# DDNet- A Deep Learning Approach to Detect Driver Distraction and Drowsiness

Panwar, Prachi
Department of CSE, GL Bajaj Institute of Technology and Management

Roshan, Prachi
Department of CSE, GL Bajaj Institute of Technology and Management

Singh, Rajat
Department of CSE, GL Bajaj Institute of Technology and Management

Rai, Monika
Department of CSE, GL Bajaj Institute of Technology and Management

他

https://doi.org/10.5109/4843120

出版情報：Evergreen. 9 (3), pp.881-892, 2022-09. 九州大学グリーンテクノロジー研究教育センター
バージョン：

# DDNet- A Deep Learning Approach to Detect Driver Distraction and Drowsiness

Prachi Panwar[1*], Prachi Roshan[1], Rajat Singh[1], Monika Rai[1], Asha Rani Mishra[1*]
Sansar Singh Chauhan[1]
[1]Department of CSE, GL Bajaj Institute of Technology and Management, Greater Noida, India

*Author to whom correspondence should be addressed:
E-mail:asha1.mishra@gmail.com

(Received June 16, 2022; Revised August 23, 2022; accepted September 12, 2022).

**Abstract**: Road accidents are the main cause of death among the human population. Distracted and drowsy driving takes thousands of lives every year around the world. Subsequently, to forestall such mishaps and save lives, there is a requirement for a system that detects both distraction and drowsiness for both day and night time. In this paper, we present a deep learning convolutional model to detect distraction and drowsiness during driving. The proposed model performs real-time video processing for monitoring the activities of drivers during driving. The model produces an alert in case of any careless driving or inappropriate behaviour of the driver with the minimum response time. For this purpose dataset for training as well as for testing were prepared. For training the model, we have used CNN model. The proposed model was able to achieve 99.95% accuracy on test dataset.

Keywords: Deep Learning, Convolutional Neural Network, Classification, ReLu, Validation loss, Real-time video processing, Drowsiness and Distracted Driving Detection System

## 1. Introduction

The World Health Organization (WHO) issued the global status report on road safety 2018, highlighting that the number of annual road traffic deaths has reached 1.35 million. Road traffic injuries are now the leading killer of people aged 5-29 years [1,32]. The leading causes of road accidents are distracted and drowsy driving. The general lifestyle has become very stressful because of which rate of mental pressure is increasing in people. This disorder also causes drowsiness in driver's behaviour which leads to a high probability of accidents. According to the National Highway Traffic Safety Administration (NHTSA), about 100,000 police-reported crashes involve drowsy driving every year. These crashes result in more than 1,550 fatalities and 71,000 injuries[2]. Drowsy-driving is a frequently occurring incident on roads and highways and occurs when the driver is too tired to stay in an alert state. According to the AAA Foundation for Traffic Safety, Researchers suggest the prevalence of drowsy driving fatalities is more than 350% greater than reported [3]. NHTSA defines drowsy driving as

- Occurs at both times of the day, people experience dips in their circadian rhythm—the human body's internal clock that regulates sleep;
- Often involve only a single driver (and no passengers) running off the road at a high rate of speed with no evidence of braking;
- Frequently occurs on rural roads and highways.

Distracted driving is defined by NHTSA, as "Distracted driving is any activity that diverts attention from driving, including talking or texting on phone, eating and drinking, talking to people in the vehicle, fiddling with the stereo, entertainment or navigation system— anything that takes your attention away from the task of safe driving." One cannot drive safely unless the task of driving has one's full attention. Any non-driving activity one engages in is a potential distraction and increases one's risk of crashing. Distracted driving is dangerous, claiming 2,841 lives in 2018 alone. Among those killed: 1,730 drivers, 605 passengers, 400 pedestrians, and 77 bicyclists [2]. The various types of distracted driving are as follows:

*Visual Distractions:* These are caused by things that take the driver's eyes and concentration off the road, even for a fraction of a second. It might alter the GPS, radio, looking outside the window, or conversing with the individuals sitting in the vehicle.

*Manual Distractions:* It happens when the driver takes either of their hands off the steering wheel. It might be to answer a call or send a text, drink, or eat a meal while driving.

*Cognitive Distractions:* It occurs when the driver is not in a good state of mind because of stress or pressure of money, family, and is unable to focus on the road while driving. This scenario is dangerous and can lead to potential road accidents.

To decrease the number of car crashes and improve transportation safety, a system that can recognize and alarm drowsy and distracted drivers is the need of the hour. This issue has attracted many researchers' attention in the past few years. This study is motivated to develop a system that can detect distracted and drowsy driving together and implement the same in real vehicles. In this paper, we aim to develop a deep learning model using the convolutional neural network (CNN) to detect distraction and drowsiness while driving. This system aims to monitor the driver during driving to maintain a safe driving environment.

The significant contributions of this proposed work are:

- Real-time tracking of the activities of the driver while driving.
- Working for both day and night time will ensure complete safety.
- Alert system in case the driver is detected to be distracted or drowsy with the minimum response time.
- Extend the safety features of the vehicle one level up.
- The system will reduce the rate of accidents caused by drowsiness and distraction while driving.

The goal of this research is to create a model that can identify both daytime and night time tired and distracted driving. The suggested model is trained on a self-created dataset that includes five labels: Attentive, Sleeping, Yawning, Using Mobile Phone, and Un-attentive to detect attentive driving. The training of the suggested model is carried out on a self-prepared dataset, and every effort has been made to cover as many real-life scenarios as possible, including various clothing styles, gender, and the individual wearing various accessories, such as a mask, glasses, cap, etc. In order to overcome the problem of late warning caused by analysis in by existing techniques, this paper's goal is to recognise driver drowsiness.

## 2. Literature Review

At present, increasing road accidents are prime concerns of many countries. Various reports and researches reveal that distracted driving is the main reason for the majority of the road accidents. Distracted driving can be defined as the state in which the focus of the driver is diverted from the driving to some other event. This may happen intentionally i.e. use mobile phones or drinking etc., or unintentionally i.e. tiredness, drowsiness etc. Many researches are going on to equip vehicles with such technologies that can detect such distractions at early stages which helps to reduce incidents in real-time may reduce the frequencies of road accidents. Distracted driving detection is the technique in which the state of the driver is detected and symptoms of distraction are observed. Some of these techniques are based on sensors, image processing techniques or AI based. Classification of these approaches is illustrated in Fig. 1.

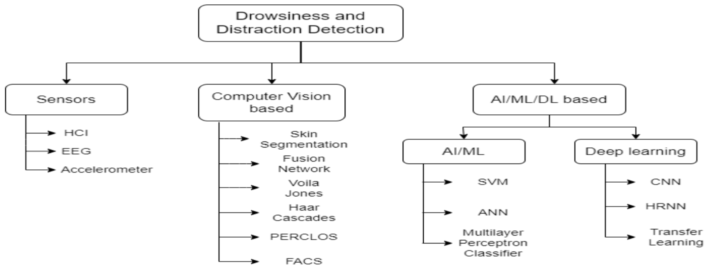


**Fig. 1:** Classification of drowsiness detection techniques.

Eden borough et al. [4] proposed a product named Driver State Monitor (DSM). The DSM is the system that analyses eye-closures and head pose to infer driver's fatigue or distraction. This system uses near-IR illumination and sensors to detect eye-closures and is designed to work in real-time. M.J. Flores et al. [5] also used the IR approach to process the visual information in order to locate, track and analyse both the driver's face and eyes to compute the drowsiness and distraction indexes. Kang [6] proposed an EEG based system to monitor the driver's state of mind during driving. He proposed EEG based approach in order to monitor the driver's driving behaviour. This approach detects non-physiological signals to detect the drowsiness symptoms in drivers. Heung-Sub et al. [7] use PPG sensors along with EEG to detect the drowsiness in the driver. The approach is based on HRV signals and physiological signals. Analysis of these signals carried out in frequency domain in order to detect drowsiness symptoms in drivers. In sensor based approach, EEG is a common technique to detect the physiological signals based drowsiness detection.

Abtahi et al. [8] developed a drowsy driving monitoring system for yawning detection. They applied face detection algorithms for more than 500 images with different scenarios of different drivers. The face gestures of drivers such as yawning, eye tiredness, and eye movements are used to categorize the driver fatigue. These gestures are detected on the basis of change in mouth geometry due to changes in driver's activity during driving. YIN et al. [9] proposed an approach for driver fatigue detection from facial image sequences. In the approach, the various algorithms and filters such as LBP and Gabor filter for the representation of multiscale features and computing dynamic features are applied on facial images of the driver. The proposed approach was tested on 600 images of 30 different people and achieves 98.33% accuracy in detection of driver fatigue.

An approach by Vesselenyi et al. [10] to detect driver's drowsiness used Artificial Neural Network (ANN) image processing. ANN layers were made to detect whether eyes are fully opened, half opened, or closed. ANN involves simple three-layer system input, hidden and output which is not as efficient as compared to CNN as CNN that involves multiple convolution and pooling layers before it. An approach proposed by Jabbar et al. [11] works on embedded systems that use Multilayer Perceptron Classifiers (MLP) for the processing of data. This MLP is a complex structure of nodes and their linking. It predicts the output using connection weights. The dataset used in training contains images of drivers from only side view. This design limits the practical usability because the upfront view of the driver is most suitable for the same task. The eyelids movement tracking is used in various researches for detection of drowsiness symptoms in drivers. AL Anizy et al. [12] use a cascade classifier along with a Support Vector Machine (SVM) to track the eyelid movements of the vehicle's driver. Features similar to Haar are used for face and eyelids detection. The SVM based classifier is used to classify the eyelids into open or closed categories. This hybrid approach is applied to each frame for the real-time video of the vehicle's driver. Aksjonov et al. [13] proposed an approach for the detection of distracted drivers that involves analyzing various in-vehicle secondary tasks that happen during the driving. This analysis is carried out by ANN which is trained on various dataset and able to classify the activities such as talking, sleeping etc. The dataset is composed of the various scenes obtained by some social experiments. This approach has limited scope as the in-vehicle secondary tasks can be subjective and depends on person to person. Also these assumptions are not strong enough to detect the drowsiness and driver's distraction. Vural et al. [14] proposed any hybrid approach composed of facial action tracking and accelerometer. They have trained a SVM model to classify 31 different types of facial actions. Also they used an accelerometer to record the movement of the driver's head. All this together is used to detect the drowsiness of the driver. Altoibi and Altoibi [15] proposed a deep learning model for detection of drowsiness. In the proposed deep learning model, they updated the traditional inception module by adding two Long Short-Term Memory (LSTM) layers along with a Residual Network (ResNet) layer of two convolution layers. Approach is tested on two different datasets i.e. State farm's dataset on Kaggle and AUC Dataset. Proposed model achieved better learning with reduced parameters as compared to existing deep learning models. A similar approach was proposed by Eraqi et al. [16]. The proposed deep learning model composed of five different CNN models connected as fully connected layers. Out of 5 CNN models, first one classifies the face of the drivers, second is used for driver's hand detection and classification, third is used for face and hand combined classification and fourth is used for skin classification. The last model is used for classification of full image of the camera view point. These five different networks were combined in a hybrid network with the help of a fully connected layer. The hybrid network classifies an image into 10 different classes related to driving activity. Manu B.N. et al. [17] also worked on skin segmentation to detect drowsiness a mechanism is developed which is divided into three junctures. In the first juncture, Viola Jones algorithm is used to detect the facial features of the subject along with eye tracking, and yawning detection. After the face detection, separation of skin part is done, only chromatic components are considered to eliminate the background images predicated on the skin color to make the system illumination invariant. Correlation coefficient template matching id used for tracking of eyes and yawning detection. The feature vectors from above junctures were considered and a binary support vector machine classifier distinguishes between the drowsy and non-drowsy frames. After a given number of drowsy frames an alarm is raised. Basubeit et al. [18] and Tamas et al. [19] have used VGG16 CNN network for training the dataset. The model is tuned on Kaggle dataset and achieves satisfactory results. Ngxande et al. [20] uses Generative Adversarial Networks (GAN) to produce synthetic images that can be reused for training. It also uses population bias that points out the region of good performance in the model. This approach was performed on a pretrained model of ResNet consisting of 50 layers. The resulting dataset of GAN was not as useful as it seemed to train the model due to its smaller size. One more way to use these transfer learning models is to use two or more models together. VGG16 is the most basic model and it requires an image size to be above a certain limit. If not so or the size of the image is small then an error occurs during the training period. New models with deeper layers have been introduced over a period of time which give better results namely ResNet50, GoogLeNet, Xception, etc. This approach is used by Sheng et al. [21]. In their paper, they used four deep CNNs consisting of AlexNet, GoogLeNet, ResNet, and VGG-16, which were trained and assessed on an embedded GPU platform. In addition, they have developed a chat alert system that alerts the driver in real time when the driver is not focused on the driving task. In addition, the results show that GoogleNet is the best model for distraction detection in the testbed driving simulator.

During years many researches are available for the detection of drowsiness and distracted driving. Some are using traditional programming approaches; others are using sensors, also machine learning is considered as the best option. In various research papers, different approaches have been used in place of machine learning. But these approaches are not fully applicable in the real world or have some downsides. Most of them are based on computer vision algorithms and image processing. They track facial expressions like eye blinking and yawning, head pose, and driver gaze. He et al. [22] proposed an approach in which they could identify fatigue by eye blinking. They are using three variables for checking the fatigue recurrence of head nods, rotation of the head, and PERCLOS.

Jun-Juh Yan et al. [23] built a model that detects fatigue to declare the driver is in a drowsy state. The model works in 3 major steps. First is to detect the eye positions of the driver from the grey scale feed received. Then a driver's personal fatigue model is developed in the second step of the process. In the third step, this fatigue model is compared to the feed to detect drowsiness and raise alarm. Extension of this method can be seen in research proposed by Galarza et al. [24] in which they also included the yawning frequency and considering changes in physiological measures such as variation in heart rate, brain activity, and body temperature. Both of these methods are using face detection algorithms and defining some rules for distinguishing between fatigue and normal behaviour. The major problem with image processing measures is that it can easily be tricked or it doesn't work under certain conditions such as if the eyes are covered with some kind of glasses or shadow by wearing a cap or not observable by the existing face or eye detection algorithms. Haisong Gu et al. [25] developed the infrared based eye detector for gaze determination with the relative spatial positions between pupil and the glint and further they did furrow detection and head position estimation to observe all the facial features of the driver. These facial features were compared to fatigue expression features and predictions were made using FACS. Vural et al. [26] also used FACS to train a SVM to detect 31 types of facial actions. They also used an accelerometer to detect the movement of the driver's head. All this together was used to detect the drowsiness of the driver. Apart from this, Maryam Hashem et al. [27] detected eyes out of the face using Viola Jones technique and cropped the image and chose one out of them. To overcome the challenge of the lighting condition, the authors used a histogram equalizer to equalize eye contrast. This image of the eye is used in NN to detect if it is closed. Similar approach was used by_Adriana Revelo et al. [28] who developed an algorithm which detects faces using Viola Jones algorithm in front images of drivers which are taken through an infrared camera. Then eyes are detected to classify them between closed and open eyes. Two techniques were used for eye classification. First, by extracting maximums and minimums of horizontals and vertical edges of the eyes and second a multi-layer perceptron neural network is used. Lastly when the eyes are closed drowsiness is detected.

There exist many approaches that involve machine learning which proved to be better than the models previously discussed. Basically, these models are using ANN and other neural network methods. Autonomous robotic systems that incorporate a novel decision-making module such as convolution neural networks (CNN) were used by Himanshu Srivastava et al.[29]. The study was conducted for the safety of highway maintenance personnel in order to prevent mishaps or a hazard in the direct emergency lane of the highway by implementing a danger sensor device known as the Signal Warning Detector (SWAD) was used by Hassan, Suhaimi, et al.[30]. Although RFID systems play an important role in asset tracking, their performance may be hampered when they are required to work in adverse weather conditions Choudhary, Shilpa, et al.[31]. To determine system parameters for simulation purpose, time series based on neural networks was used by Dief, Tarek N., and Shigeo Yoshida[32]. Phan, Anh-Cang et al.[33] proposed different conditions for alert system based on the concepts of deep learning algorithm. First method uses facial landmark MobileNet-V2 and ResNet-50V2 deep models to design adaptive neural network. Proposed method achieves accuracy of 97%. P. P. Patel et al.[35] proposed a deep learning based on CNN model along with Raspberry Pi and a mounted camera.The proposed model was able to achieve accuracy of 96%.

Most of the existing approaches are based on static features such as facial features, eye tracking and head pose. Due to these static features, these approaches faced efficiency issues in some specific scenarios such as when the driver cover eyes with sunglasses or face was covered by a cap or mask etc.

Another limitation of the existing approaches is that, these approaches are developed for a very specific context i.e. in order to be deployed in vehicles for real-time tracking; these will require major adjustments. The above literature is evident that there is a lack of approaches that can detect distracted driving and drowsiness in very early stages.

This paper proposes a solution to overcome existing limitations for drowsiness and distracted driving detection. The paper aims to propose such a model that can be adopted in vehicles without any adjustments. In this paper, a dataset is prepared which contains a variety of the images that corresponds to detect drowsiness and driving distractions together for each possible scenario.

## 3. Preparation of Dataset for DDNET

A lot of datasets related to this paper exist but these datasets are unable to fit in the proposed context; therefore, a dataset is developed to contemplate all requirements of the proposed approach. Thus, we have collected such videos for each category around the clock. This dataset contains diverse videos with changing environment, vehicles, person and poses. Maximum possible real-life scenario variations were taken care of like different cars, different clothing styles, gender, the volunteer wearing various accessories such as mask, glasses, sunglasses, cap, etc. Collecting datasets from different age-groups and gender of people adds different magnitudes to our dataset. The infrared camera is used for recording videos during the night environment. Table 1 and Table 2 describe the dataset in day and night environments.

Table 1. Videos corresponding to each category recorded into the daylight environment.

| S. No. | Activity | No. of Videos | Description |
|---|---|---|---|
| 1. | Un-attentive view | 18 | Activities where the driver is not attentive towards the driving i.e. driver looking here, there and backward frequently or talking to co-passenger. |
| 2. | Using mobile phone | 17 | Activities where the driver is on call over the mobile phone and driving with one hand only and head pose of the driver may be tilted. |
| 3. | Yawning | 15 | Activities where the driver is yawning frequently while driving and head pose of the driver may be tilted. |
| 4. | Sleeping | 17 | Activities where the driver is sleeping and head pose of the driver may be tilted. |
| 5. | Attentive view | 20 | Activities where the driver is in an attentive position and looking forward only. |
| Total | | 94 | 4200 frames per video |

Table 2. Videos corresponding to each category recorded into the night environment.

| S No. | Activity | No. of Videos | Description |
|---|---|---|---|
| 1. | Un-attentive view | 13 | Activities where the driver is not attentive towards the driving i.e. driver looking here, there and backward frequently or talking to co-passenger. |
| 2. | Using mobile phone | 20 | Activities where the driver is on call over the mobile phone and driving with one hand only and head pose of the driver may be tilted. |
| 3. | Yawning | 27 | Activities where the driver is yawning frequently while driving and head pose of the driver may be tilted. |
| 4. | Sleeping | 17 | Activities where the driver is sleeping and head pose of the driver may be tilted. |
| 5. | Attentive view | 18 | Activities where the driver is in an attentive position and looking forward only. |
| Total | | 97 | 4200 frames per video |

The proposed model focuses on attentive driving along with four types of distractions- un-attentive view, using mobile phone, yawning and sleeping. For training of the model, a precise and appropriate dataset with a front view of the driver was required. A lot of datasets were available but it did not take into consideration all four distractions. Therefore, dataset contemplating all needs and requirements was prepared. 12 volunteers were used for collection of data in the form of videos. Infrared cameras were used for recording videos. Each volunteer drove under varying degrees of lighting and was instructed to perform attentive driving and a set of distracting activities of sleeping, yawning, watching backward, using a mobile phone. The duration of the video recorded for each activity was 120 seconds.

Attentive driving includes driving the car vigilantly, looking into rear-view mirrors for a small (fixed) amount of time; if this time exceeds, the driver comes under the inattentive state. Next, un-attentive driving accounts for not looking in front for an amount of time that exceeds the attentive driving criteria. Using mobile phone orientation corresponds to the driver holding the phone to either ear. The driver is declared to be in a state of drowsiness if he is sleeping or yawning. Sleeping is detected when the driver's eyes are closed. Lastly, yawning is detected when the driver's mouth is open.

Maximum possible real-life scenario variations were taken into account for instance- different cars, different clothing styles, gender, the volunteer wearing various accessories like a mask, glasses, sunglasses, and cap. Collecting datasets from distinct age groups and gender of people adds different magnitudes to the dataset.

Each class contains 4000 frames of which 50% frames are of daylight and remaining is of night vision dataset. Identical frames were skipped and only frames with considerable variation for training to make the model learn better were used.

The pictorial description of the frames that were extracted from the dataset is outlined in Fig. 2 and sample images belonging to each category are shown in Fig. 3.

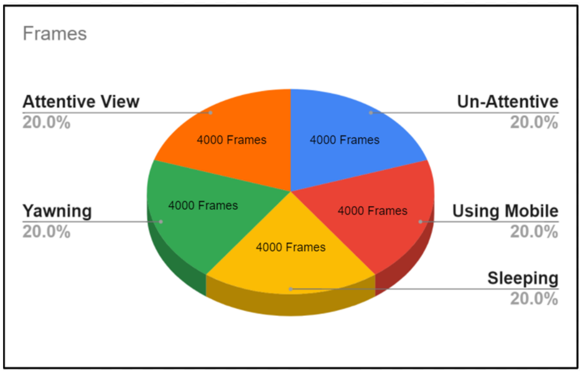


**Fig. 2:** Category wise frame distribution.


**Fig. 3:** Sample images for different views.

Each label is trained using 2000 frames for night as well as 2000 frames for daylight with a total of 4000 frames. So total frames used = (2000+2000)*5 category = 20,000 frames in the prepared dataset.

The videos for frame extraction with their total count for training and testing are shown as in Table 3 and Table 4.

Table 3. Daylight dataset statistics.

| S. No. | Activity | No. of Training Videos | No. Of Testing Videos | Total |
|---|---|---|---|---|
| 1. | Un-attentive view | 14 | 4 | 18 |
| 2. | Using Mobile Phone | 13 | 4 | 17 |
| 3. | Yawning | 10 | 5 | 15 |
| 4. | Sleeping | 13 | 4 | 17 |
| 5. | Attentive View | 16 | 4 | 20 |
| Total | | 66 | 28(21 +7 miscellaneous) | 94 |

Table 4. Night Dataset statistics

| S. No. | Activity | No. of Training Videos | No. of Testing Videos | Total |
|---|---|---|---|---|
| 1. | Un-attentive View | 11 | 2 | 13 |
| 2. | Using Mobile Phone | 16 | 4 | 20 |
| 3. | Yawning | 23 | 4 | 27 |
| 4. | Sleeping | 13 | 4 | 17 |
| 5. | Attentive View | 14 | 4 | 18 |
| Total | | 77 | 20(18+2 miscellaneous) | 97 |

## 4. Proposed Work

This paper proposes an approach for drowsiness and driving distraction detection while real-time driving. The main objective of this approach is to develop a model that can be used in vehicles and should be able to detect drowsiness and driving distraction caused due to various acts of the drivers. To accomplish the said objectives, the possible acts of the driver that can cause distraction are identified as 5 different categories. In these, the first category includes the acts in which the driver is watching forward and has full attention; this will classify acts as attentive driving. In the second category, the acts of the driver in which activity of the driver is not attentive due to watching backwards, such acts will be considered as un-attentive. Third category includes activities such as using a mobile phone while driving. The fourth and fifth categories include scenarios where the driver is frequently yawning and sleeping respectively. This paper uses a deep learning based approach to detect the act of the driver followed by a rule-based reasoning algorithm to classify the act of the driver and generate the early response. The proposed approach aims to utilize real-time video of the driver to accomplish the task. Hence, system architecture is developed for the same that illustrates the placement of the camera inside the vehicle and the viewpoint of the camera as shown in Fig. 4.

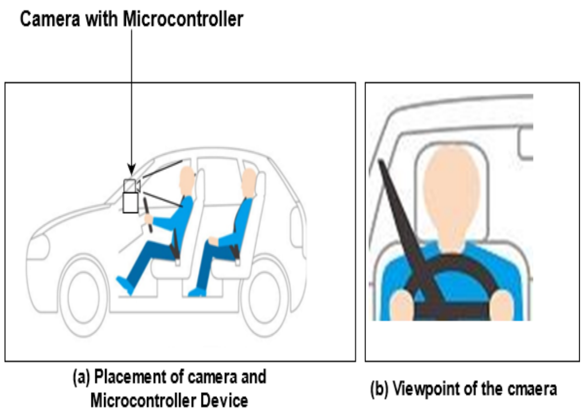


**Fig. 4:** Illustration of proposed placement of the camera and viewpoint of the camera.

The proposed approach works on the live feed received from the camera. The camera is connected to the microcontroller. This microcontroller is programmed to process the live video obtained from the camera. A trained CNN model has been deployed in the microcontroller so that it can classify the activity on frame by frame basis followed by a rule-based reasoning algorithm to detect possible events. The microcontroller also has an alarm to indicate the response of the system. Hence, this approach comprises three main phases. In the very first phase, the live feed received from the camera is preprocessed. In this preprocessing, frames are extracted from the live feed video, timestamp tagging on the frames and resizing related activities are performed. In the second phase, these series of the frames are fed into a trained CNN model. This CNN model classifies the frames as into predefined five classes. At last, the proposed rule-based reasoning algorithm classifies the activity of the driver into five different categories and generates appropriate responses. The above three phases of the approach are discussed in subsequent sections.

*Phase 1: Preprocessing of Live Feed*

In this phase, the processing of the live video feed obtained from the camera is carried out. The frames are extracted from video and timestamp tagging on respective frames is the major operation performed in this phase. Later, each frame is converted into an RGB color model and resized to specific size. Finally, a frame vector containing a series of frames, timestamp tag is made available for subsequent phases.

If the video feed received from the camera is 'V', the frame extracted from this video feed at timestamp t is f (t). Then generated frame vector is represented as:

frame_V = { (f(t), t), (f(t+1), t+1), (f(t+2), t+2), …………………}

*Phase 2: Deep Learning Model for Frame Classification*

Objective of this phase is to classify the preprocessed frames into five different categories such as attentive view, unattentive view, and use of mobile phone, yawning and sleeping. Descriptive flow chart of the proposed CNN model is shown in Fig. 5.

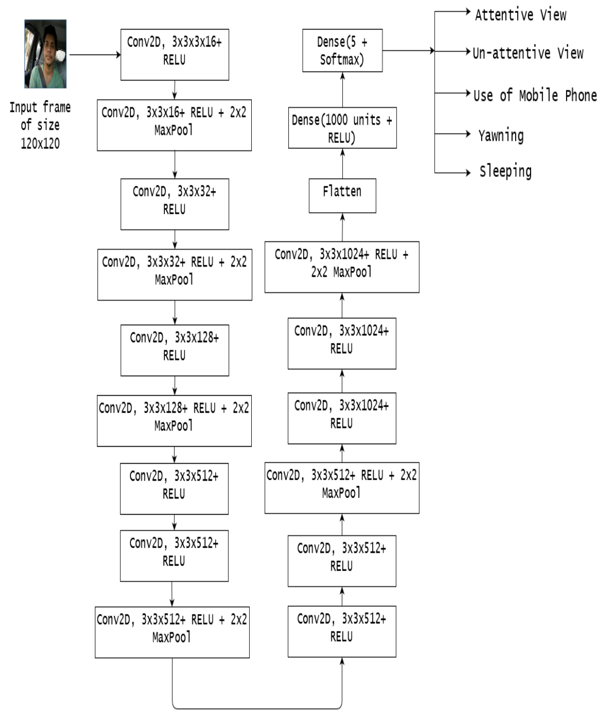


**Fig. 5:** Flowchart of the proposed model.

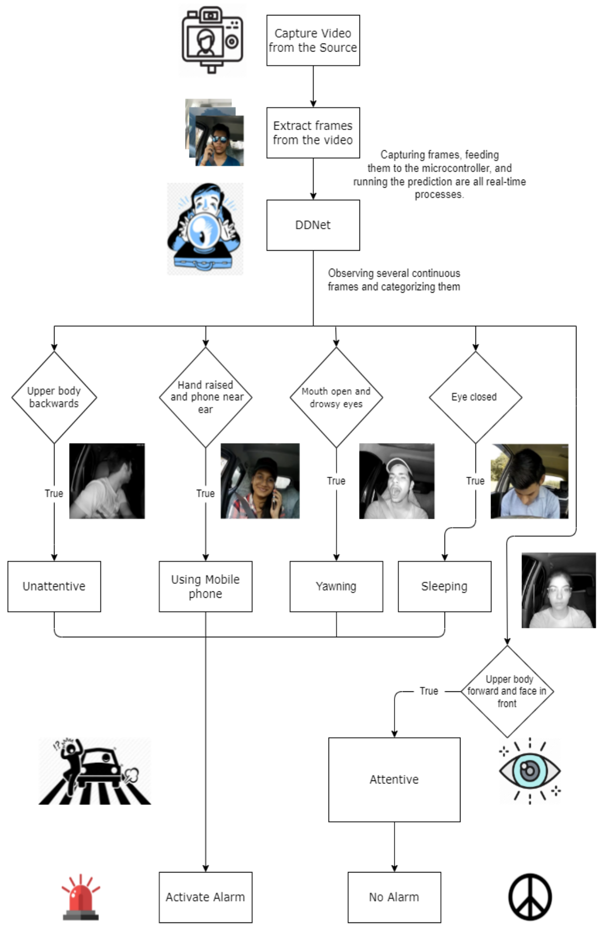


**Fig. 6:** Specification of the proposed CNN model.

Complete specification of the proposed model is shown in Fig. 6.

This model has a 15 convolution layer with varying number of filters at each layer from 16 to 1024, 1 fully connected layer (Dense) with 1000 units. The ReLu activation function as defined in Eq. 1 is used in all layers except the output layer and softmax function as defined in Eq. 2 at the output layer with 5 units. The categorical cross-entropy is an opted loss function. Finally, the RMS prop optimization algorithm is used to optimize the loss using Eq. 3.

$$R(z) = \begin{Bmatrix} z, & z > 0 \\ 0, & z \leq 0 \end{Bmatrix} \quad (1)$$

Where, z is the output at any layer.

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (2)$$

Where, z is the output of any layer, k =5 is the total number of classes.

$$\pounds(\hat{y}, y) = -\frac{1}{N}\sum_i^N [y_i log\hat{y}_i + (1-y_i)log(1-\hat{y}_i)] \quad (3)$$

Where, y, and y are predicted and actual targets, N is the total number of samples.

*Phase 3 Training of proposed CNN Model*

The proposed CNN model is trained on the self-developed dataset described earlier. During the training process, 70% dataset is used for training and the remaining 30% dataset is used for the validation process. In both cases, the videos from both sets of dataset i.e. daylight and night environment are included. After several hours of training, the stable CNN model is achieved, average 99% accuracy in training and 98% accuracy in testing. The hyper parameters used to train the CNN model are described in Table 5.

Table 5. Hyper parameters used in training of the CNN.

| **Sr. No..** | **Name of Hyperparameter** | **Value/ Function** | **Description** |
|---|---|---|---|
| 1. | Initial Learning rate | 0.0001 | Before the training |
| 2. | Learning rate decay | 0.000002 | Applied after per Epoch |
| 3. | Batch Normalization | kernel_regularizer='l2' | At the Dense layer (1000 units) and Output layer |
| 4. | Dropout | 0.5 | After Dense layer with (1000 units) |
| 5. | Batch Size | 32 | Training is carried out in total 32 batches |

To evaluate the performance of the model, the precision as defined in Eq. 4 and recall as defined in Eq. 5 is used.

$$\text{Precision} = \frac{tp}{tp + fp} \tag{4}$$

$$\text{Recall} = \frac{tp}{tp + fn} \tag{5}$$

Where, tp is true positive, fp is false positive, and fn is false negative.

***Workflow:***

Our trained model DDNET is programmed into the microcontroller to customize it for the vehicle. An infrared camera will be placed on the dashboard of the vehicle, above the steering wheel that will be connected to the microcontroller and the alarm machine. The infrared camera will capture the real-time video and extract the frames from it and will send them to the microcontroller where our model will classify the activities performed by the driver in the following classes-

- If the driver's upper body and face are backward while driving then he is put under the unattentive label. Placing the phone near either ear with a raised hand is inspected as using the phone. Open mouth and drowsy eyes for several consecutive frames is considered a Yawning. Sleeping is detected when the driver has his eyes closed for various successive frames. Attentive is detected when the driver has his upper body in forward direction and his face in front. If inattentiveness is detected as the final result for instance if the probability of un-attentive, using phone, yawning, or sleeping is greater than that of the forward, then it will raise an alarm which will alert the driver in minimum response time. The alarm connected will remain activated for a few seconds keeping human reaction time in mind. These detections will be concluded dynamically using a machine learning approach.

***System Architecture:***

Proposed CNN model DDNet is made by six sets of convolutional and pooling layers ensembled one after another. Input shape is kept 120*120 as we had produced our pickled data files with the same image size during the pre-processing step. The first set of layers consists of two convolutional layers with 16 neurons in each; similarly, the second set has two layers with 32 neurons in each along with a max-pooling layer stacked at the end of both the sets. The third, fourth, fifth and sixth sets have three convolutional layers of 128, 512, 512, 1024 neurons respectively along with a max-pooling layer connected afterwards in all the sets. Padding is used to ensure the output image's size remains the same as the input after revolving the kernel of dimension 3*3 during convolution. 'ReLu' activation function is used for convolutional layers. The pool size of 2*2 is fixed for max-pooling layers to extract the most prominent features of input images. After convolution and pooling, the flatten function is used to convert pooled feature map to a 1-d column. Subsequently, a fully connected dense layer with 1000 neurons is clamped along with the 'ReLu' activation function, and the l2 kernel regularizer to drive flattened data to the classification decision. A dropout layer with a value of 0.5 is added next to reduce over fitting. At last, an output layer consisting of 5 neurons is connected along with the softmax activation function and l2 kernel regularizer to reduce the weight of the network. The learning rate of 0.0001 together with a decay of 0.000002 is fixed with the RMSprop optimization technique to reach optimal weights and make our model learn better. Categorical cross-entropy loss functions for performance evaluation. Batch size is fixed to 32, maximum epochs to 50, and the callback function is used so that model does not iterate training epochs any further when performance gain is not optimal.

Proposed model contains layers with 1028 neurons; arranged as multiple convolutional layers with padding, activation function and max-pooling layer. Different activation functions were applied but settled upon ReLu as it gave us efficient results. Flatten layer is then added to flatten the features from previous layers into a single column to pass it to the fully connected layer. Dropout layer is added to randomly set input units to 0 with a frequency of rate at each step, which helps prevent over fitting. Optimizer functions are used to optimize the training after each epoch while the model is learning. The categorical cross-entropy loss function is used while training our model. Multiple dropout layers are used at different positions to prevent overfitting of the model. Callback function is used to obtain a best trained model.

## 5. Results and Discussion

The overall performance of the model has been evaluated based on its accuracy. The dataset has been divided into two parts in the ratio of 7:3. 70% is used for the training purpose and the remaining 30% is used for testing. Training data plays a very crucial role in any supervised machine learning technique. The same is also true for this model. The used dataset consists of high-quality frames and it is tried to keep the biasing at a minimum. The proposed CNN model after essential image processing gives an accuracy of 99% on the test dataset containing video sample.

Table 6 and Table 7 shows final statistics of the proposed model.

Table 6. Predicted and Actual Loss.

| Epochs | Accuracy | Loss | Validation Accuracy | Validation Loss |
|---|---|---|---|---|
| 1 | 0.7914 | 70.3326 | 0.4113 | 7.49 |
| 2 | 0.983 | 3.4932 | 0.9984 | 1.9541 |
| 3 | 0.9786 | 1.6101 | 0.9976 | 1.3119 |
| 4 | 0.9768 | 1.2658 | 0.9868 | 1.2641 |
| 5 | 0.9767 | 1.1403 | 0.9992 | 1.0294 |
| 6 | 0.979 | 1.0721 | 0.7216 | 2.0901 |
| 7 | 0.9759 | 1.0554 | 0.9976 | 1.0914 |
| 8 | 0.9984 | 0.4169 | 0.9992 | 0.3891 |
| 9 | 0.9982 | 0.3775 | 0.9984 | 0.3598 |
| 10 | 0.998 | 0.3669 | 0.9995 | 0.3477 |
| 11 | 0.9984 | 0.3605 | 0.9992 | 0.3431 |
| 12 | 0.9979 | 0.3576 | 0.9995 | 0.3312 |
| 13 | 0.9984 | 0.3523 | 0.9981 | 0.3368 |
| 14 | 0.9968 | 0.3514 | 0.9987 | 0.3319 |
| 15 | 0.9986 | 0.2907 | 0.9995 | 0.2738 |

Table 7. Accuracy and Loss value.

| Predicted vs. Actual | Unattentive | Using Mobile Phone | Yawning | Sleeping | Attentive | Total Frames |
|---|---|---|---|---|---|---|
| Unattentive | 11102 | 103 | 110 | 167 | 78 | 11560 |
| Using Mobile Phone | 147 | 10547 | 105 | 393 | 114 | 11306 |
| Yawning | 95 | 164 | 11167 | 190 | 132 | 11748 |
| Sleeping | 79 | 364 | 176 | 10803 | 117 | 11539 |
| Attentive | 101 | 95 | 39 | 72 | 11383 | 11690 |
| Total Frames | 11524 | 11273 | 11597 | 11625 | 11824 | 57843 |

Average Precision= {(11102 / 11524) + (10547/11273) + (11167/111597) + (10803/11625) + (11383/11824)} /5=0.95077

Average Recall= {(11102 / 11560) + (10547/11306) + (11167/11748) + (10803/11539) + (11383/11690)} / 5=0.95074

Fig. 7 and Fig. 8 show that the proposed deep model has promising results.


**Fig. 7:** Accuracy vs. Loss.


**Fig. 8**: Validation Accuracy vs. Validation Loss.

## 6. Conclusion

Drowsiness and distraction during driving have become more common and have led to potential accidents in recent years. The proposed deep learning model aims to classify the driving into five different classes. Inattentive driving includes four predefined labels, first is if the upper body is backward then the driver is put under the Backward label; second is if the driver's hand is raised and the phone is near his ear then distraction is labeled as Talking via Phone; third is if the driver's mouth is opened for several continuous frames then he is put under the Yawning label; fourth is if the driver's eyes are closed continuously for many frames then Sleeping is reported. If any label of inattentive driving is detected then DDNet raises an alarm to alert the driver in minimum response time. If the driver is sitting straight and his face is in the forward direction he is put under the fifth label i.e. Forward; in case of forwarding or attentive driving, no alarm is raised. CNN model, and the dataset used for training is self-collected and CNN model is used to train the dataset. An experimental result shows that model has achieved an accuracy of 99% for the dataset.

In the future, the scope of implementation of the system can be extended by enclosing all possible real-life environments- weather conditions and landforms using sensors. The system can be trained to dynamically switch between the various environments; this will lead to more precise monitoring and careful supervision. Data analysis can be used to build rule based algorithms for raising alarm in different environments keeping in mind driver's attention levels. More distraction labels can be added to enhance the system's usefulness. Implementation of this system in Taxis, Vending vehicles, Private vehicles, etc., can become the ultimate step of the road map.

## References


1) Global status report on road safety 2018. Geneva: World Health Organization; 2018. Licence: CC BY-NC- SA 3.0 IGO.
2) https://www.nhtsa.gov/risky- riving/drowsy-driving
3) Owens, T. A. Dingus, F. Guo, Y. Fang, M. Perez, J. McClafferty, and B. Tefft, "Prevalence of drowsy-driving crashes: Estimates from a large-scale naturalistic driving study", (2018).
4) Edenborough, N., R. Hammoud, A. Harbach, A. Ingold, Branislav Kisacanin, Phillip Malawey, T. Newman,"Driver state monitor from delphi" in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, **vol. 2,** pp. 1206-1207. IEEE, (2005).
5) Flores, MarcoJavier, JoséMaría Armingol, and Arturo de la Escalera, "Driver drowsiness warning system using visual information for both diurnal and nocturnal illumination conditions" in *EURASIP journal on advances in signal processing* (2010): 1-23.
6) Kang, Hang-Bong "Various approaches for driver and driving behavior monitoring: A review." in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 616-623, (2013).
7) Shin, Heung-Sub, Sang-Joong Jung, Jong-Jin Kim, and Wan-Young Chung. "Real time car driver's condition monitoring system"in *SENSORS, IEEE,* pp. 951-954, (2010).
8) Abtahi, Shabnam, Behnoosh Hariri, and Shervin Shirmohammadi. "Driver drowsiness monitoring based on yawning detection" in IEEE *International Instrumentation and Measurement Technology Conferenc*e, pp. 1-4. IEEE, (2011).
9) Yin, Bao-Cai, Xiao Fan, and Yan-Feng Sun, "Multiscale dynamic features based driver fatigue detection." *International Journal of Pattern Recognition and Artificial Intelligence*, **vol. 23** pp. 575-589, (2009).
10) Vesselenyi, Tiberiu, Sorin Moca, Alexandru Rus, T. Mitran, and B. Tătaru, "Driver drowsiness detection using ANN image processing." *In IOP conference series: materials science and engineering,* **vol. 252**, no. 1, p. 012097. IOP Publishing, (2017).
11) Jabbar, Rateb, Khalifa Al-Khalifa, Mohamed Kharbeche, Wael Alhajyaseen, Mohsen Jafari, and Shan Jiang," Real-time driver drowsiness detection for android application using deep neural networks techniques." *Procedia computer science* **vol.130,** pp. 400-407(2018),doi.org/10.1016/j.procs.2018.04.060.
12) AL-Anizy, Ghassan Jasim, Md Jan Nordin, and Mohammed M. Razooq,"Automatic driver drowsiness detection using haar algorithm and support vector machine techniques." *Asian J. Appl. Sci* **Vol.8**, no. 2 pp. 149-157, (2015).
13) Aksjonov, Andrei, Pavel Nedoma, Valery Vodovozov, and Eduard Petlenkov,"in-vehicle information system as a driver's secondary activity: Case study" in 15th *International Workshop on Advanced Motion Control (AMC),* pp. 523-528. IEEE, (2018).
14) Vural, Esra, Mujdat Cetin, Aytul Ercil, Gwen Littlewort, Marian Bartlett, and Javier Movellan, "Drowsy driver detection through facial movement analysis", *in international workshop on human-computer interaction*, pp. 6-18. Springer, Berlin, Heidelberg, (2007).
15) Alotaibi, Munif, and Bandar Alotaibi, "Distracted driver classification using deep learning", *Signal, Image and Video Processing* ,**vol.14**, no. 3 617-624,(2020).
16) Abouelnaga, Yehya Abouelnaga, Mohamed H. Saad Saad, Hesham M. Eraqi Eraqi, and Mohamed Moustafa Moustafa, "Driver Distraction Identification with an Ensemble of Convolutional Neural Networks", (2019).
17) Manu, B. N, "Facial features monitoring for real time drowsiness detection", *12th International Conference on Innovations in Information Technology (IIT),* pp. 1-4. IEEE, 2016.
18) Basubeit, O. G., D. N. T. How, Y. C. Hou, and K. S. M. Sahari,"Distracted driver detection with deep convolutional neural network", *Int. J. Recent Technol. Eng* , **vol.8,** no. 4 ,6159-6163, (2019).
19) Tamas, Vlad, and Vistrian Maties,"Real-time distracted drivers' detection using deep learning." American Journal of Artificial Intelligence , **vol.3**, no. 1, 1-8, (2019)
20) Ngxande, Mkhuseli, Jules-Raymond Tapamo, and Michael Burke, "Bias remediation in driver drowsiness detection systems using generative adversarial networks", *IEEE* Access 8 ,55592-55601, (2020).
21) Tran, D. Do HM, "Sheng W Bai H Chowdhary G," Real-time detection of distracted driving based on deep learning", *IET Intell Transp Syst* , **vol.12**, no. 10: 1210, (2018).
22) He, J., S. Roberson, B. Fields, J. Peng, S. Cielocha, and J. Coltea, "Fatigue detection using smartphones", *Journal of Ergonomics* ,**vol. 3,** no. 03 1-7, (2013).
23) Yan, Jun-Juh, Hang-Hong Kuo, Ying-Fan Lin, and Teh-Lu Liao, "Real-time driver drowsiness detection

system based on PERCLOS and grayscale image processing", *International Symposium on Computer, Consumer and Control (IS3C),* pp. 243-246. IEEE, (2016).

24) Galarza, Eddie E., Fabricio D. Egas, Franklin M. Silva, Paola M. Velasco, and Eddie D. Galarza, "Real time driver drowsiness detection based on driver's face image behavior using a system of human computer interaction implemented in a smartphone", *International Conference on Information Technology & Systems*, pp. 563-572. Springer, Cham, (2018).
25) Gu, Haisong, and Qiang Ji,"An automated face reader for fatigue detection", in 6th *IEEE International Conference on Automatic Face and Gesture Recognition*, Proceedings., pp. 111-116. IEEE, (2004).
26) Vural, Esra, Mujdat Cetin, Aytul Ercil, Gwen Littlewort, Marian Bartlett, and Javier Movellan, "Drowsy driver detection through facial movement analysis", in *International workshop on human-computer interaction*, pp. 6-18. Springer, Berlin, Heidelberg, (2007).
27) Hashemi, Maryam, Alireza Mirrashid, and Aliasghar Beheshti 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).
28) Revelo, Adriana, Robin Álvarez, and Felipe Grijalva, "Human drowsiness detection in real time, using computer vision," IEEE Fourth Ecuador Technical Chapters Meeting (ETCM), pp. 1-6. IEEE, (2019).
29) Srivastava, Himanshu, Abhishek Kaushal, Harish Kumar, Anurag Tripathi, and Anuj Kumar Sharma," A Design and Development of Baggage Sorting Robotic System at the Airport", *Evergreen*. **9** *(1)*, 86-92, (2022), doi.org:10.5109/4774219
30) Hassan, Suhaimi, Norbazlan Mohd Yusof, Mohamad Shah Ikhsan, Mohamad Zhairul Iqumal Jumari, Mohamad Amirul Mat Nadir, Mohd Hassan Helmi Ibrahim, Muhammad Azimuddin Mohd and Nor Azman, " Safety Working Environment at Highway: Safety Warning Detector (SWAD) System ", *Evergreen*. **8** *(3)*, 517-523, (2021), doi.org:10.5109/4491637
31) Shilpa Choudhary, Abhishek Sharma, Kashish Srivastava, Hemant Purohit and Mudita Vats, " Read Range Optimization of Low Frequency RFID System in Hostile Environmental Conditions by Using RSM approach", *Evergreen*. **7** (*3*), 396-403, (2020), doi.org:10.5109/4068619
32) P. Tungjiratthitikan, "Accidents in thai industry between 2001 and 2017," *Evergreen*. **5** *(2),* pp.86-92, (2018), doi.org:10.5109/1936221
33) Dief, Tarek N and Yoshida Shigeo, "System Identification for Quad-rotor Parameters Using Neural Network", *Evergreen.* **3** *(1),* pp.6-11*,* (2016), doi.org:10.5109/1657380.
34) Phan,Anh-Cang,Ngoc-Hoang-Quyen Nguyen, Thanh-Ngoan Trieu, and Thuong-Cang Phan,"An Efficient Approach for Detecting Driver Drowsiness Based on Deep Learning",*Applied Sciences* **vol.11,** (18),441,(2021),doi.org:10.3390/app11188441
35) P. P. Patel, C. L. Pavesha, S. S. Sabat and S. S. More,"Deep Learning based Driver Drowsiness Detection," International Conference on Applied Artificial Intelligence and Computing (ICAAIC), pp.380-386,(2022),doi:10.1109/ICAAIC53929.2022 .9793253.