International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing Driver Drowsiness Detection: A Fusion of Facial Landmarks and Modified YOLOv5 Architecture

Mohan Arava[1], Divya Meena Sundaram[*2]

Submitted: 06/11/2023 **Revised**: 25/12/2023 **Accepted**: 03/01/2024

Abstract: The driver drowsiness detection system aims to enhance road safety by preventing accidents caused by driver fatigue. Despite progress in drowsiness detection using various approaches, existing methods often lack accuracy in capturing subtle aspects like facial expressions, eye movement patterns, micro head gestures, changes in blink frequency, and variations in steering control of driver behavior can often provide crucial insights into their level of alertness and potential drowsiness while operating a vehicle. This work introduces a novel framework that combines facial landmarks and the Yolov5 architecture to enhance drowsiness detection. By extracting relevant facial features using a modified Yolov5 architecture, the system gains a comprehensive understanding of the driver's state during operation, enabling it to detect even subtle indicators of drowsiness. The framework's integration with facial landmarks allows for the observation of minute changes in facial expressions, providing valuable insights into the driver's level of alertness. The framework was evaluated on the benchmark UTA and custom dataset, where the proposed model achieved an accuracy of 95.5% and 96.4% respectively. In comparison with the state-of-the-art techniques, the proposed system achieves an improvement of 3.2%.

Keywords: *Drowsiness Detection, You Only Look Once (YOLO) v5, Face Detection, Facial Landmark, Data Augmentation.*

1. Introduction

Drowsy driving plays a vital role in road accidents.. When a driver becomes drowsy or falls asleep at the wheel, they may be unable to react quickly enough to unexpected situations or make safe decisions. This can lead to accidents that cause property damage, injuries, or even fatalities [1]. There are several reasons why driver drowsiness can occur, including lack of sleep, sleep disorders, medications, and alcohol consumption [2]. Drivers who are fatigued or sleep-deprived are particularly at risk of drowsiness. It's essential to recognize the signs of driver drowsiness, which may include yawning, blinking frequently, drifting out of lanes, missing exits or traffic signs, and difficulty focusing or keeping their eyes open [3].

As per the US National Highway Traffic Safety Administration (NHTSA)[4], drowsy driving is believed to contribute to approximately 7% of all accidents and 16% of fatal crashes in the United States.

Driver drowsiness is a significant problem in India. According to a study conducted by the Indian Ministry of Road Transport and Highways[5], driver fatigue was identified as a contributory factor in 4.6% of all road accidents in the country in 2021. Another study conducted by the Save LIFE Foundation, a non-profit organization, found that driver fatigue was a contributing factor in 37.6% of road accidents on highways in India. Most drivers and pedestrians endure catastrophic injuries as a result of drowsy driving; however, drowsiness detection technologies are still lacking in the Advanced Driver Assistance System (ADAS)[6]. Forty percent of road fatalities and injuries are caused by drivers who are too sleepy to pay attention, according to the Central Road Research Institute (CRRI)[7]. About one hundred thousand incidents occur annually due to sleepy driving, resulting in two thousand fatalities and seventy thousand injuries, as reported by the NHTSA[8]. Overall, these statistics motivate us to develop an effective driver drowsiness detection system[9].

Driver drowsiness detection systems also face significant challenges, including the need to achieve high accuracy and reliability in detecting drowsiness amidst variations in driver behavior and facial expressions [10]. System performance can be affected by environmental factors like lighting and weather, underscoring the importance of effectively handling false positives and false negatives to ensure reliability [11].The above challenges also stress the need for an effective driver drowsiness detection system. Existing drowsiness detection methods, which often utilize facial analysis and vehicle dynamics through computer vision, have limitations in capturing subtle facets of driver behavior, resulting in reduced accuracy and reliability[12]. The challenge lies in developing a novel framework that effectively detects driver drowsiness, accommodating the variability in driver behavior, and achieving real-time processing for timely intervention. Additionally, the system must handle diverse environmental conditions, and

[1,2] School of Computer Science and Engineering, VIT-AP University, Amaravati, India
* *Corresponding Author Email: divyameena.s@vitap.ac.in*

minimize false positives and negatives. This work aims to address these challenges and provide an efficient and reliable drowsiness detection system.

Understanding how to identify drivers who are too drowsy to safely operate a vehicle has become a major focus of study in recent years[13]. Technology, such as fatigue detection systems and driver assistance systems, can also play an important role in preventing road accidents due to driver drowsiness. Some vehicles feature drowsiness alert systems incorporating lane departure warnings, collision alerts, and automatic braking[14]. However, a significant limitation is the reliance on physical cues, overlooking cognitive fatigue, potentially leading to delayed interventions and compromised road safety. Wearable devices offer drowsiness detection and alerts for drivers, prompting breaks or stops[15]. Yet, a key limitation is their external placement, inhibiting direct monitoring of vital signs and potentially missing early signs of fatigue or impairment. Drowsiness detection through facial and behavioral analysis holds significant promise due to several compelling reasons. Firstly, facial analysis can capture subtle changes in expressions and eye movements, directly reflecting cognitive states like drowsiness [16]. Secondly, behavioral analysis, including steering patterns and head movements, offers real-time insights into driver engagement. Thirdly, the non-intrusive nature of these methods ensures driver comfort and facilitates widespread adoption[17]. Lastly, advancements in computer vision and machine learning empower accurate and timely alerts, contributing to enhanced road safety and accident prevention[18].

The main contribution of this work is the introduction of a novel framework that addresses the challenge of driver drowsiness detection using facial analysis. Through the integration of facial landmarks and the Yolov5 architecture, the suggested method attains a thorough comprehension of the driver's state during operation, thereby improving the precision and dependability of drowsiness detection. The integration of facial landmarks enables the system to capture subtle indicators of drowsiness by observing minute changes in facial expressions, providing valuable insights into the driver's level of alertness. The framework's robustness is demonstrated through impressive experimental results, achieving an accuracy of 95.5% and 96.4% on both the UTA and custom datasets respectively.

The rest of this paper is arranged as follows: Section II discusses the related work. The proposed methodology is described in Section III, a description of the Experimental framework in Section IV, Section V focuses on the Experimental analysis and results of the YOLOv5. The work mentioned above is concluded in Section VI.

2. Related Work

In this section, we start by discussing the various techniques available for driver drowsiness detection. We also discuss the literature on physiological, vehicular, and behavioral-based techniques. Toward the end, we list the research gaps in the driver drowsiness detection system

2.1Classification of drowsiness detection techniques

Drowsiness detection techniques may be broken down into three different subsets, including physiological measurements, vehicle-based measures, and facial analysis [19,20]. First of all, as a driver gets weary, their physical circumstances change, and this shows up in their heart rate, body temperature, pulse rate, and other physiological markers [21]. Conventional techniques for determining a person's health status often involve the use of electrocardiograms [22], electroencephalograms [23], electromyograms [24], and electrooculograms [25]. Driver comfort while wearing sensors is a major concern when using physiological approaches, which is a major drawback [26].

Secondly, measurements used within the car can pick up signs of sleepiness in the driver by monitoring things like the driver's hand placement on the wheel, instances of erratic braking, and fluctuations in speed [27]. To detect drowsiness while driving, sensors are mounted in various parts of the car to track the driver's actions and track driving patterns. The primary problem with vehicular approaches is that the vehicle's behavior is influenced by factors such as bad weather and road conditions, excessive medicine, lack of sleep, etc. To overcome the problems with the first and second techniques, machine learning and computer vision algorithms (CV) are used to analyze a person's behavior or face to determine if they are sleepy[28].

2.2Literature on physiological and vehicle-based techniques

Zuojin Li et al[29]. looked at the strategies that may be implemented in a vehicle. With the use of in-car sensors, they were able to collect information on the driver's state of fatigue, including yaw angles and steering wheel angles. The entropy is derived from a series of data used to analyze the characteristics acquired from steering wheel angles and yaw angles [30]. Using a classifier trained with input from Back-propagation Neural Networks and features approximating the driver's entropy, the study was able to accurately predict the driver's drowsiness with an accuracy of 87.21 percent. The driver is classified by the algorithm to be either sleepy or very drowsy.

2.3Literature on behavioral-based techniques

In recent years, behavioral techniques have been established to address the shortcomings of physiological and vehicle-based approaches [31]. To identify drowsiness, many new behavioral techniques have emerged in recent years. Different authors have employed different algorithms to identify facial expressions and levels of tiredness[32]. Drowsiness detection may be achieved with the help of neural network techniques such as CNN [33], ANN, Naive Bayes classifier, and GAN's.

Due to their focus on the driver's behavior instead of vehicle behavior, behavioral approaches should be more accurate than vehicular methods. Physiological methods, yield very precise outcomes but are rarely employed because of their complexity. Sherif Said et al. introduced driver fatigue of the driver drowsiness system using the Viola-Jones to identify areas of the eye and face, which represents a significant improvement over previous behavioral approaches [34]. The driver's level of drowsiness is monitored by the device, which then triggers an alarm. The system was evaluated in both indoor and outdoor settings, and it achieved 82% and 72.8% success rates in those two conditions, respectively.

FengYou developed a procedure that prepares the driver for online use by first training it in isolation. Dlib's68 point facial landmarks are utilized to determine the eye aspect ratio once the CNN in Dlib has detected the face areas. Drowsiness detection systems typically consist of two phases: offline training using an SVM classifier and continuous online monitoring of the driver. The suggested technique has an accuracy of 94.8% based on a comparative study, however, the key drawback is the SVM classifier needs training from the end users. The data augmentation technique of Generative Adversarial Networks (GAN) and the prediction technique of Convolutional Neural Networks (CNN) were both presented in [35]. The authors conducted extensive research and found that by using GAN, they were able to create fresh data samples (real-world photos), and by incorporating CNNs, they were able to improve the model's accuracy.

The method proposed by R Tamanani et al. uses a pair of sequential systems: one system that uses the Haar cascade to detect the face data from a live video feed, and an output system that uses the CNN LeNet architecture[36]. With cross-validation on UTA-RLDD, the model obtained averages of 91.8% accuracy, 92.8% precision, 92% recall, and 92% F1-score. The system has a 98% success rate on a bespoke dataset, with an accuracy of 84% on training data, 88% on validation data, and 88% on testing data.

ResNet (Residual Network) proposed by Kaiming He et al has been used in drowsiness detection system. ResNet's depth and complexity in driver drowsiness detection can lead to overfitting and demand substantial labeled data [37]. Its computational intensity and potential limitations in capturing nuanced behavioral cues might necessitate careful integration with complementary techniques for real-time and accurate drowsiness detection.

Similarly, RNNs [39], LSTMs [40], and Bi-LSTMs [41] used for driver drowsiness detection may suffer from vanishing gradient issues and limited memory cell capacity, potentially hindering their ability to capture intricate patterns. Additionally, their training complexity and susceptibility to overfitting, along with challenges in real-time operation and model interpretability, underscore the need for careful consideration and potential combination with other techniques.

Out of all the neural networks, YOLO (You Only Look Once) has emerged as a prominent choice for driver drowsiness detection, often integrated with complementary techniques such as Haar Cascade and facial landmarks [42], This multi-pronged approach leverages YOLO's real-time object detection capabilities, enabling accurate identification of key facial features and contextual cues associated with drowsiness. The synergy between YOLO and these supplementary techniques enhances the overall robustness and efficacy of drowsiness detection systems, showcasing a dynamic fusion of modern computer vision and traditional machine learning methodologies to tackle the critical issue of driver safety. In summary, the following research gaps are identified:

i). ***Poor Performance in Low Lighting:*** Many established algorithms exhibit suboptimal performance in low-light environments, leading to an elevated rate of false positives. This deficiency inhibits their practical applicability in real-world contexts marked by variable lighting conditions.

ii). ***Inability to Detect Eyes with Eyeglasses***: Some models encounter difficulties in accurately detecting drivers' faces when eyeglasses are worn. This issue introduces the risk of false negatives, compromising the reliability of the entire system and raising concerns about its effectiveness in scenarios involving drivers who wear glasses.

iii). ***Challenges with Masked Drivers:*** Certain models reliant on mouth-related features struggle to detect drowsiness in drivers wearing masks, as exemplified by the COVID-19 pandemic. This limitation introduces uncertainties in drowsiness detection accuracy, particularly when mask usage is prevalent, underscoring the need for improved methodologies.

To address these pressing research gaps, this work introduces an innovative model that leverages the robust capabilities of YOLO v5 and facial landmarks for precise driver drowsiness detection. Notably, this approach presents a novel solution for countering the challenges posed by poor lighting conditions through the strategic implementation of histogram equalization and contrast equalization techniques. These enhancements enhance the clarity and quality of input data, allowing the model to achieve more accurate and reliable drowsiness detection outcomes. By seamlessly integrating cutting-edge technology and sophisticated image processing techniques, this proposed framework promises to bridge the existing gaps and significantly advance the field of driver drowsiness detection, ultimately contributing to enhanced road safety and accident prevention.

3. Proposed Methodology

In this section, we discuss the proposed methodology for the driver drowsiness system which has two phases namely, face extraction and drowsiness detection. First, we extract faces using facial landmarks, and then detect drowsiness using YOLOv5. We also use appropriate data augmentation techniques to increase the data and thereby improve the performance. The schematic representation of the proposed system is illustrated in Figure. 1.

3.1Data Augmentation:

In this work, we use data augmentation for the custom dataset alone and not on the benchmark UTA dataset. Appropriate data augmentation techniques are applied to the custom dataset to increase the size and thereby improve the model performance. Data augmentation is represented as A: X → X', where X' is the augmented set of images. The augmented dataset X' = $\{x_1', x_2', x_3' \ldots, x_n'\}$ is obtained through various transformations, such as rotation, scaling, flipping.

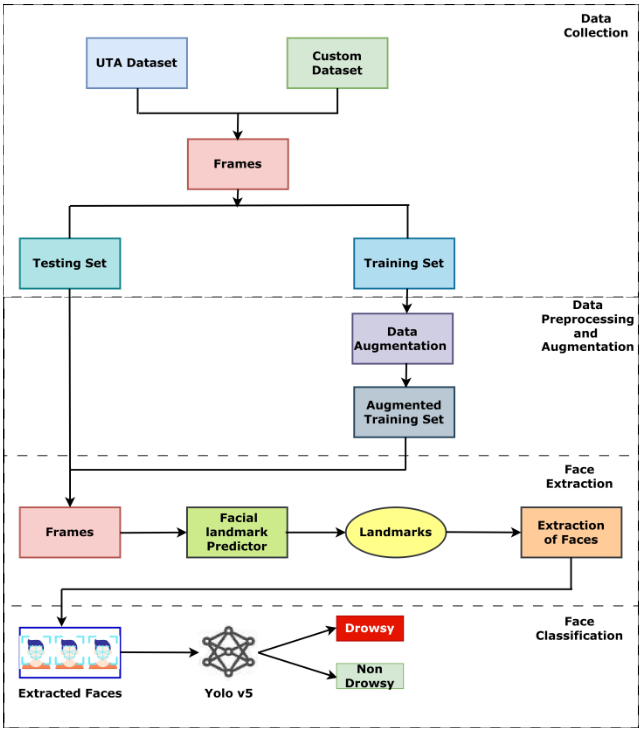


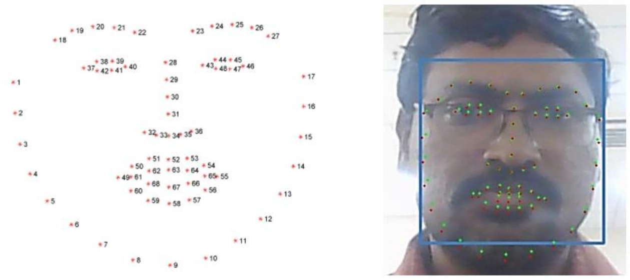
Fig. 1. Illustration of proposed drowsiness detection framework with data collection, data augmentation, facial landmark identification, and YOLOv5-based drowsiness detection modules.

3.2Face extraction using facial landmark

Face alignment is the process of identifying key facial landmarks such as the eyes, nose, and mouth, and then transforming the image so that these landmarks are in a standardized position relative to each other. This can be achieved using various computer vision techniques, such as facial landmark detection and affine transformations. Facial landmark detection involves using algorithms to detect and locate key facial features, such as the corners of the eyes, the tip of the nose, and the corners of the mouth. These points are then used as reference points for the alignment process. Affine transformations are mathematical operations that can be used to rotate, scale, and translate an image. Once the facial landmarks have been identified, the image can be transformed so that these landmarks are in a standardized position relative to each other. For example, the eyes can be aligned horizontally, and the mouth can be centered vertically. The resulting aligned image can be used for a variety of applications, such as face recognition, facial expression analysis, and virtual try-on.

Given a set of images X = $\{x_1, x_{2,}x_3 \ldots, x_n\}$, where x_i represents the i_{th} image containing the driver's face, and the corresponding labels Y = $\{y_1, y_{2,}y_3 \ldots, y_n\}$, where y_i is the binary label for the image x_i indicating drowsy (1) or non-drowsy (0) state. For each image x_i, the custom face extraction model is applied to identify facial landmarks, represented as $L_i = \{l_1, l_2, l_3, \ldots, l_m\}$, where l_j represents the j_{th} facial landmark point. The output of the face extraction phase results in a set of facial landmarks L = $\{L_1, L_2, L_3, \ldots, L_n\}$, corresponding to each image x_i in the dataset. Facial landmarks are used here to detect faces. We incorporated two landmark predictors: (i) a 68-point landmark predictor (refer to Figure 2), and (ii) a 5-point landmark predictor.

The performance of face landmarking was assessed using the normalized root mean square error (NRMSE), which was normalized concerning the inter-ocular distance (IOD). The IOD is defined as the distance between the centers of the two eyes. By normalizing the errors in landmark localization with the IOD, the evaluation of performance becomes impartial to the size of the face or the zoom factor of the camera. The normalized distance δ is computed as the Euclidean distance $d(.,.)$ between the ground-truth landmark coordinates (x, y) and the predicted landmark coordinates$(\tilde{x}, \tilde{y})$, normalized by the IOD.



a) b)

Fig. 2. a) 68-point landmark predictors b) Face landmarks (Red dots represents ground truth landmarks, green dots represent predicted landmarks).

Equation 1 shows the formula for the normalized distance of each landmark, where the subscript k indicates one of the landmarks.

$$\delta_k = \frac{d\{(x_k, y_k), (\tilde{x}_k, \tilde{y}_k)\}}{IOD} \quad (1)$$

$$NRMSE_{local} = \sqrt{\frac{\sum_{k=1}^{n} \delta_k^2}{n}} \quad (2)$$

NRMSE of each image ($NRMSE_{local}$) is calculated by the formula shown in Eqn. 2, where n is the number of landmarks.

3.3Drowsiness detection using YOLOv5

The drowsiness detection phase involves using YOLOv5 architecture to detect drowsiness based on the facial landmarks extracted in the previous phase. Mathematically, we define the drowsiness detection model as; Let D be the set of all driver drowsiness instances in the dataset, X be the set of all input images containing the driver's face and Y be the set of corresponding binary labels indicating drowsy (1) or non-drowsy (0) states for each image in X. The input to the YOLOv5 model is the set of facial landmarks L = $\{L_1, L_2, L_3, ..., L_n\}$. The model processes the landmarks and performs binary classification to determine whether each driver is drowsy or not, resulting in a set of predicted drowsiness labels P = $\{p_1, p_2, p_3, ..., p_n\}$, where p_i the predicted drowsy (1) or non-drowsy (0) state for each driver based on their facial landmarks.

The modified YOLO network we developed is based on the YOLOv5 architecture, renowned for its rapid detection speed. Our modifications to the model focused on adapting it specifically for driver drowsiness detection. This involved altering both the backbone and the loss function to enhance its suitability for the task. The backbone of our network is darknet-53, which combines features from darknet-19 and ResNet With 53 convolution layers, darknet-53 is more robust and powerful than darknet-19, while maintaining greater efficiency compared to ResNet-101 or ResNet-152. It delivers comparable performance to ResNet-152 but achieves twice the speed.

To integrate multi-scale information effectively, we adopted a design inspired by feature pyramid networks (FPN) to merge low-level features with high-level features. This approach allows us to utilize information from different scales in the image, thereby enhancing the performance of multi-scale detection. As the features of small-scale targets become highly compressed after multiple dimensional reductions, subsequent layers may not capture sufficient information, resulting in decreased efficiency and accuracy. In order to address this challenge, we implemented changes to the network architecture of darknet-53 by augmenting the number of layers in the first two residual blocks. This adjustment facilitated the capture of more comprehensive small-scale facial features and led to a significant improvement in the performance of driver drowsiness detection.

In the training phase, YOLO employs a multi-part loss function encompassing regression loss, classification loss, and confidence loss, even when detecting no objects. In the original YOLO implementation, the sub-losses are weighted equally at a ratio of 1:1:1:1, which is suitable for multi-class object detection. However, in the context of driver drowsiness detection, which is a binary classification problem, we found it necessary to adjust the weights to better align with this specific task. Through empirical analysis, we revised the weights to 2:1:0.5:0.5, assigning greater importance to regression and confidence losses. This adjustment aimed to enhance the model's performance in detecting driver drowsiness accurately.

In conventional practice, the evaluation of optimizer performance relies on the intersection over union (IoU) metric, which measures the overlap between the predicted location and the corresponding ground truth. The mean squared error (MSE) function is used for regression loss. The prior research has identified a discrepancy between optimizing MSE and maximizing the IoU value, particularly when dealing with non-overlapping bounding boxes. To overcome this limitation, the authors introduced a new metric known as generalized IoU (GIoU), which demonstrates a strong correlation with optimizing the MSE function. Drawing inspiration from this study, we enhanced the regression loss by adding the original ln-norm error and a weighted GIoU loss. This modification resulted in a more effective optimization objective for our driver drowsiness detection model.

The modified YOLO model uses the darknet framework on an NVIDIA GeForce GTX 1080Ti GPU. The training process involved using a batch size of 64 samples and setting the input image size to 416×416 pixels. To optimize the model, stochastic gradient descent with momentum was employed, starting with an initial learning rate of 0.001, which decayed exponentially every 4,000 steps. To enhance the model's versatility and ability to generalize, various data augmentation techniques were applied during training. These techniques involved adjusting the saturation, brightness, and hue of the input images. The training process was conducted over 20,000 steps to ensure the model reached a desirable level of performance. Following training, the model was evaluated on different datasets to assess its overall performance and effectiveness in various real-world scenarios. The evaluation aimed to gauge the model's capabilities and measure its accuracy in detecting objects in images.

The proposed methodology mainly contains three steps and are outlined as follows:

Input: Image X

Output: C {Drowsy, Non drowsy}

Dataset: D {UTA U, Custom X}

1. Data Augmentation:

For each image X in the custom dataset:

Apply a set of transformations to X, creating an augmented image X'

Augmentation A is defined as, A: X → X'

Augmented dataset X' = $\{x_1', x_2', x_3' \ldots, x_n'\}$ with transformations (A) applied

Dataset $D^{`} = \{U \,\&\&\, X'\}$

2. Face Extraction using Facial Landmark:

For each image $D^{`}$ in the dataset:

Apply point based landmark predictors to identify facial landmarks $L = \{L_1, L_2, L_3, \ldots, L_n\}$

Extract faces F from Landmark L, $F = \sum_{i=1}^{n} L_i$

For the extracted faces, check the NRMSE,

$$NRMSE_{local} = \sqrt{\frac{\sum_{k=1}^{n} \delta_k^2}{n}}$$

Where δ_k is defined in Eqn. 1

3. Drowsiness Detection using YOLOv5:

For the extracted faces F in the dataset $D^{`}$:

Apply YOLOv5 on the extracted faces, F to predict the drowsiness labels P = $\{p_1, p_2, p_3, \ldots, p_n\}$

Pseudocode for the proposed methodology are shown as follows:

Input: Dataset D {UTA U, Custom X}

Output: C {Drowsy, Non drowsy}

--------**Data Collection and splitting** -------------

1- $U \rightarrow \{U_1, U_2, U_3, \ldots\ldots, U_N\}$; $U_i \in R^{L*W*C}$

2- $X \rightarrow \{X_1, X_2, X_3, \ldots\ldots, X_M\}$; $X_i \in R^{L*W*C}$

3- $F \leftarrow \{U \,||\, X\}$; $F \rightarrow \{F_1, F_2, F_3, \ldots\ldots, F_{M+N}\}$

4- $\{F_{train}\} \,||\, \{F_{test}\} \leftarrow F$

--------------**Data Augmentation** --------------------

5- *For each* F_i *in* F_{train}:

6- *Augmentation (set of transformations)* $A: F \rightarrow F'$

7- $F_i' \leftarrow A(F_i)$

8- Augmented training set: $F_{train}' = \{F_1', F_2', F_3', \ldots, F_{train_{N+M}}'\}$

9- END For

--------------**Face Extraction using Facial Landmark**------------

10- $F^{(+)} \leftarrow \{F_{train}'\} || \{F_{test}\}$

11- *For each* $F_i^{(+)}$ *in* $F^{(+)}$:

12- *facial landmarks:* $L_i = \{L_{1i}, L_{2i}, L_{3i}, \ldots, L_{ni}\} \leftarrow$ *landmark predictors*$(F_i^{(+)})$

13- *Extracted faces:* $EF_i \leftarrow \sum_{i=1}^{n} L_i$

14- *For* EF_i, $NRMSE_{local} = \sqrt{\frac{\sum_{k=1}^{n} \delta_k^2}{n}}$

15- *END For*

--------------**Face classification using YoloV5**------------

16- *For each* EF_i *in EF:*

17- *drowsiness labels:* $P_i = \{p_{1i}, p_{2i}, p_{3i}, \ldots, p_{ni}\} \leftarrow$ *YOLOv5* (EF_i)

18- C_i *{Drowsy, Non drowsy}* $\leftarrow P_i$

19- *END For*

Since most of the datasets have wide-angle, the accuracy of the model is enhanced by extracting a Region of Interest (ROI) from these images. The YoloV5 framework is used for automated recognition of faces and cropping in wide-angle images. YoloV5 is a hybrid network that combines elements of both the Cross Stage Partial (CSP) and Darknet (Darknet) architectures. Annotating ROI for a collection of wide-angle frames of UTARLDD dataset allows us to generate the weights and locations for the YoloV5 architecture. From the annotated input image, the CSP and Darknet grid extract features and target information. The input vector is partitioned into A × A grids for face recognition. If the geographic center falls inside a certain grid, the grid is responsible for determining the target. The face's regression box's coordinates are determined using the following equation (3):

$$R_{p,q} C_p^q = R_{p,q} \times IOU_{\text{Predicted}}^{True} \quad (3)$$

Where C_p^q represents the confidence score (C.S) of the qth bounding box of the pth grid. $R_{p,q}$ represents whether there exists a target in the qth bounding box of the pth grid. Then $R_{p,q}$ will be equal to one if not zero. The IOU True Predicted is a commonly used metric that describes the Intersection Over Union (IOU). A high IOU score indicates that the predicted box is located more accurately.

4. Experimental Framework

In this section, we delve into the specifics of implementation, outline the experimental setup, provide details about the dataset, and elaborate on the evaluation parameters

4.1Experimental setup

The system was developed and evaluated using a workstation running Windows 10 (64-bit) with a single NVIDIA GeForce 940 M GPU, using the Version 376.82 driver. The workstation is equipped with an Intel(R) Xeon® W-2022 CPU running at 2.90-GHz and 1-TB RAM. For backup, a 2 TB Seagate hard disk is used. The system was run on a Docker environment on Windows, using TensorFlow.

4.2Datasets

Two video datasets are used to evaluate the proposed system, namely 1) The Real-Life Drowsiness Dataset from the University of Texas at Arlington (UTARLDD) and 2) a custom dataset. For drowsiness detection, the proposed method just needs one frame. Therefore, the framework's input consists of images that have been retrieved from the video dataset.

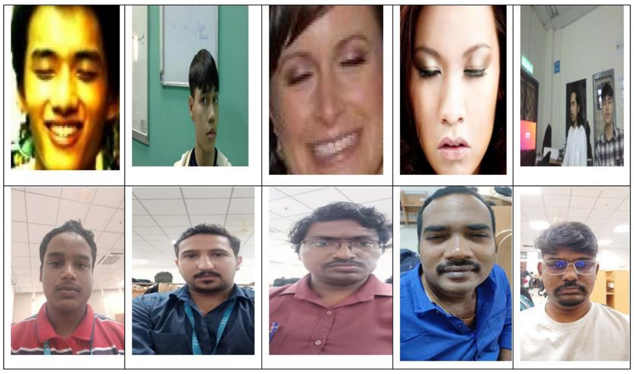


Fig. 3. Samples images from the dataset top are UTA dataset and the bottom is a custom dataset

To detect drowsiness, the system processes individual frames extracted from the two video datasets. The experiments were conducted using the UTA-RLDD dataset, which is the most significant publicly available benchmark dataset for drowsiness detection. This dataset includes videos of 36 participants in distinct scenarios, including drowsiness-related symptoms such as yawning and nodding, as well as non-drowsiness activities like talking phone, laughing, and gazing. Each video is around 90 seconds long. From each participant, Random frames were collected and binarily labeled based on the drowsiness status. The frames have a resolution of 640 × 480, indicating a relatively high level of image clarity. The dataset exhibit transformations like scale, pose, and perceptions, making it an appropriate benchmark dataset to demonstrate the proposed system's efficiency and performance in practical scenarios.

The custom dataset consists of 39 subjects captured using a high-resolution DSLR camera. This dataset introduces variations in posture, view of angle, and orientation, making it more challenging than the UTA-RLDD dataset. In total 20 subjects in the age range of 18 to 60 were chosen for the dataset. Of the 20, 10 were male and others were female. We had an equal number of people wearing eyeglasses. The subjects were taken during both the day and nighttime.

4.3Performance metrics

Driver drowsiness detection is an important application of machine learning that aims to prevent accidents caused by tired drivers. Several performance metrics can be used to evaluate the effectiveness of driver drowsiness detection systems. Accuracy is the most common metric used to evaluate classification models, which predict whether a driver is drowsy or alert. It measures the proportion of correctly classified instances over the total number of instances. However, accuracy can be biased in cases where the dataset is imbalanced (i.e., there are more instances of one class than the other).

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (4)$$

The second important measure is Precision, which measures the proportion of true positives (i.e., instances where the driver is drowsy and the model predicts drowsiness) over the total number of instances predicted as positive (drowsy). Recall measures the proportion of true positives over the total number of instances that are positive (drowsy). These metrics are useful when the cost of false positives or false negatives is different.

$$Precision = \frac{TP}{(TP + FP)} \quad (5)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (6)$$

In this context, T_P refers to the number of images that are correctly identified as being in a drowsy state. F_N represents the number of images that are incorrectly classified as not being in a drowsy state. F_P denotes the number of images that are erroneously classified as being in a drowsy state, despite actually being in a non-drowsy state. T_N is the identified image of a non-drowsy state.

$$F1\ Score = \frac{2 * Precision * Recall}{(Precision + Recall)} \quad (7)$$

The F1 score, being the harmonic mean of precision and recall, offers a unified metric that strikes a balance between these two measures. It proves particularly valuable in scenarios of imbalanced datasets, as it takes into consideration both false positives and false negatives.

5. Experimental Analysis and Results

The YoloV5 architecture serves as the core architecture in our system, configured to accurately detect faces and extract ROIs from input images. The model is trained with 250 epochs, and the confidence score is set to 0.75. Fig. 4 depicts the sample drowsiness detection results in both day and night time.

Fig. 4. Detection results in daytime and nighttime

The metrics like precision, recall, and mean average precision metric (mAP) at a 0.50 threshold are plotted in Fig. 5. Fig. 6 shows the detection effect of the trained architecture. After that, we calculated the speed of the trained YoloV5 to be 51.90 images per second. Once the bounding box locations are obtained, the ROI is cropped.

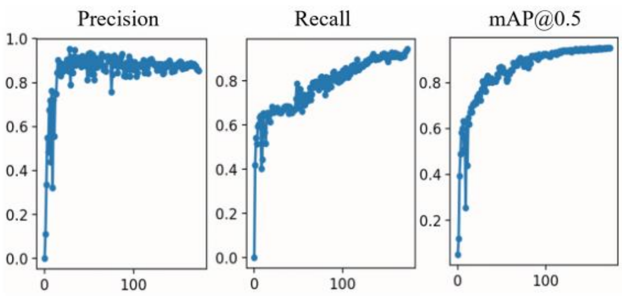


Fig. 5. YoloV5 Performance for 250 epochs

Our Proposed framework is evaluated using typical benchmarks for classification models. The learning curves for accuracy and loss are shown in Fig. 6. The plots indicate a well-fitted learning algorithm. To improve performance, we incorporated three missions simultaneously during the training of our efficient proposed model: computing output, debugging errors, and tuning hyperparameters. Finally, we achieved 95.20% and 96.40% maximum training and validation accuracies.

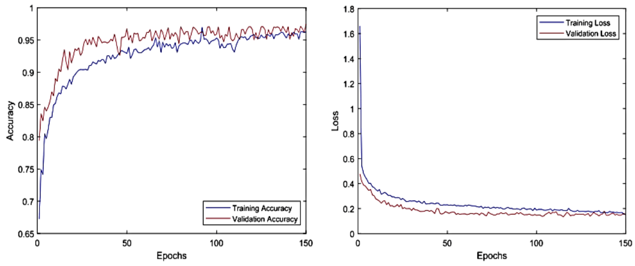


Fig. 6. Training and validation accuracy and loss for 250 epochs

We consider two video samples one is 30 seconds and another one is 90 seconds. Got an accuracy of 96.42% for Non drowsy, 91.66% for fatigue, and 93.33% for fully drowsy these are the results for 30-second video samples. For 90-second video sample accuracy 94.44 for Non drowsy, 92.50 for fatigue, and 91.11 for full drowsy. The results are represented in Table 1.

Table 1. Accuracy of the proposed model on 30 seconds and 90 seconds video samples

State	Accuracy (30 sec videos) (%)	Accuracy (90 sec videos) (%)
Non-Drowsy	96.42	94.44
Fatigue	91.66	92.50
Mildly drowsy	90.90	93.33
Hypo vigilance	92.30	87.50
Fully drowsy	93.33	91.11

To further evaluate the performance of our proposed classification model, we calculated the hamming loss (H.L) and binary cross-entropy (B.C.E). Our trained model has B.C.E and H.L are 0.6906 and 0.0672, respectively. These values indicate good results, with the log loss being closer to zero. The cross-entropy loss penalizes inaccurate predictions excessively, we calculated the precision, recall, and F1-score, and their values are shown in Table 2.

Table 2. Performance of the proposed model on UTA dataset

States	Precision (%)	Recall (%)	F1-Score (%)
Non-Drowsy	96.0	94.0	94.0
Partially drowsy	97.0	94.0	96.0
Mildly drowsy	94.0	96.0	94.0
Moderately drowsy	95.0	95.0	94.0
Fully drowsy	97.0	97.0	96.0

We also evaluated the proposed system on different head positions. Table 3 describes the accuracy of driver drowsiness with the effect of head positions like Head Right, Head Left, Head Up, and Head Down.

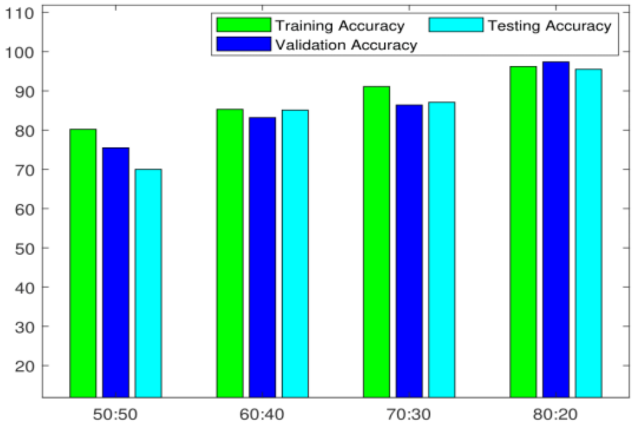


Fig. 7. The Impact of Data Splits on 4 different proportions like 50:50, 60:40, 70:30 and 80:20

We created a custom dataset consisting has 3 different scenarios: face, face with spectacles, and face with sunglasses captured during the day and night time. We

Table 3. Effect of head position in detecting driver drowsiness

State	Non-Drowsy	Fatigue	Mildly Drowsy	Hypo Vigilance	Fully Drowsy	Accuracy (%)
Head Right	Y	N	Y	N	Y	95.80
Head Left	N	Y	Y	N	Y	96.07
Head Up	N	Y	N	Y	N	93.18
Head Down	Y	Y	N	Y	Y	96.92

Table 4 describes the accuracy of drivers with glasses and without Glasses. With glasses, the accuracy is 95.50% and without glasses, the accuracy is 97.70%. The primary goal is to assess the performance of the proposed model on the custom dataset.

Table 4. Effect of Eyeglasses on different states of drowsiness

State	Non-Drowsy	Fatigue	Mildly Drowsy	Hypo Vigilance	Fully Drowsy	Accuracy (%)
With Glasses	Y	Y	Y	N	N	95.50
With Out Glasses	N	N	Y	Y	Y	97.70

Four different combinations of data splitting are considered. Each set has (80-20, 70-30, 60-40, and 50-50) combinations. Finally, 95.20%, 96.40%, and 94.50% are the best training, validation, and testing accuracies with a split of 80-20, as seen in Fig. 7.

Table 5. The model's average accuracies for daytime, evening, and nighttime images showed significant differences.

The model's average accuracy for daytime, evening time, and night time was 94.8%, 96.6%, and 91.9%, respectively. The model performed better in the evening time, primarily due to the optimal lighting conditions. The final accuracy of the custom dataset was 94.6%.

Table 5. Performance of the proposed model on custom dataset to test the effects of eyeglasses during daytime and night time

	Day time		**Night time**	
Scenario	**Non-drowsy (%)**	**Drowsy (%)**	**Scenario**	**Non-drowsy (%)**
With eyeglasses	94.0	95.0	With eyeglasses	94.0
Without eyeglasses	96.0	97.0	Without eyeglasses	96.0
Average	95.0	96.0	Average	95.0

with good occlusion handling but limited lighting robustness. Anh-Cang et al. [46] employed SSD Network with MobileNet-V2 and ResNet-50V2, reaching 96.9% accuracy with strong lighting robustness but poor eyeglass occlusion handling. The proposed framework, adopting YOLOv5 with facial landmark detection, outperformed all others with 97.3% overall accuracy, excelling in both occlusion handling and lighting robustness. In conclusion, the proposed framework demonstrated the best overall performance among the compared methods, with high accuracy in both occlusion handling and robustness to different lighting conditions.

6. Conclusion

In conclusion, this work presents a novel framework for driver drowsiness detection that addresses the limitations of existing methods by combining facial landmarks and the Yolov5 architecture, as it is crucial to select the appropriate combination of face detection and classification architectures to achieve optimal performance in identifying human drowsiness.

Table 6. Comparative study with respect to various factors like occlusion handling and lighting conditions

References	**Dataset**	**Detection techniques**	**Occlusion Handling with Eyeglass**	**Robustness to Different Lighting**	**Overall Accuracy**
Bakheet et al. [43]	NTHU	Haar Cascade with HOG, Naïve Bayes	✓	✓	84.61 %
Shreyans M et al. [44]	UTA	Logistic Regression	×	✓	74.77 %
R Tamanani et al. [45]	UTA	Haar Cascade with LeNet CNN	✓	×	92.0 %
Anh-Cang et al. [46]	Mixed	SSD Network, MobileNet-V2 and ResNet-50V2	×	✓	96.9%
Proposed Framework	UTA, Custom dataset	YOLOv5 with facial landmark	✓	✓	**97.3 %**

Table 6 presents a comparison of different research papers and their proposed frameworks for object detection. The objective is to evaluate the performance of various techniques in handling occlusion with eyeglasses and robustness to different lighting conditions, ultimately determining the overall accuracy of the detection. Researchers evaluated various object detection techniques on datasets like NTHU and UTA, comparing their performance in handling occlusion with eyeglasses and robustness to different lighting conditions. Bakheet et al. [43] achieved an 84.61% overall accuracy using Haar Cascade with HOG and Naïve Bayes. Shreyans M et al. [44] utilized Logistic Regression and attained a 74.77% accuracy with robust lighting handling but struggled with eyeglass occlusion. R Tamanani et al. [45] used Haar Cascade with LeNet CNN, achieving a 92.0% accuracy

By extracting relevant facial features and observing subtle indicators like facial expressions, eye movement patterns, and micro head gestures, the proposed system achieves a comprehensive understanding of the driver's state during operation. The evaluation of the UTA and custom datasets demonstrate the system's effectiveness, achieving high accuracies of 95.5% and 96.4% respectively. Compared to state-of-the-art techniques, the proposed framework showcases a significant improvement of 3.2% in drowsiness detection accuracy. This enhancement in accuracy and the ability to capture subtle aspects of driver behavior make the proposed system a promising solution for enhancing road safety and preventing accidents caused by driver fatigue.

6.1 Acknowledgment:

We thank VIT-AP University for providing the seed grant for this research.

Author contributions

Mohan Arava: Data collection, Methodology, Coding, Writing

Divya Meena Sundaram: Reviewing and Editing.

Declaration of Interest Statement:

We wish to confirm that there are no known conflicts of interest associated with this publication that could have influenced its outcome.

References

[1] Wörle, J., Metz, B., Thiele, C., & Weller, G. (2019). Detecting sleep in drivers during highly automated driving: The potential of physiological parameters. IET Intelligent Transport Systems, 13(8), 1241-1248.

[2] De Mello, M. T., Narciso, F. V., Tufik, S., Paiva, T., Spence, D. W., BaHammam, A. S., ... & Pandi-Perumal, S. R. (2013). Sleep disorders as a cause of motor vehicle collisions. *International journal of preventive medicine*, *4*(3), 246.

[3] Dong, Y., Hu, Z., Uchimura, K., & Murayama, N. (2010). Driver inattention monitoring system for intelligent vehicles: A review. *IEEE transactions on intelligent transportation systems*, *12*(2), 596-614.

[4] Ziebarth, M. (2020). NY v. national highway traffic safety admin., 974 F. 3d 87 (2d Cir. 2020). *Transp. LJ*, *47*, 87.

[5] Kumar, A., Kalita, D. J., & Singh, V. P. (2020, February). A modern pothole detection technique using deep learning. In *2nd International Conference on Data, Engineering and Applications (IDEA)* (pp. 1-5). IEEE.

[6] Capitaine, M. P., & Cárdenas, H. M. G. (2023). Artificial intelligence and advanced driver assistance systems absorption (ADAS) in Mexico. *Ciencia Nicolaita*, (88).

[7] Tibrewal, M., Srivastava, A., & Kayalvizhi, R. (2021). A deep learning approach to detect driver drowsiness. *Int. J. Eng. Res. Technol*, *10*, 183-189.

[8] Panwar, P., Roshan, P., Singh, R., Rai, M., Mishra, A. R., & Chauhan, S. S. (2022). DDNet-A Deep Learning Approach to Detect Driver Distraction and Drowsiness.

[9] Cao, Y., Li, F., Liu, X., Yang, S., & Wang, Y. (2023). Towards reliable driver drowsiness detection leveraging wearables. *ACM Transactions on Sensor Networks*, *19*(2), 1-23.

[10] Yang, C., Yang, Z., Li, W., & See, J. (2022). FatigueView: A Multi-Camera Video Dataset for Vision-Based Drowsiness Detection. *IEEE Transactions on Intelligent Transportation Systems*, *24*(1), 233-246.

[11] Bassil, M., Perrine, K. A., & Machemehl, R. B. (2023). Synthesis of Automated Multimodal Data Collection Techniques and Applications in the US. In *International Conference on Transportation and Development 2023* (pp. 480-488).

[12] Kotseruba, I., & Tsotsos, J. K. (2022). Attention for vision-based assistive and automated driving: a review of algorithms and datasets. *IEEE transactions on intelligent transportation systems*.

[13] Rajawat, A. S., Goyal, S. B., Bhaladhare, P., Bedi, P., Verma, C., Florin-Emilian, Ț., & Candin, M. T. (2023, May). Real-Time Driver Sleepiness Detection and Classification Using Fusion Deep Learning Algorithm. In *Proceedings of International Conference on Recent Innovations in Computing: ICRIC 2022, Volume 1* (pp. 447-457). Singapore: Springer Nature Singapore.

[14] Ayyasamy, S. (2022). A Comprehensive Review on Advanced Driver Assistance System. *Journal of Soft Computing Paradigm*, *4*(2), 69-81.

[15] Ibrahim, M. S., Kamat, S. R., & Shamsuddin, S. (2023). THE APPLICATION OF DRIVING FATIGUE DETECTION AND MONITORING TECHNOLOGIES IN TRANSPORTATION SECTOR: A REVIEW. *International Journal of Technology Management and Information System*, *5*(2), 30-42.

[16] Zhang, Z., Ning, H., & Zhou, F. (2022). A systematic survey of driving fatigue monitoring. *IEEE transactions on intelligent transportation systems*.

[17] Manfreda, A., Presbury, R., Richardson, S., Melissen, F., & King, J. (2023). Walking the talk: A High Engagement Research implementation framework in the qualitative study of tourism and hospitality experiences. *Tourism Management Perspectives*, *48*, 101142.

[18] Mohsan, S. A. H., Khan, M. A., Noor, F., Ullah, I., & Alsharif, M. H. (2022). Towards the unmanned aerial vehicles (UAVs): A comprehensive review. *Drones*, *6*(6), 147.

[19] Ansari, S., Du, H., Naghdy, F., & Stirling, D. (2022). Automatic driver cognitive fatigue detection

based on upper body posture variations. *Expert Systems with Applications*, *203*, 117568.

[20] El-Nabi, S. A., El-Shafai, W., El-Rabaie, E. S. M., Ramadan, K. F., Abd El-Samie, F. E., & Mohsen, S. (2023). Machine learning and deep learning techniques for driver fatigue and drowsiness detection: a review. *Multimedia Tools and Applications*, 1-37.

[21] Li, K., Yu, R., Liu, Y., Wang, J., & Xue, W. (2023). Correlation analysis and modeling of human thermal sensation with multiple physiological markers: An experimental study. *Energy and Buildings*, *278*, 112643.

[22] Fan, J., Yang, S., Liu, J., Zhu, Z., Xiao, J., Chang, L., ... & Zhou, J. (2022). A High Accuracy & Ultra-Low Power ECG-Derived Respiration Estimation Processor for Wearable Respiration Monitoring Sensor. *Biosensors*, *12*(8), 665.

[23] Sharma, G., Joshi, A. M., Yadav, D., & Mohanty, S. P. (2023). A Smart Healthcare Framework for Accurate Detection of Schizophrenia using Multi-Channel EEG. *IEEE Transactions on Instrumentation and Measurement*.

[24] Sharma, M., Darji, J., Thakrar, M., & Acharya, U. R. (2022). Automated identification of sleep disorders using wavelet-based features extracted from electrooculogram and electromyogram signals. *Computers in Biology and Medicine*, *143*, 105224.

[25] Farooq, H., Jain, A., & Shukla, M. K. (2022, August). Classification of Sleep using Polysomnography. In *Journal of Physics: Conference Series* (Vol. 2327, No. 1, p. 012064). IOP Publishing.

[26] Sar, I., Routray, A., & Mahanty, B. (2023). A Review on Existing Technologies for the Identification and Measurement of Abnormal Driving. *International Journal of Intelligent Transportation Systems Research*, *21*(1), 159-177.

[27] Ponnan, S., Theivadas, J. R., HemaKumar, V. S., & Einarson, D. (2022). Driver monitoring and passenger interaction system using wearable device in intelligent vehicle. *Computers and Electrical Engineering*, *103*, 108323.

[28] Bandele, O. O. (2022). *Real-time drowsiness detection using computer vision and deep learning techniques* (Doctoral dissertation, Dublin, National College of Ireland).

[29] Jan, M. T., Hashemi, A., Jang, J., Yang, K., Zhai, J., Newman, D., ... & Furht, B. (2022, October). Non-intrusive drowsiness detection techniques and their application in detecting early dementia in older drivers. In *Proceedings of the Future Technologies Conference* (pp. 776-796). Cham: Springer International Publishing.

[30] Guo, B., Hua, Q., Jin, L., Xie, X., Huo, Z., & Wang, H. (2022). Analysis of driving control characteristics in typical road types. *Sustainability*, *14*(2), 782.

[31] Shaik, M. E. (2023). A systematic review on detection and prediction of driver drowsiness. *Transportation Research Interdisciplinary Perspectives*, *21*, 100864.

[32] Mu, Z., Jin, L., Yin, J., & Wang, Q. (2022). Research on a driver fatigue detection model based on image processing. *HUMAN-CENTRIC COMPUTING AND INFORMATION SCIENCES*, *12*.

[33] Basit, M. S., Ahmad, U., Ahmad, J., Ijaz, K., & Ali, S. F. (2022, December). Driver Drowsiness Detection with Region-of-Interest Selection Based Spatio-Temporal Deep Convolutional-LSTM. In *2022 16th International Conference on Open Source Systems and Technologies (ICOSST)* (pp. 1-6). IEEE.

[34] Krishna, G. S., Supriya, K., & Vardhan, J. (2022). Vision transformers and YoloV5 based driver drowsiness detection framework. *arXiv preprint arXiv:2209.01401*.

[35] Jilani, U., Asif, M., Rashid, M., Siddique, A. A., Talha, S. M. U., & Aamir, M. (2022). Traffic congestion classification using GAN-Based synthetic data augmentation and a novel 5-layer convolutional neural network model. *Electronics*, *11*(15), 2290.

[36] Krishna, G. S., Supriya, K., & Vardhan, J. (2022). Vision transformers and YoloV5 based driver drowsiness detection framework. *arXiv preprint arXiv:2209.01401*.

[37] Lee, C., & An, J. (2023). LSTM-CNN model of drowsiness detection from multiple consciousness states acquired by EEG. *Expert Systems with Applications*, *213*, 119032.

[38] Ghosh, L., Dewan, D., Chowdhury, A., & Konar, A. (2021). Exploration of face-perceptual ability by EEG induced deep learning algorithm. *Biomedical Signal Processing and Control*, *66*, 102368.

[39] Jegham, I., Alouani, I., Khalifa, A. B., & Mahjoub, M. A. (2023). Deep learning-based hard spatial attention for driver in-vehicle action

monitoring. *Expert Systems with Applications*, *219*, 119629.

[40] Safarov, F., Akhmedov, F., Abdusalomov, A. B., Nasimov, R., & Cho, Y. I. (2023). Real-time deep learning-based drowsiness detection: leveraging computer-vision and eye-blink analyses for enhanced road safety. *Sensors*, *23*(14), 6459.

[41] Barakoti, S. (2023). Enhancing driving safety using artificial intelligence technology.

[42] Li, M. L., Sun, G. B., & Yu, J. X. (2023). A pedestrian detection network model based on improved YOLOv5. *Entropy*, *25*(2), 381.

[43] S. Bakheet and A. Al-Hamadi, "A framework for instantaneous driver drowsiness detection based on improved hog features and na¨ıve bayesian classification," Brain Sciences, vol. 11, p. 240, 02 2021

[44] S. Mittal, S. Gupta, Sagar, A. Shamma, I. Sahni, and N. Thakur, "Driver drowsiness detection using machine learning and image processing," in 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2021, pp.

[45] R. Tamanani, R. Muresan, and A. Al-Dweik, "Estimation of driver vigilance status using real-time facial expression and deep learning," IEEE Sensors Letters, vol. 5, no. 5, pp. 1–4, 2021

[46] A.-C. Phan, N.-H.-Q. Nguyen, T.-N. Trieu, and T.-C. Phan, "An efficient approach for detecting driver drowsiness based on deep learning," Applied Sciences, vol. 11, no. 18, 2021