. 12. Similar patterns are traced in the attentiveness recognition time measurement. The more faces in an image, the longer it takes for the ED to recognize them attentively. The ES, however, causes a moderate increase in recognition time. In the image, according to the number of faces, Fig. 12 compares the overall processing time between the ES and ED. Regarding ED, the FD and fatigue recognition times internal to the ED are added to determine the total processing time.

The RLDD dataset is used in the research, and the DLIB library features are considered. The performance measures, like accuracy, speed, and time for ES and ED, are analyzed. The processing time for attentiveness recognition in the ED and ES is illustrated in Figs. 10–11. The ED takes longer to FR carefully when more people are in an image. However, the ES brings a slight increase in recognition time. To forecast the nature of listening, it is vital to identify student reactions quickly, and it is essential to compare the performance speeds of the proposed and current models. In comparison to existing models, the proposed model predicts the result in a short amount of time. The accuracy of classification is compared, and EBOL outperforms the *state-of-the-art* techniques.

Table 1
Accuracy analysis.

Algorithms	Accuracy (%)	Speed ($ms \cdot f^{-1}$)
Model 1	93.18	53.61
Model 2	90.88	61.54
Proposed Model	95.29	47.23

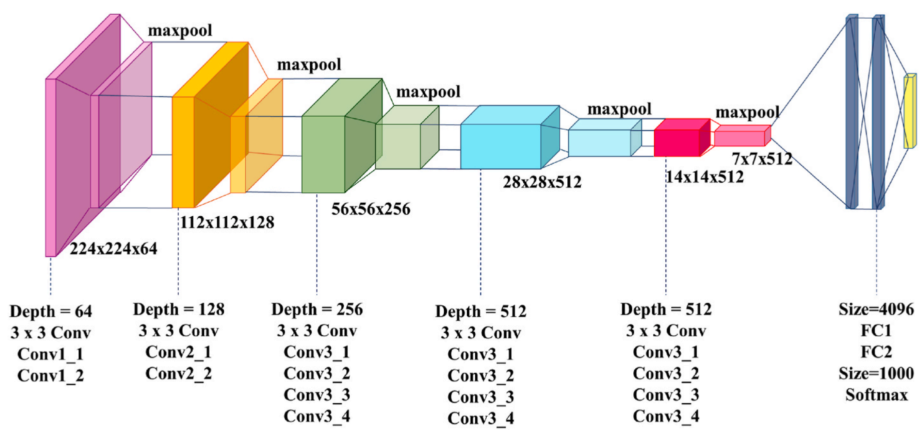


Fig. 7. Diagram of the network model of VGG-19.

Fig. 8. The sample of the RLDD dataset.

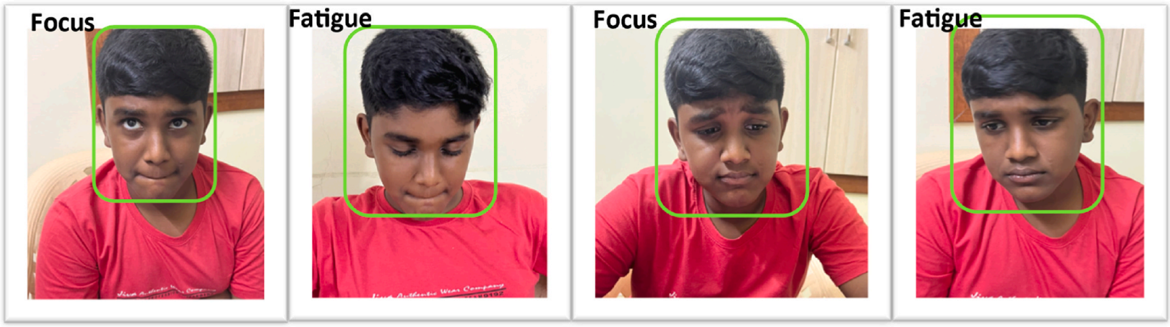


Fig. 9. Example of detection results.

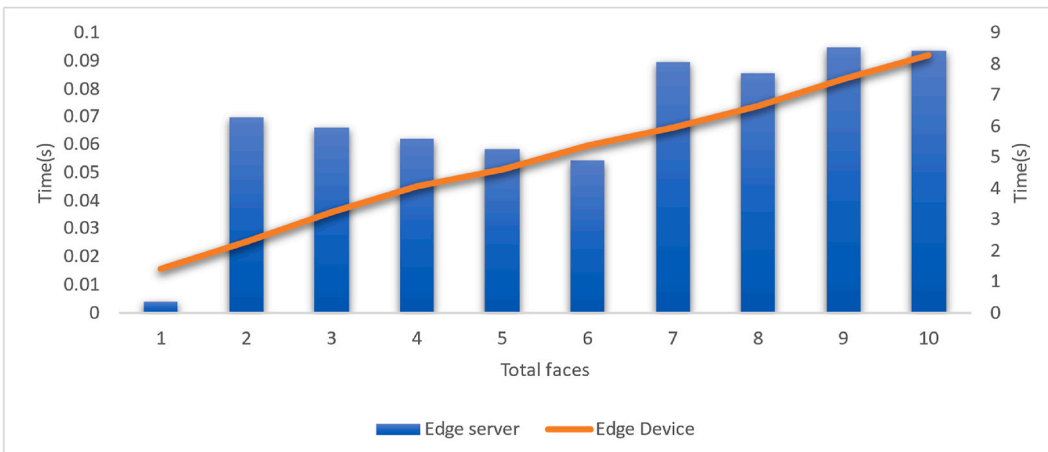


Fig. 10. The ES's and ED's-FD time.

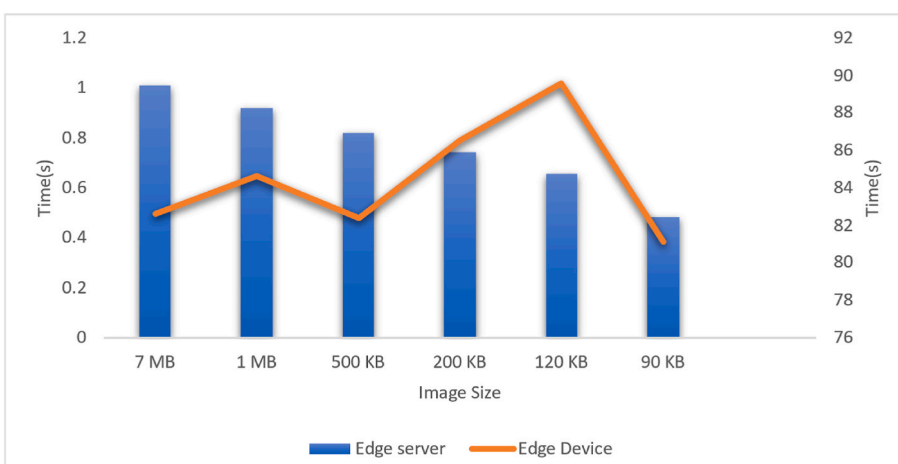


Fig. 11. Variation in FD time, including image file size.

5. Conclusion and future work

In conclusion, the development of Edge-based Online Learning (EOL) has given students more ease and flexibility in their academic endeavors. While taking online classes, measuring, and assessing students' involvement has been particularly challenging. This problem can be effectively solved by the proposed Edge-Based Student Attentiveness Analysis System (EBSAAS), which enables teachers to precisely measure and assess pupils' levels of attention using innovative Deep Learning (DP) models. The system uses innovative technology to analyze student attentiveness more quickly and effectively, enhancing learning outcomes and the efficiency of Online Learning (OL). The proposed model states whether the student is fatigued or focused, and the detection is achieved with a speed of 47.23, which is minimal than other models. The proposed model stands for a significant improvement in the evolution of contemporary education by giving teachers a potent tool to enhance their educational approaches and guarantee that students reach their full potential.

Future research can focus on developing models that can generalize well across different domains, including different ethnicities, age

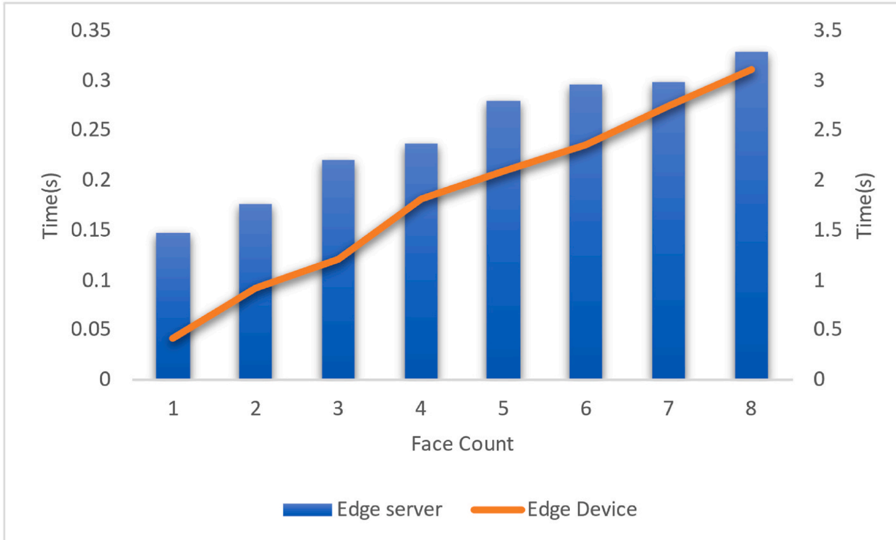


Fig. 12. Performance comparison of the complete processing time between the ES and ED.

groups, and cultural contexts. While current Face Detection (FD) algorithms have achieved high accuracy, there is always research direction for improvement. Research can focus on developing more accurate models that can handle challenging scenarios, such as occlusions, extreme poses, variations in lighting conditions, and diverse demographics.

Credit author statement

Rasheed Abdulkader: Software, Firas Tayseer Mohammad Ayasrah: Resources, Venkata Ramana Gupta Nallagattla: Supervision, Kamal Kant Hiran: Data Curation, Pankaj Dadheech: Visualization, Project administration, Vivekanandam Balasubramaniam: Validation, Sudhakar Sengan: Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Writing - Original Draft, Investigation.

Declaration of competing 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.

Data availability

No data was used for the research described in the article.

References

[1] Pallavi S, Smruti SR. Internet of Things: architectures, protocols, and applications. J Electrical Comput Eng 2017;2017:1–25. https://doi.org/10.1155/2017/9324035.
[2] Kumar S, Tiwari P, Zymbler M. Internet of Things is a revolutionary approach for future technology enhancement: a review. J Big Data 2019;6:111.
[3] Kim J, Maher ML. Conceptual metaphors for designing smart environments: device, robot, and friend. Front Psychol 2020;11:198.
[4] Greengard S. The internet of things. MIT Press; 2021.
[5] Oke A, Fernandes FAP. Innovations in teaching and learning: exploring the perceptions of the education sector on the 4th industrial revolution (4IR). J Open Innovat: Technol Market Complex 2020;6(2):31.
[6] Bajracharya B, Blackford C, Chelladurai J. J. Prospects of Internet of Things in education system. Prospects 2018;6(1):1–7.
[7] Mah PM, Skalna I, Munyeshuri E, Akoko J. Influence of Internet of Things on human psychology (internet of Thoughts) for education, healthcare, and businesses. EAI Endorsed Trans Mobile Commun Appl 2022;7(2).
[8] Fanian F, Khatibi V, Shokouhifar M. A new task scheduling algorithm using firefly and simulated annealing algorithms in cloud computing. Int J Adv Comput Sci Appl 2018;9(2):228–31.
[9] Dias de M, Assunção M, da Silva A, Veith A, Buyya R. Distributed data stream processing and edge computing: a survey on resource elasticity and future directions. J Netw Comput Appl 2018;103(8):1–17.
[10] Simamora RM. The Challenges of online learning during the COVID-19 Pandemic: an Essay Analysis of performing arts education students. Stud Learn Teach 2020;1(2):86–103.
[11] May SG. The excess of content and the decreasing attention span in students Get our Weekly. 2020. p. 1–5.
[12] Balan AK, Jacintos AR, Montemayor T. The influence of online learning towards the attention span and motivation of college students. Mapua University; 2020 (Unpublished undergraduate research).
[13] Hartnett M, Hartnett M. The importance of motivation in online learning. In: Motivation in online education. Massey Univeristy; 2016. p. 5–32. https://doi.org/10.1007/978- 981-10-0700-2_2.
[14] Deng Q, Wu Z, Students' Z. Attention assessment in eLearning based on machine learning. Earth Environ Sci 2018;26(4):741–52.
[15] Huang L, Su J, Pao T. A context-aware smart classroom architecture for smart campuses. Appl Sci 2019;9:1837.
[16] Francisti J, Balogh Z, Reichel J, Magdin M, Koprda S, Molnár G. Application experiences using IoT devices in education. Appl Sci 2020;10:7286.
[17] Sun S, Liu R. A review of research on attentiveness recognition. Sci Tech Inf 2021;4:6–8.
[18] Wang P, Wang D. Research on the application of focus recognition technology in teaching monitoring. Comput Knowl Technol 2020;16:38–42.
[19] Chang Z. Analysis and reflection on concentration of online teaching in vocational colleges. J. Wuhan Eng. Inst. 2020;32:91–4.
[20] Yang SZW, Hu. Design of fatigue monitoring system based on face recognition technology. Instrum Technol 2020;8.
[21] He Z. Research on application of face recognition technology in quality monitoring of college students' lectures. China CIO News 2018;3:102.
[22] Chen H. Research on application of face recognition in learning attentiveness monitoring. J Ind 2020;10(9):70–1.
[23] Zhong ZL. He. Study on online education focus degree based on face detection and fuzzy comprehensive evaluation. Computer Science 2020;47(S2):196–203.
[24] Zhang K, Zhang Z, Li Z, Qiao Y. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process Lett 2016;23(10):1499–503.
[25] Ji Y, Wang S, Zhao Y, Wei J, Lu Y. Fatigue state detection based on multi-index fusion and state recognition network. IEEE Access 2019;7:64136–47.
[26] Bezerra G, Gomes R. Recognition of occluded and lateral faces using MTCNN. dlib and homographies; 2018. p. 1–4.
[27] Sagonas C, Tzimiropoulos G, Zafeiriou S, Pantic M. 300 faces in-the-wild challenge: the first facial landmark localization challenge. In: International conference on computer vision workshops; 2013. p. 397–403.
[28] Karen S, Andrew Z. Very Deep convolutional networks for large-scale image recognition. International Conference on Learning Representations; 2015.
[29] Adhinata FD, Rakhmadani DP, Wijayanto D. Fatigue detection on face image using FaceNet algorithm and K-nearest neighbor classifier. J Inf Syst Eng Bus Intell 2021;7(1):22–30.
[30] Chen L, Xin G, Liu Y, Huang J. Driver fatigue detection based on facial key points and LSTM. Security and Communication Networks; 2021. p. 1–9.
[31] Abhishek G, Alagan A, Ling G, Ahmed SK. Deep learning for object detection and scene perception in self-driving cars: survey, challenges, and open issues. Array 2021;10:100057. https://doi.org/10.1016/j.array.2021.100057.
[32] Charalambos T, Vladan V, Vladimir D, Fredi N. Visual SLAM algorithms and their application for AR, mapping, localization and wayfinding. Array 2022;15:100222. https://doi.org/10.1016/j.array.2022.100222.