



**ISSN: 2454-9940**



**INTERNATIONAL JOURNAL OF APPLIED  
SCIENCE ENGINEERING AND MANAGEMENT**

**E-Mail :**  
**editor.ijasem@gmail.com**  
**editor@ijasem.org**

**[www.ijasem.org](http://www.ijasem.org)**

# Drive Alert AI: Facial-Based Driver Fatigue Detection System

<sup>1</sup>Polepally Ramya, <sup>2</sup>Ch Mahesh

<sup>1</sup>M.Tech Scholar, Dept. of CSE, Malla Reddy Technical Campus, Malla Reddy Vishwavidyapeeth, Maisammaguda, Hyderabad, Telangana 500100, India, [polepallyramya3@gmail.com](mailto:polepallyramya3@gmail.com)

<sup>2</sup>Assistant Professor, Dept. of CSE, Malla Reddy Technical Campus, Malla Reddy Vishwavidyapeeth, Maisammaguda, Hyderabad, Telangana 500100, India, [chekurimahesh364@gmail.com](mailto:chekurimahesh364@gmail.com)

---

## Abstract:

There are an increasing number of occupations that need undivided attention these days. In order to respond swiftly to unexpected circumstances, drivers must maintain a tight focus on the road. Fatigued drivers are a leading cause of many accidents on the road. The frequency of incidents caused by driver weariness may be drastically reduced if technology could identify and alert drivers to their poor psychophysical state. Nevertheless, there are other challenges associated with the accurate and rapid identification of driver tiredness symptoms that arise throughout the development of such systems. The use of vision-based approaches is one of the technological alternatives for implementing systems that detect driver sleepiness. The many systems that are presently in use to detect driver sleepiness are presented in this article. In this case, we are assessing the driver's visual system in order to identify his tiredness.

---

**Keywords:** Artificial neural networks; Deep feature extraction; Driver drowsiness dataset; Driver fatigue detection

---

## Introduction:

We will monitor the driver's every action in future autonomous cars to address the issue of driver weariness and yawning, which are now the top causes of catastrophic accidents. By doing so, we can detect when the driver's eyes are on the road. Using a Kinect sensor—which can recognize three-dimensional images—this system produces accurate results. More than that, we can identify the driver's exact actions using this technology since it can follow their head movements.

Driver drowsiness detection is a kind of car safety equipment that can detect whether a driver is becoming sleepy. Research suggests that drowsiness contributes to 20% of all traffic accidents and 50% on certain roads. There is an alarmingly high rate of automobile accidents involving drivers who are fatigued. According to current statistics, accidents resulting from drivers being too sleepy cause more than 1,200 deaths and 76,000 injuries annually. Finding a way to detect or stop drivers from being drowsy is one of the major challenges in accident prevention systems. Because being drowsy while

driving is dangerous, we need to figure out how to prevent it. When drivers aren't paying attention, it might be because they're too tired or distracted to pay full attention to the road. To be distracted while driving is to allow one's attention to wander to anything other than the road ahead. Unlike driver distraction, which starts with an isolated occurrence, driver drowsiness is a cumulative effect of the driver's increasing inattention to the road and traffic conditions. Driving when distracted or drowsy increases the risk of accidents, decreased reaction times, and poor performance overall. offers a comprehensive system diagram. The technology combines the front-facing camera's video gathering capabilities together with real-time analysis of an incoming video stream to ascertain the driver's level of tiredness. By sensing the eyes, an alert will be delivered if the tiredness is assessed. Here, the exact driver is located using the human skeletal identification method. An individual's joints may be used as a kind of identification. such that the many parts that move throughout the body may be easily identified. Using this technology, we can positively

identify the human skeleton. By taking a succession of images with a standard camera in real-time, computers and mathematical equations may determine whether the driver is indeed asleep or just yawning. We can record the expressions on people's faces, including their eyes and lips, by using image processing techniques. Then, a threshold value may be generated using the eye aspect ratio (EAR). whether the driver is asleep, we may use this threshold number to determine whether they need to be constantly beeped at to avoid deadly accidents, injuries, or collisions with immobile objects. Afterwards, a wake-up call is issued to the appropriate team or the driver's loved ones by automated email.

Noise reduction and other image processing techniques are usually the first steps in object detection in photos. The next step is (low-level) feature extraction, which involves finding lines, areas, and sometimes even locations with certain textures. The clever part is to handle clusters of these shapes like distinct objects, as moving cars on a road, boxes on a conveyor belt, or cancer cells under a microscope. This is probably a problem with AI, as the way an item looks might alter significantly depending on factors like lighting and perspective. The difficulty increases when trying to differentiate between actual object features and those that are just background details, shadows, etc. Although the human visual system often does these tasks instinctively, computers need very skilled programmers and enormous amounts of processing power to reach human levels of proficiency. Changing information via visual representation in a variety of possible ways. Patterns seen on screens (such as those in movies, TV shows, slides, or pictures) often lead people to believe that visuals are two-dimensional arrays of brightness values. Digital or optical image processing may be done on a computer. The first stage of digital processing is to convert an image to a numerical format that a computer can understand and use. The brightness value at a certain location is indicated by the various numerical components that comprise a picture, known as pixels. A typical digital image may have dimensions of  $512 \times 512$ , or around 250,000 pixels, even if larger and larger images are becoming more common. When working with digital images on a computer, there are essentially three main options. An output pixel value for a point operation is determined by a single pixel value in the input image. Many neighboring pixels in the input image decide

the value of each output picture pixel in local processes. In a global operation, each pixel in the input image is converted into a pixel value in the output image. These methods, used singly or in combination, enhance, restore, or compress the image. Image enhancement may mean many things to different people. One meaning is making the image seem better. Another is making the information in the image more visible. Eliminating background noise is one instance. To eliminate noise in an image, median filtering may be applied using a  $3 \times 3$  pixel window. What this means is that the values of the eight pixels immediately around each noisy pixel are also recorded. In order to determine the median number to utilize as the pixel value in the new image, sort these nine integers by size. In order to construct the filtered image, the  $3 \times 3$  window is moved pixel by pixel over the noisy image.

## Literature Survey

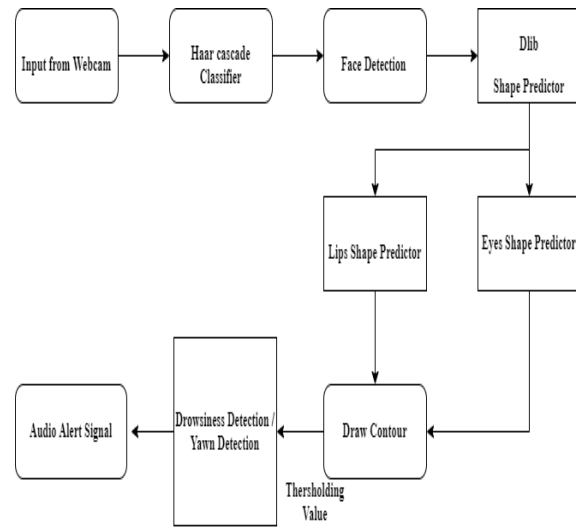
An ideal device would be one that could monitor a person's attention levels in real-time, allowing us to prevent accidents caused by microsleep and broad lethargy. This article explains how to build a video-based system that uses computer vision to assess face metrics in order to identify behavioral microsleep and visual markers of tiredness without invasiveness or light sensitivity. The developers of the device used a camera mounted on a remote and lit with near-infrared light to record the video. In order to determine the eye closure ratio, the next step is to employ computer vision algorithms to gradually identify the positions of the eyes, face, and eyelids. Nine people with diverse facial characteristics and varying eye closure and gaze ratios were photographed frontally in both bright and dark settings to test the procedure. The accuracy in discriminating between fully closed, partly closed, and open eyes was shown in preliminary trials to be promising.

In this paper, we provide a machine learning approach to visual object identification that can swiftly analyze images while maintaining high detection rates. What makes this research unique are three primary contributions. To begin, we have revolutionized picture representation with the introduction of the "integral image" that substantially streamlines the calculation of detector characteristics. The second is a learning method based on AdaBoost that generates highly effective classifiers by distilling

a large set of visual features to a manageable subset. Thirdly, we present a mechanism to "cascade" more complex classifiers, which allows for the quick removal of image backgrounds and the redistribution of computing resources to possible object-like regions. Unlike previous approaches, the cascade provides statistical guarantees that the excluded regions are unlikely to include the target object; it may be seen as a mechanism for object-specific focus-of-attention. The system performs similarly to previous top systems when evaluated on face detection. Picture differencing and skin color recognition detectors slow things down, letting real-time apps run at 15 fps.

Drivers' fatigue and yawning is said to be a leading cause of accidents on roads and highways. Consequently, a system that can identify when a motorist is becoming too sleepy to drive safely should be developed. In an effort to lessen the occurrence and impact of car accidents, this research describes the design and implementation of a system that can identify and document the level of drowsiness and yawning in an average driver. A Celoxica RC203E FPGA-based processor communicates with an infrared camera placed on the vehicle to process live video feeds. In particular, the technique takes use of the "bright pupil" effect, which occurs when 850 nm radiation reaches the retina and causes it to reflect 90% of that light. During video acquisition, a processing chain is executed to detect the driver's eyes and compute a PERCLOS (Percentage of Eye Closure) function; this allows us to ascertain the driver's yawn and degree of tiredness. The results show that the system can process a PAL video stream at 60 frames per second in real time, with a whole 720\*576 frame requiring just 16.7 milliseconds. The proposed device proved its worth by tracking the eyes of a real-life driver and detecting signs of fatigue.

## Methodology



System proposed

## Face detection

Face detection has been the subject of several proposed approaches due to its popularity as a research topic. These methods generally address face detection with the same level of rigor as any other binary classification issue. The goal, therefore, is to identify facial features in a given image fragment. A classifier trained on example photographs may then decide whether the converted features resembling a human face represent the given location. A window-sliding technique is also often employed since faces may appear in various sizes and locations. Regardless of the size or position of the picture, the objective is to train the classifier to distinguish between faces and non-faces. An input photo is used to produce consistently sized and intensity-normalized face regions using this approach. These features of one's look are derived from the recognized face component, which signals changes to the face such as furrows and wrinkles (skin texture). This system paradigm makes use of an executable file, which is a dynamic link library (.dll), to extract the face area. To identify faces, the process relies on an adaptive boosting method and traits similar to haar.

## Pre-processing:

In order to prepare the raw input video for processing by the subsequent phases, this step comprises a set of basic image processing operations. Early processing steps use basic spatial or temporal smoothing to

mitigate precipitation-related noise. An essential component of pre-processing is the data format used by the background removal method. The majority of the methods deal with the pixel-wise luminance intensity scalar. Nevertheless, there has been a recent uptick in the use of color images in background removal research, whether in RGB or HSV color space [8, 15]. In these articles, the authors state that when it comes to detecting things in low-contrast environments and reducing the effect of shadows created by moving objects, color is superior than brightness. Incorporating edge and motion information is sometimes accomplished using pixel-based picture attributes like spatial and temporal derivatives, in addition to color. As an example, the Kalman filter may be used for background tracking by combining intensity data and spatial derivatives into a single state space. A constant velocity background model for identifying fast cars is created by Plessy et al. using both spatial and temporal derivatives. One major issue with using derived features or color in background modeling is the additional 130 CS & IT complexity required for estimating model parameters. Most backdrop modeling solutions use a separate model for each pixel, which may lead to a noticeable increase in complexity.

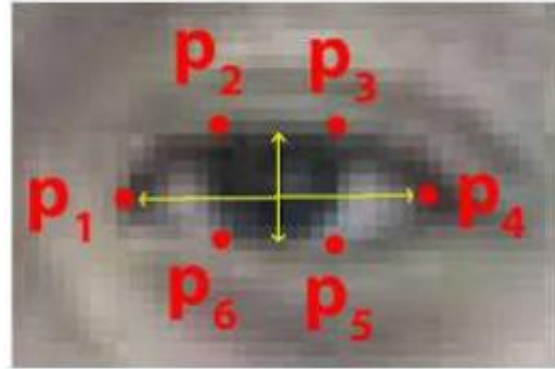
### Fatigue Detection:

One piece of automotive safety equipment that may assist avoid accidents caused by fatigued drivers is yawn and driver drowsiness detection. Depending on the study, sleepiness might be the cause of as many as half of all traffic accidents (20% overall and 50% on certain roadways). Some of the existing systems are able to learn driving habits and identify signs of driver fatigue.

### Eye Detection and drowsiness Detection:

Eyes seen from the face undergo further processing to ascertain whether the driver is drowsy. In this section, we will examine the eyes in particular. An algorithm that can detect eye blinks in camera video sequences in real time is used by this proposed system. Even when subjected to camera shake, illumination changes, and human emotions, modern landmark detectors hold their own. An estimate of the eye-opening level may be provided by the project's landmark detection, which is sufficiently precise. The proposed technique evaluates the landmark positions and derives a number known as the eye aspect ratio

(EAR) to characterize the opening of the eye in every frame.



This method detects the opening and shutting of the eye by making use of several landmarks. These landmark detectors are able to pick up on the majority of identifying features in a picture of a human face. A quick closure and reopening of the eye is called a blink. The blinking pattern varies somewhat from person to person. The length of a blink, the amount of pressure on the eye, and the rate of shutting and opening the eye are all variables that vary between patterns. Blinking of the eyes takes from 100 to 400 milliseconds. We estimate the eye opening state using the eye aspect ratio (EAR), which is derived from the landmarks identified in the picture. The eye landmarks are identified for each frame of the video. We calculate the eye's aspect ratio, which is the ratio of its height to its breadth. The ocular landmarks are shown in figure 2 as P1, P2,..., P6.

The EAR is almost nil when one closes their eye and is rather steady when one opens their eye. Because blinking occurs simultaneously in both eyes, we take the EAR of both and average it. A sleepy driver is one whose EAR value remains below the limit for more than two or three seconds after its acquisition. To fix the driver's erratic conduct, the system's buzzer takes action.

### Drowsiness detection:

Many methods have been suggested for face detection as it is a hot study subject. The majority of these approaches treat face detection like any other binary classification problem. In other words, the objective is to determine whether a particular picture segment represents a face. To do this, the provided area is first transformed into features, and then a

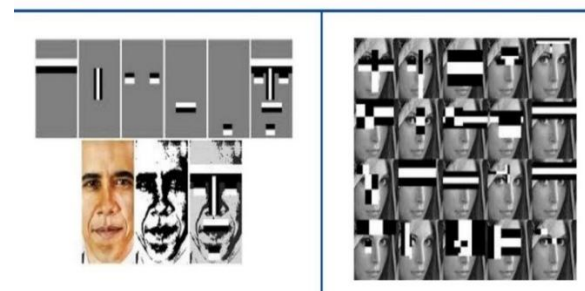
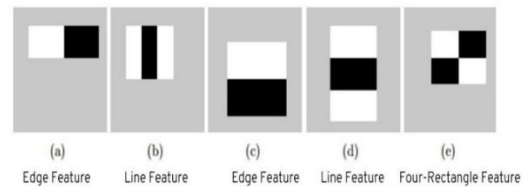
classifier trained on sample photos determines whether or not these characteristics depict a human face. Because faces might show up in different sizes and places, a window-sliding approach is typically used as well. The goal is to train the classifier to identify faces and non-faces in images across all sizes and locations. The method involves obtaining uniformly sized and intensity-normalized facial areas from an input picture.

The identified facial component indicates changes of the face, such as furrows and wrinkles (skin texture), and these appearance elements are taken from that. • To extract the face area in this system model, an executable (.dll-dynamic link library) file is used. The adaptive boosting approach and haar-like characteristics form the basis of the face detection procedure. Measurements of forward head tilt angle, eye aspect ratio (to identify eye closing), and mouth aspect ratio (to identify yawning) are all part of the computer vision algorithm's trifold method to tiredness detection.

### Haar Cascade Algorithm:

Digital picture characteristics that are employed for object detection are known as Haar-like features. When it comes to detecting features in images, Haar features are not dissimilar to convolution kernels. At a given spot in a detection, a Haar-like feature takes into account neighboring rectangular areas. The white area in a Haar feature is substituted with -1 and the black area with +1. When first testing these characteristics in a picture, the Viola Jones method use a 24x24 window size. (For instance, Type 1) We compute almost 160,000 features in this 24x24 frame if we account for all potential Haar feature properties, such as location, scale, and type. Haar Algorithms Classifiers begin extracting facial characteristics as soon as they get training data. The presence or absence of a certain feature on an image may be verified using convolution kernels, which are known as Haar features. With a basic window size of  $24 \times 24$ , the Integral Image Haar Cascade classifier generates around 180,000 features. Finding the average pixel difference for each face feature is challenging. The sum of all the pixels, taking into account the values of every rectangle's four corners, yields an integral picture. Adaboost, since the integral image step generates 180,000 features. But face detection doesn't need all of those attributes. Subsequently, Adaboost employs an ML algorithm to choose the optimal collection of face attributes. As a result, the number of features drops from over

180,000 to around 6,000. Deterioration The characteristics are grouped into distinct classifier stages and applied one by one since it is impossible to retain all 6,000 extractions on a single window. In the event of a main stage failure, the window is eliminated. After the time limit has elapsed, go to the second stage by applying the face traits. To identify a facial area, look for the window that goes through each stage.



### Eye Aspect Ratio

Eye Aspect Ratio & Alert Alarm—This method makes use of many landmarks to identify when the eye opens and closes. An picture of a human face may have most of its distinguishing features picked up by this landmark detector. A quick closure and reopening of the eye is called a blink. The blinking pattern varies somewhat from person to person. The length of a blink, the amount of pressure on the eye, and the rate of shutting and opening the eye are all variables that vary between patterns. Blinking of the eyes takes from 100 to 400 milliseconds. We estimate the eye opening state using the eye aspect ratio (EAR), which is derived from the landmarks identified in the picture.

### Measuring EAR

Based on the picture's detected features, we calculate the eye aspect ratio (EAR), which we then use to

estimate the eye opening state. For every frame in the video, the ocular landmarks are named. The aspect ratio of the eye is determined by dividing its height by its width. See P1, P2,..., P6 in Figure 2 for the ocular landmarks. You may use this to get the ratio of the distances between the horizontal and vertical eye landmarks. As a first step in creating a blink detector, facial landmark identification is executed to find the eyes in a certain frame of a video stream. Using the two sets of facial landmarks, we can determine the eye aspect ratio, which provides a single value that connects the vertical and horizontal distances between the two sets of landmarks.

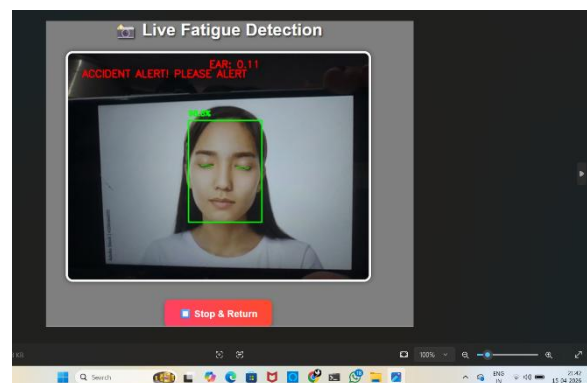
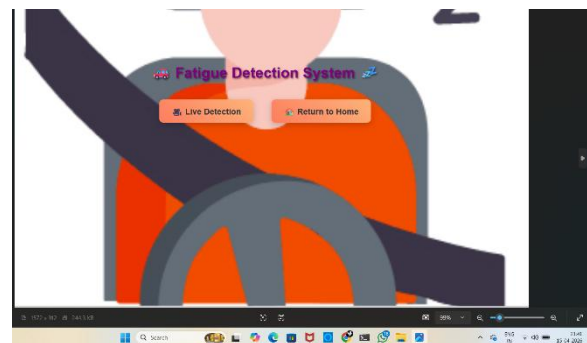
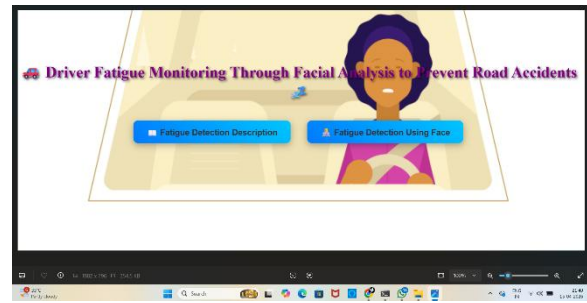
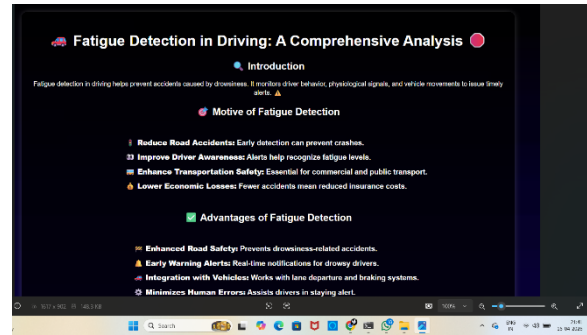
### Mouth detection

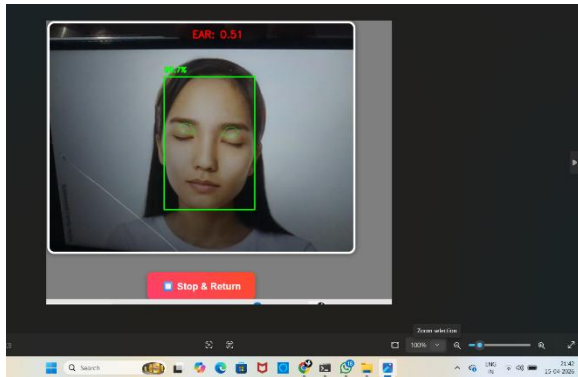
Locating the mouth and lips is the next stage in the suggested techniques. One way to find the area that contains the mouth, often called the mouth window, is to use intensity or color data. The following regions may be approximately inferred from the position of the face using previous knowledge: 1) the bottom half of the face area is where the mouth is; 2) the corners of the lips are at a particular distance from the boundary of the face region; and 3) the lower lip does not touch the chin. By reducing the searching area and avoiding upsetting the backdrop pixels with similar hue, the estimated region is seeking space for mouth region detection.

### Shape Predictor Algorithm:

Python, OpenCV, and dlib for face landmark detection. The "rect\_to\_bb" utility function (which means "rectangle to bounding box") is the first one: A bounding box rectangle generated by a dlib detector (such as the face detector) is presumed to be the only input that this function takes, rect. With the detection's (x, y)-coordinates included in the rectangle object. Nevertheless, for the sake of convenience, the rect\_to\_bb function converts this rect\_object into a 4-tuple of coordinates as that is how bounding boxes are often conceptualized in OpenCV. To get the 68 (x, y)-coordinates of the face landmark areas, the dlib face landmark detector will provide you a shape object. We can transform this object into a NumPy array by using the shape\_to\_np method. In order to identify faces in photos, we made use of these two auxiliary functions.

## Results





SVM		Predicted	
		Drowsy	Non Drowsy
Actual	Drowsy	21,826	522
	Non Drowsy	1010	18,435

RF		Predicted	
		Drowsy	Non Drowsy
Actual	Drowsy	22,322	26
	Non Drowsy	66	19,379

(a) (b)

ANN		Predicted	
		Drowsy	Non Drowsy
Actual	Drowsy	22,341	7
	Non Drowsy	4	19,441

Confusion matrix results for each model (a)

SVM; (b) RF; (c) ANN.

A crucial component of intelligent vehicles is the advancement of technologies for perceiving and modeling their environments. The current status of technology for environment perception and modeling is summarized in this project. The next section will include scene understanding, common algorithms and approaches for modeling tiredness and yawn detecting behavior. Environment perception and modeling technology codes and publicly available datasets are also detailed. Complex outside surroundings need effective ways for real-time perception, which is a current problem for environment perception and modeling technologies. Intelligent cars continue to face substantial obstacles in dealing with dynamic environmental factors, complicated backdrops (particularly those including occluding objects), and unpredictable lighting and

weather. In addition, knowing how to identify a road in an off-road setting is crucial. Building the whole system is the next phase, because the technological stage of environment perception and modeling connects the tasks of localization and map construction, route planning and decision-making, and motion control.

### CONCLUSION

The suggested method uses the driver's eye closure and yawning as indicators to identify fatigue and yawning. The system is evaluated across many domains using a small set of test instances. It has a 95% success rate when detecting faces and the necessary facial characteristics. Under improved lighting conditions, the accuracy of detection is shown to rise. Fatigue and yawning are quickly identified in every instance of positive feature recognition. Eye closure detection works flawlessly even when obstructed by glasses or other eyewear. To overcome the challenges caused by poor illumination, this might be expanded. When external obstacles lead to disappointments in location forms, our approach may be used to detect languor in exceptional cases as a future scope.

### Reference

- [1]. Zhang, Y., et al. (2021). Driver Fatigue Detection Using Convolutional Neural Networks.
- [2]. Singh, A., & Kumar, R. (2022). Driver Drowsiness Detection Using Image Processing and Machine Learning Techniques.
- [3]. Patel, S., et al. (2021). Real-Time Driver Monitoring System Using OpenCV and Python.
- [4]. Li, X., et al. (2023). Hybrid CNN-LSTM Model for Driver Fatigue Detection.
- [5]. Sharma, P., & Verma, S. (2022). Driver Fatigue Detection Using Facial Landmark Analysis.
- [6]. Kim, H., et al. (2021). Infrared Camera-Based Driver Drowsiness Detection System.
- [7]. Rao, K., et al. (2023). IoT-Based Driver Fatigue Monitoring System.
- [8]. Ahmed, M., et al. (2022). Deep Neural Network-Based Real-Time Fatigue Detection.



- [9]. Gupta, R., & Singh, T. (2021). Comparative Study of Machine Learning Algorithms for Fatigue Detection.
- [10]. Chen, L., et al. (2023). Multi-Modal Driver Fatigue Detection System Using Facial and Vehicle Data.
- [11]. Mehta, D., et al. (2022). Eye Aspect Ratio-Based Driver Drowsiness Detection.
- [12]. Wang, Z., et al. (2021). Real-Time Yawning and Eye Closure Detection Using Deep Learning.
- [13]. Reddy, V., et al. (2023). Low-Cost Driver Fatigue Detection System Using Raspberry Pi.
- [14]. Ali, S., & Hassan, M. (2022). Facial Expression-Based Fatigue Detection Using Machine Learning.
- [15]. Park, J., et al. (2021). Driver Fatigue Detection Using Head Pose and Eye Movement Analysis.
- [16]. Nair, R., et al. (2023). Cloud-Based Driver Fatigue Monitoring System.
- [17]. Joshi, M., & Patil, A. (2022). Real-Time Driver Alert System Using Buzzer and Vibration.
- [18]. Brown, T., et al. (2021). Artificial Intelligence for Driver Safety and Fatigue Detection.
- [19]. Khan, A., et al. (2023). Deep Learning-Based Driver Fatigue Detection Under Various Conditions.
- [20]. Das, S., & Roy, P. (2022). Hybrid Image Processing and Machine Learning Approach for Fatigue Detection