

## Research Article

# An Efficient Approach of Face Detection and Prediction of Drowsiness Using SVM

Ratnesh Kumar Shukla,<sup>1</sup> Arvind Kumar Tiwari,<sup>2</sup> and Ashish Kumar Jha <sup>3</sup>

<sup>1</sup>Dr. APJ Abdul Kalam Technical University Lucknow, Lucknow, India

<sup>2</sup>Kamla Nehru Institute of Technology, Sultanpur, Uttar Pradesh, India

<sup>3</sup>Nepal Engineering College, Pokhara University, Lekhnath, Nepal

Correspondence should be addressed to Ashish Kumar Jha; ashishkj@nec.edu.np

Received 27 August 2022; Revised 7 October 2022; Accepted 10 October 2022; Published 17 April 2023

Academic Editor: Dinesh Kumar Saini

Copyright © 2023 Ratnesh Kumar Shukla et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article investigates an issue of road safety and a method for detecting drowsiness in images. More fatal accidents may be averted if fatigued drivers are using this technology accurately and the proposed models provide quick response by recognising the driver's state of falling asleep. There are the following drowsiness models for depicting the possible eye state classifications as VGG16, VGG19, RESNET50, RESNET101 and MobileNetV2. The absence of a readily available and trustworthy eye dataset is perceived acutely in the realm of eye closure detection. On extracting the deep features of faces with VGG16, 98.68% accuracy has been achieved, VGG19 provides an accuracy of 98.74%, ResNet50 works with 65.69% accuracy, ResNet101 has achieved 95.77%, and MobileNetV2 is achieving 96.00% accuracy with the proposed dataset. The put forth model using the support vector machine (SVM) has been used to evaluate several models, and the present results in terms of loss function and accuracy have been obtained. In the proposed dataset, 99.85% accuracy in detecting facial expressions has been achieved. These experimental results show that the eye closure estimation has a higher accuracy and cheap processing cost, as well as the ability of the proposed framework for drowsiness.

## 1. Introduction

As people reside for comforts, circumstances keep on getting improved. Nowadays, everyone feels a primary requirement for a vehicle and thus the number of vehicles has grown rapidly in the society. It has resulted in the growth of a number of enterprises and enhanced economic prosperity.

As the number of vehicles on the road are increasing, the possibilities of the traffic accidents increase particularly due to the reason of sleepy drivers. Long-distance driving requires drivers to travel for extended periods of time to improve the efficiency of driving. Driving a car for an extended period of time exhausts and distracts the driver. It may lead to a fatal traffic accident. Drivers not getting enough sleep on a consistent basis may feel sleepy while driving. Most often such drivers may become victims of accident. Drowsy driving is the leading cause of automobile

accidents globally. The best approach to avoid sleepy driving accidents is to warn the motorist ahead of time. Drowsiness can be diagnosed using a variety of methods. Drowsiness is detected using the face detection and eye recognition [1]. These have sparked scholarly interest in sleepiness detection and deployment of insurance and productivity industries. Facial recognition algorithms have been improved to recognise drowsiness occlusions, and hair and skin color has some of the sources of drowsiness. The possible downstream has impacts of performance on drowsiness, based on potential sleepiness detection technologies, misapplications for driving safety, insurance claim skimming, coverage avoidance, worker monitoring, and job instability. Driver related characteristics are commonly used in the current techniques for identifying a driver's drowsiness. As a result of autonomous driving, the availability of these measures is limited. Physiological signal-based techniques have

appeared to be a feasible choice [2]. Only minimally intrusive methods are appropriate in a dynamic driving environment and convolutional neural network (CNN) has tried to solve this problem in the form of deep learning. The performance of open-source brain-computer interface (OpenBCI) Cyton electroencephalogram (EEG) has detected device and the Arduino open source electrical platform. The driver plays such a crucial part in the vehicle operation on the road. To prevent road accidents, the early warning system will help to wake a driver from the state of sleepiness. Drowsiness detection differs based on the equipment and the method used [3]. Deep learning-based systems have detected tiredness that decreases the road accidents. A tiredness detection technique based on the CNN has been suggested in the article [4]. People might suffer from a variety of mental diseases, such as schizophrenia, depression, and intellectual difficulties. In today's world, social media is the one thing that cannot be replaced. Through online chats, users may express their feelings, opinions, and thoughts to others on this platform. The data that are gathered everyday on the internet and are growing daily are the focus of many different strategies today [5]. The system discriminated human and nonhuman person's eyes by training the first network, then used the second network to identify the location of the feature points of the eye. Estimating the eye opening position based on the features, the network points out the score of the driver's sleepiness state [6]. The most prevalent kind of "synthetic media," which includes images, sounds, and films that appear to be created using conventional techniques but actually were created using the cutting-edge software, appears to be Deepfakes. In this chapter, we looked at a few publications to learn more about DF and its uses [7].

## 2. Related Work

One of the leading causes of accidents is drowsy driving. Major accidents can occur when sleepy drivers are unable to react to hazardous conditions. It is critical to identify driver's tiredness early and correctly in order to reduce accidents caused by sleepy driving. Early identification of sleepy driving is crucial to reduce such incidents. The effects of driver sleepiness on driving performance, behavioural indicators, and physiological markers have been found. This demonstrates the feasibility of implementing a sleepiness detection model based on the hybrid sensing and without contact sensors, as well as the prospect of extremely accurate early identification of a fatigued driver [8]. This model has detected sleepiness which in turn detected all levels of exhaustion from mild to severe. The posture index was especially beneficial when paired with the blink data since it was more sensitive to mild tiredness than traditional information and could compensate for the limitations of blink data [9]. Practical application of the proposed technique is one of the work limitations. Using 32 channels in a real-world context is not an opinion, this is a physical constraint that makes these devices hard to commercialize [10]. This technique is used for detecting tiredness based on the detection yawns and blinks at the same time. Establishment and foundation have many

discriminating physiological inputs at the same time. These are identified automatically from sources to merge corresponding multioutputs, short-term energy, and kurtosis. Finally, tiredness is recognised by monitoring many yawn and blink signals at the same time [11]. To discriminate between tiredness and alertness, deep neural networks (DNNs) are used. The best suited channels of brain activity detection are determined by CNN using colour map pictures. DNNs from the right lateral prefrontal cortex are used to get the findings. The CNN architecture has achieved a 99.3% average accuracy, showing that the model can distinguish between sleepy and nondrowsy pictures [12]. CNN has employed different aims in real-time application to obtain high accuracy and recognise the driver's falling asleep status as an indication of tiredness. The outcomes of the experiments reveal that eye closure estimate has high accuracy and low processing complexity. The proposed framework is able to detect sleepiness [13]. Twitter is the biggest storehouse for data collected from various blogging platforms including Facebook, Weibo, YouTube, Instagram, and others. Additionally, text, audio, and video were gathered from repositories. The sentiment analysis collects user attitudes through opinion mining and categorises them as favourable, negative, or neutral [14]. By feeding, the recurrences are plotted as input features into a CNN. The efficiency has been found for drowsy/awake categorization studied in CNN. In drowsy/awake categorization, the CNN models based on rectified linear units and recurrence plots outperformed other conventional models, with ECG accuracy enhanced by 6–17% and photo pothys monogram accuracy is observed to have been improved by 4–14% [15]. The wearable dry electrode prototype, a rather pleasant forehead setup, which captures images in vigilance dynamics easily. The suggested approach achieves the best mean and correlation coefficients of 71.18% and 66.20%, respectively, in real-time circumstances and practically in laboratory. To assess the simulated-to-real generalisation, cross-environment tests are applied, and the best correlation coefficient is determined to be 53.96% [16]. Mathematics of SVM, tree-based classifiers, and neural networks will go into great detail about ensemble algorithms with appropriate applications [17]. A deep cascaded convolutional neural network (DCNN) has been created to recognise face area and it overcomes the problem in low accuracy using an artificial feature extraction. The Dlib toolkits are utilised to identify the frontal facial landmarks in frame. In this process, the authors have used a new parameter known as eyes aspect ratio (EAR) to analyse the tiredness in the current frame, according to the eyes landmarks. Practically, there are different modules to identify the changes in the size with the help of EAR. The first module used SVM with the EAR as input to build a unique weariness state classifier [18]. After the features have been determined, they are classified using the following classifiers: bagging reduced error pruning, K-closest neighbour, AdaBoost, Gaussian Naive Bayes (GNB), stochastic gradient descent, and SVM. Thus, stacking of LGBM and random forest has been found to produce the best forecasts across all three datasets [19].

The above Table 1 shows the different positions of sleepiness and expected levels thereof. In the article, all of the expected eye positions have been provided for the calculation of the current position of opened and closed eyes using the EAR. The main aspect of this table is to carry out the investigation of the position of eye for assessing the state of sleepiness [20]. The position of eyelids is the outcome of the calculation of EAR for detecting sleepiness. Actually eyes present different states such as open eyes, closed eyes, and partially open or closed in a random time. The closed and partially opened or closed levels are marked as sleepiness [21].

### 3. Methods and Data Description

Drowsiness detection systems have been developed by using subjective, vehicle-based, behavioural, and physiological data. The subjective measurements have been commonly used to establish a sleepy baseline or reference point. Observation and self-ratings are examples of these measurements. Experts or qualified ratters' keep an eye on the driver to assess his or her present level of sleepiness based on observer evaluations. A driver's sleepiness can be detected by monitoring the driver. The upper body or facial region of the driver is examined for sleep-induced drowsiness. Measures like steering wheel movement and lane position standard deviation are used by vehicle-based systems to assess driving behaviour. In behavioural-based approaches, a camera is utilised to monitor the driver's face. Image processing technologies have been utilised to discover drowsiness signs and early symptoms [21]. The eye movements, facial emotions, and head posture are all evaluated. Data from brain activity, heart activity, eye activity, and muscle tone are used in the physiological methods.

*3.1. Datasets of Drowsiness.* The facial drowsiness dataset is included for closed and open eyes, and to design for an improved performance of facial recognition in the existing models. There are several sleepiness datasets that can offer decent results in real-time. There are the following datasets used in sleepiness detections.

*3.1.1. ZJU Dataset.* ZJU gallery from the ZJU eye blink database has a collection of photos of 80 video clips in video libraries. They made the following four movies for each participant: first without including glasses, second one has thin rim glasses, third one has black frame glasses, and fourth one with and without spectacles of left and right eyes photographed individually. Table 2 are containing 4,841 photos of the collection, including 2,458 open and 2,383 closed eyes and the trained networks with subsample pictures with their shots geometrically normalised to  $24 \times 24$  pixels [22].

*3.1.2. MRL Dataset.* The recognition of eyes, their components, estimation of the gaze, and frequency of eye-blinking are all critical challenges in computer vision. Solving the

issues in the field of driver behaviour, which has resulted in the acquisition of a huge quantity of testing data acquired in real world scenarios. The MRL eye dataset possesses a large database of human eye pictures. These collections are included for low and high resolution infrared images captured in a variety of lighting conditions and a variety of equipment. Table 3 contains 4,841 images, comprising of 24,000 closed eyes and 24,000 open eyes. The dataset is suitable for testing a wide range of features or trainable classifiers [23].

*3.1.3. Proposed Dataset.* The proposed dataset in Table 4 contains 3,893 images of open eyes and closed eyes, and they are cropped and labelled by using the annotations supplied. 2,701 open-eye pictures and 1,192 closed-eye shots have been covered up. The authors cropped the annotated faces with margin around the face and eye bounding boxes in their suggested datasets. Figure 1 depicts the collection of the proposed dataset pairings of closed eyes and open eyes pictures.

*3.2. Model Description.* This article has used the previous pertained models as feature extractors of the faces to extract eyes features from informative regions. The driving time is critical for improving the model's performance. Obviously, the longer a driver drives in monotonous conditions, the more likely he or she is to fall asleep. This is why highway drivers are usually advised to take a rest after two hours of driving. As a consequence, the model has been made to learn a linear relationship between elapsed time and the amount of time before the critical state occurs. Some people may feel drowsy at one point, and then get awake again. The behaviour shows that there is no obvious linear relationship between the driving duration and the time before reaching a certain degree of drowsiness.

*3.2.1. VGG16.* The VGG model is using the ImageNet dataset for image classification. K. Simonyan and A. Zisserman from the University of Oxford who proposed VGG16 have used a CNN model in their article "Very Deep Convolutional Networks for Large-Scale Image Recognition." ImageNet dataset contains 14 million pictures divided into thousand classes, and this model achieved 92.7% accuracy in the top-5 tested model. ImageNet dataset comprises of more than 15 million high-resolution images organised into more than 22,000 categories. The photographs are collected from the Internet and categorised by users using Amazon's Mechanical Turk crowd-sourcing technology [21].

*3.2.2. VGG19.* A deep learning has made use of CNN architecture for image categorization. The VGG architecture has included with 16 convolutional layers and three fully linked layers. The next great step has been put forward after Alexnet's ImageNet Challenge triumph in 2012. The VGG19 abbreviation refers to the fact that there are 19 different strata in all. VGG19 is a 19-layer convolutional neural network that has trained on millions of image samples that

TABLE 1: Description of different subject and level of sleepiness position in images.

Level	Subject	Level of sleepiness
1	Feel busy	Not feeling sleepy
2	Critical, alert and wakeful	Extremely sleepy and fighting sleep
3	Work smart but not concentrate on the peak level	Some effort for sleep alert
4	Wakeful and relaxed but not alert	Optimal sleepiness of open
5	Small sparkle	Showing sleepiness features
6	Difficult to wake up, foggy, and lose the track	Alert
7	Woozy and sleepy	Very alert
8	Struggling for sleep	Extremely sleep
9	Not awake after alert	Alert at extreme level

TABLE 2: Analysis of open eyes and closed eyes in ZJU dataset.

Open eyes	Closed eyes	Total images
2383	2458	4841

TABLE 3: Extraction of open eyes and closed eyes in MRL dataset.

Open eyes	Closed eyes	Total images
24000	24000	48000

TABLE 4: Proposed dataset classification with open and closed eyes.

Open eyes	Closed eyes	Total images
2701	1192	3893

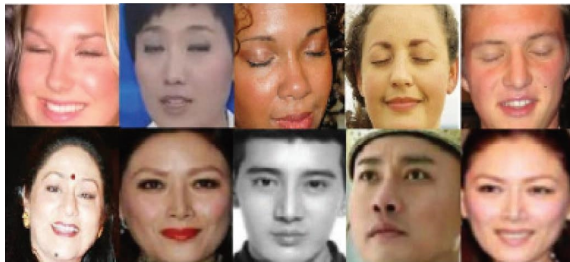


FIGURE 1: Sample set of the proposed dataset with open eyes and closed eyes.

normalise images using the zero-centre architectural style. The fully linked layer has included ReLU, Dropout, Softmax, and classification output.

**3.2.3. RESNET50.** A CNN with 50 layers like the frequently used ResNet-50 model is a deep residual network. ResNet has artificial neural networks (ANNs) that build networks by stacking residual blocks on top of each other. ResNet comes in a variety of flavours, each with a different number of layers but the same basic idea. ResNet 50 is a ResNet variation capable of working with up to 50 neural network layers. ResNet was built specifically to address this issue. To increase model accuracy, deep residual nets use residual blocks. The strength of this form of neural network is the notion of “skip connections,” which lies at the heart of the residual blocks [24]. As a reference architecture for learning food-domain

properties, we used a residual network with 50 layers. The proposed database features outperform over the learnt ones from previous food databases and the massive ImageNet picture database [25].

**3.2.4. RESNET101.** ResNet101 is a deep convolutional neural network with 101 layers. ResNet101 is trained with ImageNet database including million photographs with a pretrained version of the network. This network can classify photographs into thousands distinct image categories, such as mouse, keyboards, pencils, and a wide range of animals. ResNet101 has amassed a library of rich feature extraction for a wide range of images including  $224 \times 224$  pixels size images [26].

**3.2.5. CNN.** CNN is a computationally efficient model that analyses features to discover image-related issues using specific convolution and pooling procedures. CNN is a classification framework that divides the pictures into labelled categories. The many layers of CNN extract picture information before learning to identify the images. As a result, typical outputs of CNN indicate the classes or labels of the classes that the CNN has learned to categorise [27].

**3.2.6. SVM.** SVM is a linear model used for classification and regression problems of the objects. SVM is very useful for solving the problems of linear and nonlinear categories and may be used in a variety of situations. The approach separates the data into classes by creating a line or hyperplane, which is a core notion in SVM. The SVM is a machine learning method that takes data and, if possible, draws a line that separates the classes. Assuming that the user supplies a dataset of red rectangles and blue ellipses (positives from negatives) to SVM to differentiate, and the SVM creates a hyperplane to identify the dataset into two groups (red and blue) [28].

**3.2.7. MobileNetV2.** MobileNetV2 is an extended version of CNN designed to function on mobile devices. It has included residual networks to solve the problem of inverted residual structure with bottleneck. MobileNetV2 has intermediate expansion layer filters that extract features using lightweight and depth wise convolutions as a source of nonlinearity. It boosts mobile model performance across a wide range of

activities and benchmarks, as well as model sizes. A new module with an inverted residual structure has been added to MobileNetV2. Nonlinearity in thin layers is no longer an issue. The state-of-the-art object recognition and semantic segmentation results are also achieved using MobileNetV2 as the backbone for feature extraction [29].

#### 4. Approaches of the Proposed Model

The proposed model is working on detection of the drowsiness position. The authors are applying different steps to find out the correct position of the eyes in the faces of the persons. Eye is identified as the exact value of the driver, when driving the car. Another feature is not giving good result to identify the position of drivers. To begin, the authors are collecting the input image using camera or database and detect the face image.

After collection of face image, the useful image is identified. Then, multitask cascaded convolutional networks (MTCNN) is applied for classifying the faces with their features like eyes, mouth, and nose from the facial region. MTCNN having been applied can detect the eyes and apply the EAR theorem to detect the accurate position of the eyes such as open or close [30]. In the previous section, there is a different level of position to identify the status of the eyes and new metric termed as eyes aspect ratio (EAR) is created [31]. It is determined with the use of facial landmark coordinates (open or closed). Two sets of the EAR are fetched from the face recognised by MTCNN in images and trained for particular face, representing eyes-open and eyes-closed. After that, the SVM classifier compares the features of eyes and classifies if drowsiness exists or not. The landmarks of the face are then obtained using the Dlib toolkit [32]. The detection is carried out with the help of logistic regression. The image orientation detection algorithm was tested using over 100,000 pictures from the Scene UNDERstanding (SUN) collection. This technique surpasses similar algorithms utilizing art, and it recognises the accuracy of correct orientation of visual contents comparable to human observers [33]. The aspect ratio of the eyes may be computed using the eye coordinates. Finally, the one-of-a-kind SVM can classify the present condition of the frame's eyes [34]. The performance of the algorithms may be significantly improved by using the SVM and the primary contributions in this investigation are as follows:

- (1) Detection of face from image dataset using digital camera.
- (2) Application of MTCNN to classify the faces from images, which improves the performance significantly.
- (3) Detection of the eyes and application of EAR to determine the accurate position of the eyes. EAR works in different levels and the suggested technique improves significantly in both accuracy and speed.
- (4) Application of SVM to classify the eyes feature for their openness or closeness.
  - (a) If eyes are closed, next step is followed.

- (b) If eyes are open, step 1 is followed.

- (5) Finding a way out of the sleepiness position.

The suggested method takes individual differences into account, and it is more reasonable and accurate, according to the results of comparative studies. It is the same as a start-up procedure. Face traits, particularly the size of the eyes, exhibit notable variances, as is commonly observed. As a result, a simulator is employed to collect data by asking the user to open and close his or her eyes for a brief interval of time. MTCNN is built to detect the face of the driver in the current frame by applying the picture data. Facial landmarks can also be retrieved using the Dlib toolbox. As illustrated in Figure 2, the following two types of EAR may be determined: when the eyes are open, EAR1 is computed, and when the eyes are closed, EAR2 is calculated. Finally, a customised SVM is trained for the given person to determine whether the eyes are open or closed using the two forms of EAR as positive and negative examples.

##### 4.1. MTCNN (Multitask Cascaded Convolutional Networks).

Face detection is one of the most significant computer vision methods for sleepiness detection. In practice, a system for detecting sleepiness must not only be accurate but also be rapid. Deep learning technologies, notably the convolutional neural network model, improves photo recognition accuracy dramatically, and the algorithm performance, on the other hand, is delayed by the intricate network structure. The MTCNN has been developed for face detection. Different loss functions are employed for different tasks. The loss function quantifies the difference between index of expected output and input label that was actually marked and applied face classification tasks, as well as face area box bounding fitting tasks, are part of the training process, and the image is converted into matrix form with height, width, and colour [35]. After that, a rectangle box is drawn around the face and crucial spots such as the eyes, nose, and mouths are chosen. It is a straightforward approach to tackle and apply traditional feature-based algorithms such as the MTCNN classifier. An excellent job in image datasets has been proven by it [36]. Deep learning algorithms have set a desired level performance in the recently state-of-the-art on common benchmark for face detection datasets. Face detection and recognition is a challenging task in computer vision problem that entails recognising and finding persons in image photographs.

Figure 3 is showing the output after applying the OpenCV package and the traditional feature-based cascade classifier may be used to recognise faces. MTCNN has undergone three tasks to train the used CNN detectors like face/nonface classification, bounding box regression, and facial landmark localization.

##### 4.2. Face Classification.

The learning goal of expression has made use of two-classified problems for cross-entropy loss. Cross-entropy determines the amount of bits necessary to describe and transmit to average event from one distribution and compared to another distribution based on the concept

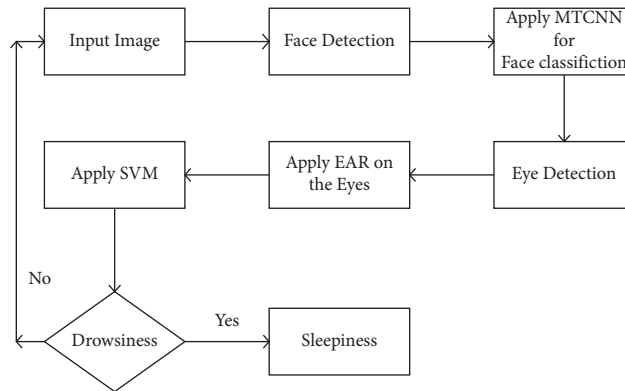


FIGURE 2: Apply proposed model to detect drowsiness condition.

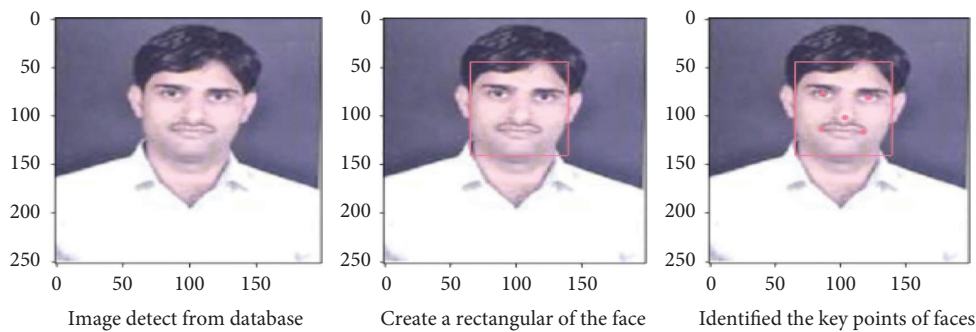


FIGURE 3: The figure given above is showing output after applying the MTCNN and found out the rectangular and key points of the faces.

of entropy in the information theory. There are additional bits necessary in cross-entropy of  $Q$  from  $P$  which is the number to represent an event, and it used  $Q$  instead of  $P$  when analysing a target of underlying probability distribution  $P$  and a close approximation of the target distribution  $Q$  is carried out.

$$H(P, Q) = -\sum x \ln x P(x) * \log(Q(x)), \quad (1)$$

where  $H(P, Q)$  is indicating a cross-entropy function,  $P$  indicates a target distribution, and  $Q$  indicates approximation of the target distribution.  $P(x)$  is the probability of event  $x$  in  $P$ ,  $Q(x)$  is the probability of event  $x$  in  $Q$ ,  $\log$  is the base-2 logarithm, and the results are calculated in bits.

**4.3. Bounding Box Regression.** It predicts the candidate window and the nearest ground truth bounding boxes left top, height, and breadth in images. Each sample is treated with the Euclidean loss, and the learning goal is stated as a regression problem. The Euclidean loss is minimised by posing facial landmark recognition as a regression problem, similar to the bounding box regression task. Three variables absolutely necessary are set up as follows:

- (i) To detect the faces, the frontal face detector method can identify one or more faces.
- (ii) A predictor that can recognise keypoints in faces inside the variable and shape.

- (iii) For capturing photos from the webcam, the cv2.Video is used to capture the object. A parameter with the value 0 was also specified for taking photographs from a camera.

**4.4. EAR (Eye Aspect Ratio).** Using the Dlib toolbox, face landmarks were collected. As indicated in Figure 4, there are six dots dispersed around each eye to identify its location. Between open and closed states, the distribution of ocular landmarks differs significantly. The programme eye aspect ratio was used to track the frequency of blinking.

The following formula may be used to calculate EAR based on the position of ocular landmarks:

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

The coordinates of eye landmarks are indicating  $P_i$  where integer  $i$  equals to 1, 2, 3, 4, 5, and 6. When the eyes are observed to be open, the EAR is more than 0.2, as demonstrated in Figure 4. The EAR, on the other hand, is less than 0.2 for the closure of the eye. As previously mentioned, the condition of the eyes might indicate whether or not he or she is sleepy, because there are huge disparities in the amount of time that eyes are closed while he or she is awake and when he or she is tired. To define the shape of the pupil, a method of ellipse fitting is presented. This approach is segmenting the pupil using a standard imaging process as seen in Figure 4 given above. The white pixels indicate a form

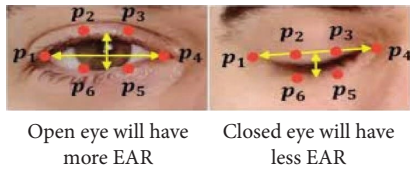


FIGURE 4: Detecting the value of open eyes and closed eyes after applying eye aspect ratio (EAR).

of eyes that are then placed into the ellipse. Finally, the eyes were assessed using the ratio of major and minor axes of ellipse.

**4.5. Drowsiness Detection.** An experiment is conducted to show the detection. Individual distinct events are taken into account in this experiment to create a unique classifier. It is a soft hand start-up procedure and an adjustable threshold rather than a set one. To make it work, images for open and close eyes for a few seconds are taken, and then the MTCNN and Dlib toolbox to acquire two sets of EAR are applied. The SVM classifier is trained by using two sets of data as input. The structural risk minimization criteria are applied in SVM, a machine learning model. It is a linear classifier model with a feature of being the biggest interval of defined space. We leverage deep features derived by CNNs linked with support vector machines as an alternative to end-to-end classification because the labelled data is typically scarce in real-world applications. Data with open eyes are labelled as positive, whereas data with closed eyes are labelled as negative. In an images eyes cannot abruptly changes, i.e., eyes cannot shifted from open to closed and closed to open in a single frame with immediate time. The EAR is containing six consecutive frames, which are concatenating into a single feature vector.

## 5. Experimental Analysis and Results

Face detection is an essential study direction in this target detection publication. The face's position in the provided picture is returned. Facial recognition using deep learning requires three phases: data input, feature extraction, and face feature detection (including eye and mouth), with feature extraction being the most important. Rather than a complex network, the design is made up of subnetworks. The algorithm can run quicker since each subnetwork has fewer filters but better filter discrimination. However, we observed that having worthless facial landmark information might be an impediment to its performance. As a result, we create a one-of-a-kind convolutional neural network that can recognise a driver's face. Because of the lightweight frame and the lack of facial landmarks, it is more efficient. We are building an image drowsiness model in the figure given above to see if the driver is drowsy or not. The authors utilise a number of algorithms to classify the driver's weariness. The recommended model will provide a better image drowsiness result. As a result, scaling the original image to various sizes is an option. The data from video or picture is initially collected in the proposed diagram. Then they use MTCNN to figure out who is in the video or picture. We apply the

facial landmarks algorithm on the faces after successfully collecting the real-time image. This programme recognises the many features of the faces. This algorithm recognises facial features such as stance, eye, and mouth. These three facial features are crucial in determining whether or not someone is drowsy. We use CNN to extract the features of the face, mouth, and eyes after collecting these three characteristics. The eyes are a very crucial part of the faces when they are drowsy. So, in order to get the exact value of the eyes, we are using the EAR. We use SVM once we have collected these three characteristics. Machine learning (ML) algorithms for classification are known as SVM and SVM has a supervised learning method for classifying objects including many categories. We use EAR in SVM to determine if the eyes are closed or open, as well as whether the mouth is yawning or not. Following this identification, we apply the SVM on the faces and if the faces are not facing forward, they are in a drowsy position. If these characteristics are true, then they forward it to the next stage as sleepiness, and another on is sent to step one and the procedure continues to the next image. It has the ability to alter the behaviour and pose a significant risk of injury to other users. By tracking the length to identify the state of objects, it is feasible to prevent the accidents caused by the driver's fatigue to some extent. However, this technology cannot determine whether or not an object is sleepy in real time. A detecting system has emerged as a realistic option with the progress of information technology. As a result, research into intelligent drowsy object detection has real-world ramifications. In general, regularisation is the only necessity when our network is at risk of overfitting. This can occur if a network is too large, and if the user trains for too long or if the user lacks sufficient data. Dropout is simple to create if the convolutional network has fully linked layers at the end. Convolutional layers require less regularisation to begin with since they have fewer parameters. Activations can also become strongly linked as a result of the spatial correlations stored in feature maps. Dropout becomes unsuccessful as a result of this. VGG16, VGG19, RESNET50, RESNET101, and MobileNetV2 were large models with completely linked layers at the network's conclusion. Dropout between completely linked layers was used to counteract overfitting. Modern convents have decreased the model size while enhancing performance by replacing dense layers with global average pooling. The batch-normalized the model's good performance backup the idea, that batch normalisation should be utilised between convolutions. Furthermore, dropout between convolutions should be avoided, as models with dropout performed worse than the control model. Generic dropouts are working well, when nodes/pixels have spatial relationship, dropping whole activations/features at random works extremely well. It is a simple Keras Dropout layer parameter change that works with a variety of convolutional models.

In the above Figure 5 is showing the model accuracy of VGG16, VGG19, RESNET50, RESNET101, MobileNeV2, and SVM model using proposed dataset. This performance ratio is performing better result using SVM. Also Figure 6, graphs are showing the comparative analysis of the

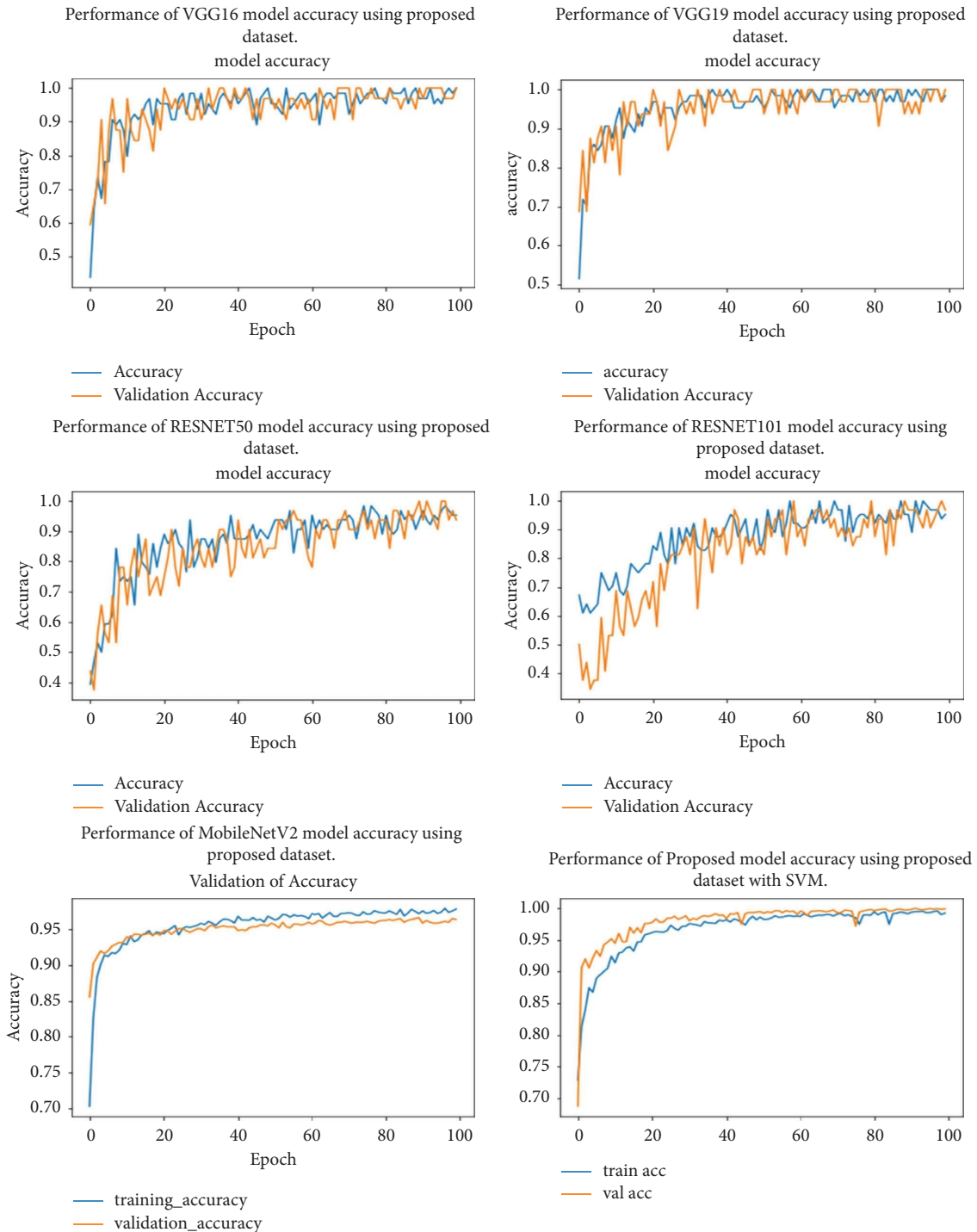


FIGURE 5: Performance accuracy of VGG16, VGG19, RESNET50, RESNET101, MobileNetV2 and proposed model using proposed dataset.

performance of loss function and validation loss in different models. VGG16, VGG19 is performing better than RESNET50 and RESNET101. These models are pretrained models perform with proposed datasets. But they are not providing good result compare than MobileNetV2 and proposed model with SVM. The pretrained model is working on softmax to categories the object in proposed model with SVM to identifying the object. SVM is classifying the features of the object in different class. The proposed model

with SVM is providing better result in loss-function and validation loss-function in compare the above VGG16, VGG19, RESNET50, RESNET101 and MobileNetV2. The proposed model with SVM is providing better result in loss function and in accuracy compare to above five pretrained models.

The Figure 7 is showing the result of after compiling 100 and 500 epochs in proposed model with proposed dataset. When the epochs are increasing the value of loss function is



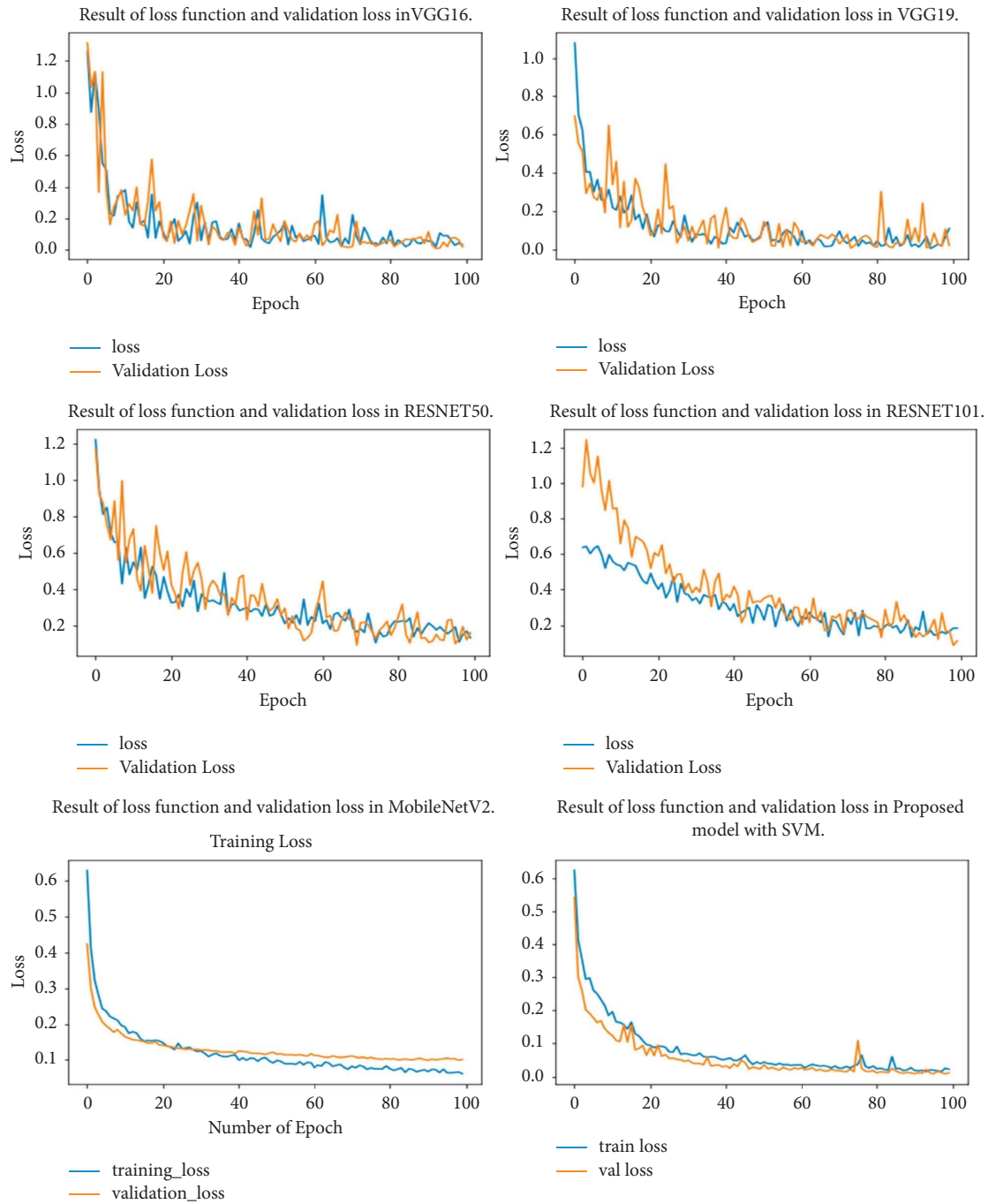


FIGURE 6: Performance of loss function in VGG16, VGG19, RESNET50, RESNET101, MblenetV2, and proposed model with proposed dataset.

decreasing and accuracy is also increasing, after comparing the 100 and 500 epochs. Epochs are working a set of feature called cluster to work together. After compiling the 500 epochs, we find an accurate performance in loss function and accuracy. So we can say the proposed model with SVM is working well compare to VGG16, VGG19, RESNET50, RESNET101 and MobileNetV2. Proposed approach has provided better performance compare to another five pre-trained model showing in the Figure 8. They will be providing exact value with proposed dataset.

## 6. Comparative Analysis of Different Models with Datasets

The Table 5 and Figure 8 exhibit the performance of the VGG16, VGG19, RESNET50, RESNET101, MobileNetV2 and proposed model with SVM. VGG16 and VGG19 yield a better result compared to RESNET50 and RESNET101. But proposed model generates better result compared to VGG16, VGG19, RESNET50, RESNET101 and MobileNetV2. They show different model with working different

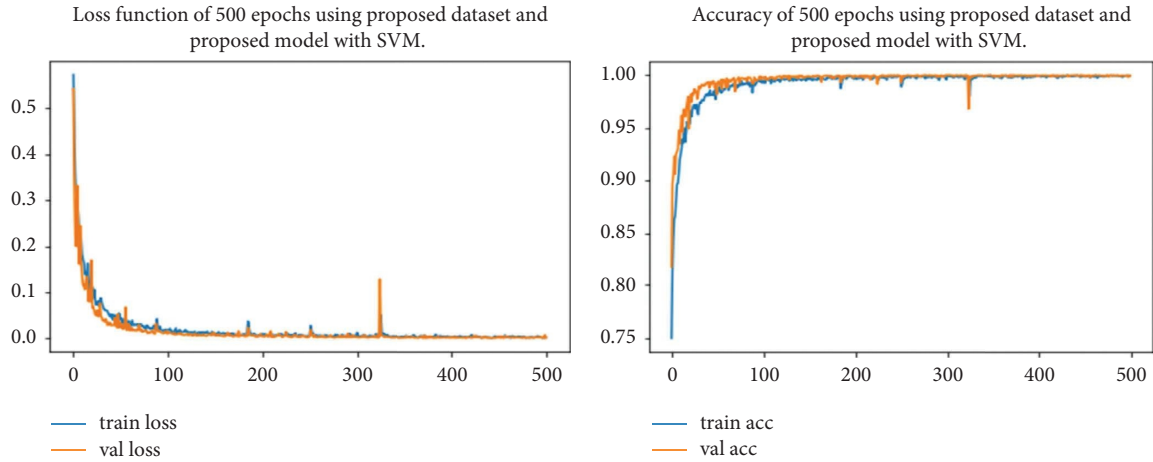


FIGURE 7: Performance of accuracy and loss function of proposed model with proposed dataset using 500 epochs.

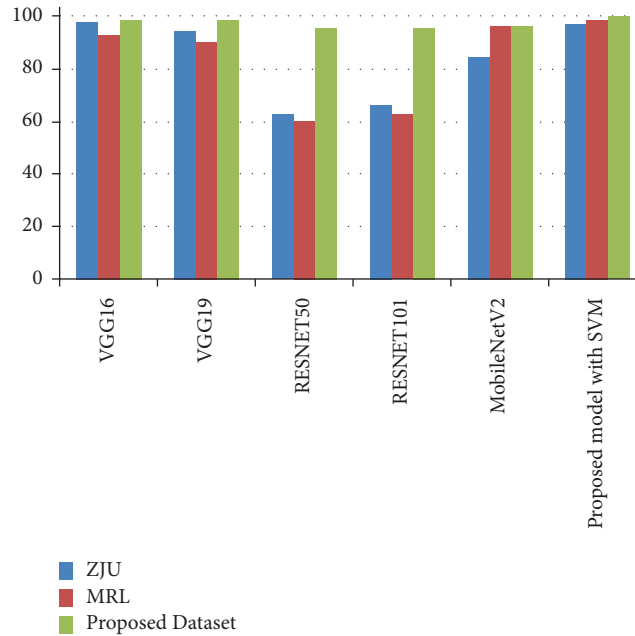


FIGURE 8: Above figure is showing the graphical representation of the VGG16, VGG19, RESNET50, RESNET101, MOBILENETV2, and the proposed model with SVM.

TABLE 5: Performance of the models VGG16, VGG19, RESNET50, RESNET101 and proposed model including different dataset.

S.No.	Technique	Dataset	Epochs	Accuracy
1	VGG 16	ZJU [22]	100	98.15
		MRL [23]	100	92.56
		Proposed dataset	100	98.68
2	VGG19	ZJU [22]	100	94.96
		MRL [23]	100	90.23
		Proposed dataset	100	98.74
3	RESNET50	ZJU [22]	100	62.50
		MRL [23]	100	60.33
		Proposed dataset	100	95.69
4	RESNET101	ZJU [22]	100	65.89
		MRL [23]	100	62.31
		Proposed dataset	100	95.77

TABLE 5: Continued.

S.No.	Technique	Dataset	Epochs	Accuracy
5	MobileNetV2	ZJU [22]	100	84.70
		MRL [23]	100	95.88
		Proposed dataset	100	96.00
6	Proposed model with SVM	ZJU [22]	100	97.54
		MRL [23]	100	98.76
		Proposed dataset	100	99.56
			500	99.85

datasets. The proposed approach is applied to the datasets from ZJU, MRL and proposed database; they have achieved recognition accuracy of 99.85% with proposed dataset. This article employs the five preprogrammed models VGG16, VGG19, ResNet50 and ResNet101. Extracting the deep features of eyes with VGG16 has resulted in 98.68% accuracy, VGG19 shows 98.74% accuracy, ResNet50 shows 95.69% accuracy, ResNet101 exhibits 95.77% accuracy and MobileNetV2 shows accuracy of 96.00% with own dataset. All result shown in this paper with accuracy complies with their approved results.

## 7. Conclusion

Research on sleepy driving detection algorithms is one of the most essential strategies to prevent road accidents. As it is aware to all, there are substantial individual variances between people, particularly when it comes to the eyes. While researching computer vision algorithms, individual modifications of Table 5 and Figure 8 must be addressed. In this paper, a novel picture of sleepiness detection technique takes individual attributes into account. To begin, an MTCNN is created to retrieve a face from an image, avoiding the artificial feature extraction process used in typical face identification algorithms. According to experimental data, face detection accuracy can reach 99.65% in 100 epochs and 99.85% in 500 epochs. In comparison to VGG16, VGG19, RESNE50, RESNET101, and MobileNetV2, the proposed model with SVM produced superior results. Based on the Dlib toolkit, a new parameter EAR is offered to analyse the status of the object's eyes. Experiments indicate a high association between the EAR and the size of an object's eyes, confirming our assumptions. Finally, in the study, a module is developed while taking individual variances in the eyes into account. A unique classifier based on SVM is constructed for each image, and the condition of the eyes is analysed using a prelearned classifier. Experiment findings show that put forth approaches consistently outperforms state-of-the-art methods while retaining real-time performance on both publicly available and self-built datasets. The comparative study reveals for the suggested model to have outperformed all five previous existing models. The fundamental contribution of this study is to monitor with drowsiness.

## Data Availability

ZJU dataset is available at (<https://github.com/elmino9ykl/ZJU-Dataset>) and MRL dataset (<https://mrl.cs.vsb.cz/eyedataset>).

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this study.

## References

- [1] M. Zhu, J. Chen, H. Li, F. Liang, L. Han, and Z. Zhang, "Vehicle driver drowsiness detection method using wearable EEG based on convolution neural network," *Neural Computing & Applications*, vol. 33, no. 20, pp. 13965–13980, 2021.
- [2] X. Zhao, C. Meng, M. Feng, S. Chang, and Q. Zeng, "Eye feature point detection based on single convolutional neural network," *IET Computer Vision*, vol. 12, no. 4, pp. 453–457, 2018.
- [3] S. S. Sharan, R. Viji, R. Pradeep, and V. Sajith, "Driver fatigue detection based on eye state recognition using convolutional neural network," in *Proceedings of the 2019 International Conference on Communication and Electronics Systems (ICCES)*, pp. 2057–2063, IEEE, Singapore, July, 2019.
- [4] F. Zhang, J. Su, L. Geng, and Z. Xiao, "Driver fatigue detection based on eye state recognition," in *Proceedings of the 2017 International Conference on Machine Vision and Information Technology (CMVIT)*, pp. 105–110, IEEE, Singapore, February, 2017.
- [5] B. Chilwal and A. K. Mishra, "Extraction of depression symptoms from social networks," *The Smart Cyber Ecosystem for Sustainable Development*, vol. 29, pp. 307–321, 2021.
- [6] J. Gwak, A. Hirao, and M. Shino, "An investigation of early detection of driver drowsiness using ensemble machine learning based on hybrid sensing," *Applied Sciences*, vol. 10, no. 8, p. 2890, 2020.
- [7] S. Rastogi, A. K. Mishra, and L. Gaur, "Detection of DeepFakes using local features and convolutional neural network," in *DeepFakes* CRC Press, Boca Raton, FL, USA, 2022.
- [8] M. Sunagawa, S. I. Shikii, W. Nakai, M. Mochizuki, K. Kusukame, and H. Kitajima, "Comprehensive drowsiness level detection model combining multimodal information," *IEEE Sensors Journal*, vol. 20, no. 7, pp. 3709–3717, 2020.
- [9] J. R. Paulo, G. Pires, and U. J. Nunes, "Cross-subject Zero calibration driver's drowsiness detection: exploring

- spatiotemporal image encoding of EEG signals for convolutional neural network classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 905–915, 2021.
- [10] C. Zhang, X. Wu, X. Zheng, and S. Yu, "Driver drowsiness detection using multi-channel second order blind identifications," *IEEE Access*, vol. 7, pp. 11829–11843, 2019.
- [11] M. A. Tanveer, M. J. Khan, M. J. Qureshi, N. Naseer, and K. S. Hong, "Enhanced drowsiness detection using deep learning: an fNIRS study," *IEEE Access*, vol. 7, pp. 137920–137929, 2019.
- [12] M. Hashemi, A. Mirrashid, and A. B. Shirazi, "Driver safety development: real-time driver drowsiness detection system based on convolutional neural network," *SN Computer Science*, vol. 1, no. 5, pp. 1–10, 2020.
- [13] H. Lee, J. Lee, and M. Shin, "Using wearable ECG/PPG sensors for driver drowsiness detection based on distinguishable pattern of recurrence plots," *Electronics*, vol. 8, no. 2, p. 192, 2019.
- [14] V. Arya, A. K. M. Mishra, and A. González-Briones, "Analysis of sentiments on the onset of COVID-19 using machine learning techniques," *Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 11, no. 1, pp. 45–63, 2022.
- [15] W. L. Zheng, K. Gao, G. Li et al., "Vigilance estimation using a wearable EOG device in real driving environment," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 1, pp. 170–184, 2020.
- [16] F. You, X. Li, Y. Gong, H. Wang, and H. Li, "A real-time driving drowsiness detection algorithm with individual differences consideration," *IEEE Access*, vol. 7, pp. 179396–179408, 2019.
- [17] H. Khan, A. Srivastav, and A. K. Mishra, "Use of classification algorithms in health care," in *Big Data Analytics and Intelligence: A Perspective for Health Care* Emerald Publishing Limited, West Yorkshire, UK, 2020.
- [18] A. Dasgupta, D. Rahman, and A. Routray, "A smartphone-based drowsiness detection and warning system for automotive drivers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 11, pp. 4045–4054, 2019.
- [19] A. K. Mishra and A. K. Dash, "Development impact bonds in developing countries: an emerging innovation for achieving social outcomes," *Journal of Social and Economic Development*, vol. 21, pp. 1–27, 2022.
- [20] M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, M. Ilyas, and A. Mahmood, "A survey on state-of-the-art drowsiness detection techniques," *IEEE Access*, vol. 7, pp. 61904–61919, 2019.
- [21] S. Kaplan, M. A. Guvensan, A. G. Yavuz, and Y. Karalurt, "Driver behavior analysis for safe driving: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3017–3032, 2015.
- [22] M. Singh and A. S. Arora, "A novel face liveness detection algorithm with multiple liveness indicators," *Wireless Personal Communications*, vol. 100, no. 4, pp. 1677–1687, 2018.
- [23] S. Mohanty, S. V. Hegde, S. Prasad, and J. Manikandan, "Design of real-time drowsiness detection system using dlib," in *Proceedings of the 2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, pp. 1–4, IEEE, Bangalore, India, November, 2019.
- [24] A. Jain, R. Kumar Dwivedi, H. Alshazly, A. Kumar, S. Bourouis, and M. Kaur, "Design and simulation of ring network-on-chip for different configured nodes," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 4085–4100, 2022.
- [25] S. Zhou, C. Chen, G. Han, and X. Hou, "Deep convolutional neural network with dilated convolution using small size dataset," in *Proceedings of the 2019 Chinese Control Conference (CCC)*, pp. 8568–8572, IEEE, Guangzhou, China, July, 2019.
- [26] K. Crowston, "Amazon mechanical turk: a research tool for organizations and information systems scholars," in *Shaping the Future of Ict Research. Methods and Approaches* Springer, Berlin, Germany, 2012.
- [27] A. Jain and A. Kumar, "Desmogging of still smoggy images using a novel channel prior," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1, pp. 1161–1177, 2021.
- [28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <https://arxiv.org/abs/1409.1556>.
- [29] S. Kumar, A. Jain, A. Kumar Agarwal, S. Rani, and A. Ghimire, "Object-based image retrieval using the U-Net-Based neural network," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 4395646, 14 pages, 2021.
- [30] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.
- [31] C. Wang, Y. Xiao, X. Gao, L. Li, and J. Wang, "A framework for behavioral biometric authentication using deep metric learning on mobile devices," *IEEE Transactions on Mobile Computing*, vol. 22, no. 1, pp. 19–36, 2023.
- [32] A. M. R. Tolba, "Trust-based distributed authentication method for collision attack avoidance in VANETs," *IEEE Access*, vol. 6, pp. 62747–62755, 2018.
- [33] A. Bhola and S. Singh, "Visualisation and modelling of high-dimensional cancerous gene expression dataset," *Journal of Information and Knowledge Management*, vol. 18, no. 1, Article ID 1950001, 2019.
- [34] N. Alioua, A. Amine, M. Rziza, and D. Aboutajdine, "Driver's fatigue and drowsiness detection to reduce traffic accidents on road," in *Proceedings of the International Conference on Computer Analysis of Images and Patterns*, pp. 397–404, Springer, Berlin, Heidelberg, August, 2011.
- [35] R. Jabbar, K. Al-Khalifa, M. Kharbeche, W. Alhajyaseen, M. Jafari, and S. Jiang, "Real-time driver drowsiness detection for android application using deep neural networks techniques," *Procedia Computer Science*, vol. 130, pp. 400–407, 2018.
- [36] A. Bhola and S. Singh, "Gene selection using high dimensional gene expression data: an appraisal," *Current Bioinformatics*, vol. 13, no. 3, pp. 225–233, 2018.