A Minor Project Report

On

Driver Monitoring System Using Deep Learning

Submitted to

Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G.), India

In partial fulfilment

For the award of degree

of

Bachelor of Technology

in

Computer Science Engineering

By

ABHISHEK BAGHEL (301602219002)

ANCHAL NAGWANSHI (301602219007)

PRATIK KUJUR (301602219034)

Under the guidance of

Mrs. Arzoo Dixit

Asst. Prof. Department of Computer Science and Engineering

Department of Computer Science and Engineering

Government Engineering College,Raipur

Sejbahar,Raipur (C.G.)

Session 2022-2023

DECLARATION

I, the undersigned solemnly declare that the report of the thesis work entitled "**Driver Monitoring System Using Deep Learning"** is based on my own work carried out during the course of my study under the supervision of **Assistant Professor Mrs. Arzoo Dixit.**

I assert that the statements made and conclusions drawn are an outcome of the project work. I further declare that to the best of my knowledge and belief that the report does not contain any part of any work which has been submitted for the award of Bachelor of Technology degree or any other degree/diploma/certificate in this University/deemed University of India or any other country. All helps received and citations used for the preparation of the thesis have been duly acknowledged.

(Signature of the Student)	(Signature of the Student)	(Signature of the Student)
ABHISHEK BAGHEL	**ANCHAL NAGWANSHI**	**PRATIK KUJUR**
301602219002	**301602219007**	**301602219034**
BH3565	**BH3570**	**BH3597**

CERTIFICATE OF THE SUPERVISOR

This is to certify that the work incorporated in the project **"Driver Monitoring System Using Deep Learning"**, is a record of project work carried out by **Abhishek Baghel (301602219002), Anchal Nagwanshi (301602219007),** and **Pratik Kujur (301602219034)** under my guidance and supervision for the award of Degree of Bachelor of Technology in Computer Science and Engineering of Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G.), India.

To the best of my knowledge and belief the project

1. Embodies the work of the candidate herself/himself.
2. Has duly been completed.
3. Fulfills the requirement of the Ordinance relating to the Bachelor of Technology Degree of the University, and
4. Is up to the desired standard growth in respect of contents and language for being referred to the examiners.

(Project Guide)
Asst. Prof. Mrs. Arzoo Dixit
Dept. Of Computer Science & Engineering,
Government Engineering College,
Sejbahar, Raipur (C.G.)

(Project Coordinator)
Asst. Prof. Pushpendra Dwivedi
Dept. Of Computer Science & Engineering,
Government Engineering College,
Sejbahar, Raipur (C.G.)

(H.O.D)
Prof. Dr. R.H.Talwekar
Dept. Of Computer Science & Engineering,
Government Engineering College,
Sejbahar, Raipur (C.G.)

Forwarded to Chhattisgarh Swami Vivekanand Technical University, Bhilai

CERTIFICATE BY THE EXAMINERS

This thesis entitled **"Driver Monitoring System Using Deep Learning"**, submitted by **Abhishek Baghel (301602219002), Anchal Nagwanshi (301602219007),** and **Pratik Kujur (301602219034)** and has been examined by the undersigned as a part of the examination and is hereby recommended for the award of the Degree of Bachelor of Technology in Computer Science and Engineering of Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G.), India.

Internal Examiner	**External Examiner**
Date:23/01/2023	**Date:23/01/2023**

ACKNOWLEDGEMENT

\We would like to thank my guide **Asst. Prof. Mrs. Arzoo Dixit**, Department of Computer Science and Engineering for her immense support and enlightened guidance for my project which we have developed as a B.Tech student. We are very grateful for the inspiring discussions with all my faculties. Their valuable support and path-guiding suggestions have helped me to develop this project. We are thankful to **Prof. Dr. R.H. Talwekar**, HOD (CSE), for giving thoughtful suggestions during my work and for their inspiration and constant encouragement that enabled me to present my work in this form. We would like to take this opportunity to express my thanks to everyone in the Computer Laboratory of Government Engineering College, Sejbahar, Raipur (C.G.).

We especially thank my colleagues, batchmates, and friends for their support and help that they offered me during the development of this project.

(Signature of the Student)	(Signature of the Student)	(Signature of the Student)
ABHISHEK BAGHEL	**ANCHAL NAGWANSHI**	**PRATIK KUJUR**
301602219002	**301602219007**	**301602219034**
BH3565	**BH3570**	**BH3597**

ABSTRACT

Fatigue reduces the driving attention of the driver. Due to drowsiness the driver may make serious driving action error that causes fatal road accidents. Accidents due to drowsiness could be avoided if the eye closure and yawning activity of driver could be detected in real time and notification could be delivered in the form of sound that would help in regaining the attention level of the driver.

In Driver Monitoring System we chose eye blinking pattern and yawning as two major factors to estimate the drowsiness state. We proposed an approach in which we take the images of driver in real time using the camera to capture the facial region and predict the facial landmark coordinates using deep learning model. Out of the facial region eye and mouth is taken as ROI (Region of Interest) to measure the percentage of eye closure and mouth openness state during yawning.

Eye percentage closure is calculated by measuring the distance between the upper lid and lower lids of the eye using eye region landmarks and similarly for yawning the distance between the upper lips and lower lips is calculated using mouth region landmarks.

We proposed two threshold metric EAR (Eye Aspect Ratio) and MAR (Mouth Aspect Ratio) that would measure the openness of the eye and mouth.

TABLE OF CONTENTS

❖

.

LIST OF FIGURES

LIST OF ABBREVIATIONS

(1) MAR - Mouth Aspect Ratio

(2) EAR - Eye Aspect Ratio

(3) CNN - Convolutional Neural Networks

(4) (LaPa) - Landmark guided face Parsing

Chapter-1

INTRODUCTION

With increase in transportation industries, road accidents figures have also increased. There are various factors that causes road accidents. Drowsiness is one major reason to road accidents. According to National Crime Records Bureau (NCRB), fatigue or sleepiness leads to around 4% of road accidents in India. National Highway Traffic Safety Administration (NHTSA), estimated 100,000 police-reported crashes each year in the United States, resulting in an estimated 1,550 deaths and 40,000 injuries due to drowsy driving.

There are several factors that lead to drowsy driving, including a lack of sleep, long working hours or night shifts or certain medications that induces sleep. Drowsiness impairs a driver's ability to make safe and effective driving decisions. When a driver is drowsy, they may have difficulty staying alert, maintaining their focus, and reacting quickly to changing road conditions. This can lead to an increased risk of accidents, including collisions and lane departures.

The need for drowsiness detection arises from the fact that drowsiness can be a major cause of accidents, particularly in situations where a person is operating a vehicle. Drowsy driving is a serious problem, as it can lead to a lack of attention, slower reaction times, and poor decision-making, all of which can increase the risk of accidents.

By providing an early warning of drowsiness, drowsiness detection systems can help to prevent accidents and improve safety by alerting the driver to take a break or pull over before they become too drowsy to drive safely. It can also have application in industrial settings, where it can help to prevent accidents by detecting when an operator is becoming drowsy and taking appropriate action, such as slowing down or stopping the machinery.

Few of the automobile car brands provides drowsy detection features that indirectly detects the drowsy state by measuring the steering angle, braking and vehicle speed pattern some system uses infrared camera mounted in top center of the windscreen to measure the head movement.

There are several types of drowsiness detection systems that can be used in cars, including:

- Video-based systems: which use cameras to track the movement of the eyes, face, and head, and use changes in these movements as an indicator of drowsiness.

- Physiological-based systems: which use sensors to measure changes in heart rate, respiration rate, and skin conductance as an indicator of drowsiness.

- EEG-based systems: which use electroencephalography (EEG) to measure brain activity, and use changes in the EEG signal as an indicator of drowsiness.

In this project we present a hypothesis where we take driver facial image as input and using deep learning model facial landmark prediction is done and EAR and MAR are calculated if either of its value is in the threshold value range, it will be the drowsy state and an alert notification will be sent to the driver.

Chapter-2

LITERATURE REVIEW

[1] Predicting drowsiness accidents from personal attributes, eye blinks and ongoing driving behavior

This paper states there are various reason for drowsiness which leads to poor performance in driving. People gradually perform more poorly on tasks performed for extended periods of time at night and following loss or disturbance of sleep. Depending on the assumed causes, several terms have been used to refer to the cause of this performance deterioration. Poor performance has been attributed to drowsiness, sleepiness, fatigue and inattentiveness. Some have argued that drowsiness is a consequence of fatigue. Others noted that monotony may also cause drowsiness. The best predicting measures for poor driving were the frequency of eye-closures exceeding 1 s and the number of times that time-to-line crossings were below 0.5 s

[2] Detecting driver drowsiness based on sensors: A Review

Driver drowsiness detection is a technology used to detect when a driver is becoming fatigued or drowsy while driving. There are several different methods that can be used to detect driver drowsiness, including:

1. Vehicle-Based Measures:

   1.1 Steering Wheel Movement (SWM): This technology uses sensors to track the movement of the steering wheel, and can detect when the driver's steering becomes less precise, indicating that they are becoming drowsy. This technology uses cameras or sensors to track the movement of the driver's eyes, and can detect when the driver's eyelids are drooping or closing.

   1.2 Standard Deviation of Lane Position (SDLP): This technology uses cameras to detect when the vehicle is drifting out of its lane, and can alert the driver if it detects that the driver is drowsy.

2. Behavioral Measures: Behavioral measures of drowsiness are methods used to detect signs of drowsiness in a person's behavior, rather than through physiological means. Some examples of behavioral measures of drowsiness include

   2.1 Yawning: This is a classic sign of drowsiness and is often used as an indicator of fatigue.

   2.2 Head nodding: This is when a person's head droops or nods, usually indicating that they are about to fall asleep.

2.3 Eye closure: This is when the eyes of a driver start to close or blink more frequently, indicating that they are drowsy.

3. Physiological Measures: Physiological measures for drowsiness are methods used to detect signs of drowsiness by measuring changes in the body's physiology. Some examples of physiological measures of drowsiness include:

   3.1 Electroencephalography (EEG): This measures the electrical activity in the brain and can be used to detect changes in brain waves that indicate drowsiness.
   3.2 Electrooculography (EOG): This measures the electrical activity in the eye muscles, and can be used to detect changes in the patterns of eye movement that indicate drowsiness.
   3.3 Electrocardiography (ECG): This measures the electrical activity of the heart, and can be used to detect changes in heart rate that indicate drowsiness.
   3.4 Electromyography (EMG): This measures the electrical activity in muscles, and can be used to detect changes in muscle tone that indicate drowsiness.

These systems can be used separately or together, to give a more accurate detection.

[3] "Eye Aspect Ratio for Real-Time Drowsiness Detection to Improve Driver Safety."

The Eye Aspect Ratio (EAR) is a measure used in computer vision to detect drowsiness based on the facial features of a person. It is calculated by taking the ratio of the distances between the inner and outer corners of the eye, and the distance between the upper and lower eyelids. When a person is awake or blinks the ratio changes.

The EAR is a simple and effective measure for detecting drowsiness, as it only requires a single camera and can be calculated in real-time.

It can be affected by various factors such as lighting conditions, glasses, facial hair and even the person's facial expressions. It should be combined with other measures such as head nodding or slow reaction time, to increase the accuracy of drowsiness detection.

[4] Real Time Driver Fatigue Detection Based on SVM Algorithm

Mouth Aspect Ratio (MAR) is a metric to detect drowsiness or yawning based on the position of lips. It is calculated by taking the ratio of the distances between the inner and outer corners of the mouth, and the distance between the upper and lower lips. When a person yawns it ratio changes

due to the change in the distance between the lips and corners.
The MAR can be used in conjunction with the Eye Aspect Ratio (EAR) as an additional measure to detect drowsiness, yawning or fatigue.

[5] Facial Landmark Detection: A Literature Survey

Facial landmark detection algorithms into three major categories: holistic methods, Constrained Local Model (CLM) methods, and the regression-based methods. They differ in the ways to utilize the facial appearance and shape information. The holistic methods explicitly build models to represent the global facial appearance and shape information. The CLMs explicitly leverage the global shape model but build the local appearance models. The regression-based methods implicitly capture facial shape and appearance information.
Deep learning-based methods typically use convolutional neural networks (CNNs) to learn the facial landmarks from a large dataset of labeled images. One popular deep learning-based facial landmark detection method is the Multi-task Cascaded Convolutional Networks (MTCNN) algorithm. This algorithm uses three cascaded CNNs to detect and align faces, and then uses another CNN to detect facial landmarks.

Chapter-3

METHODOLOGY

3.1 Software and libraries requirements-

3.1.1 Python-

Python is a high-level, interpreted programming language that is widely used for web development, scientific computing, data analysis, artificial intelligence, and other applications. It was first released in 1991 by Guido van Rossum and has since grown to become one of the most popular programming languages in the world.

One of the main advantages of Python is its simplicity and readability, making it easy for new programmers to learn and for existing code to be maintained. It also has a large and active community which has developed a wide range of libraries and frameworks for specific tasks such as NumPy for numerical computations, pandas for data manipulation, and TensorFlow for machine learning.

Python also supports multiple programming paradigms such as object-oriented, functional, and procedural programming. It runs on various platforms such as Windows, Mac, Linux, and Unix. It's widely used in various fields such as Data Science, Web development, Machine Learning, Automation, and many more.

3.1.2 OpenCV-

OpenCV (Open-Source Computer Vision) is a library of programming functions mainly aimed at real-time computer vision. It was developed by Intel and is now maintained by a non-profit organization. OpenCV is written in C++ and has interfaces for C++, Python, and Java.

OpenCV provides a wide range of functionality for image and video processing, including:

- Image processing and manipulation, such as filtering, thresholding, and colour space conversions.
- Object detection and recognition, such as Haar cascades, HOG, and deep learning-based object detection.
- Camera calibration and 3D reconstruction.

OpenCV also provides a number of pre-trained deep learning models, including MobileNet, ResNet , which can be used for various computer vision tasks such as object detection and semantic segmentation.

3.1.3 TensorFlow and Keras-

TensorFlow is an open-source software library for dataflow and differentiable programming across a range of tasks. It is a powerful library for numerical computation, particularly well

suited and fine-tuned for large-scale Machine Learning and Deep Learning. TensorFlow allows developers to create data flow graphs—structures that describe how data moves through a model, and how the model should change to improve its performance. It also provides tools for deploying machine learning models on a variety of platforms, including web and mobile.
Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow. It was developed with a focus on enabling fast experimentation. It provides a user-friendly API which makes it easy to build and train models, while still allowing for customization through the use of TensorFlow. Keras abstracts the complexities of building a deep neural network and allows users to focus on developing and evaluating their models. It is widely used for quickly prototyping, building, and training deep learning models, and it's a popular choice for researchers and practitioners in the field of deep learning.

3.1.4 MobileNetV2-

MobileNetV2 is a lightweight neural network architecture for image classification and object detection tasks, designed to run on mobile and embedded devices. It was developed by Google and introduced in the paper "MobileNetV2: Inverted Residuals and Linear Bottlenecks" in 2018. MobileNetV2 is a variation of the original MobileNet architecture and is designed to be more efficient in terms of computational resources while maintaining high accuracy. It uses depth wise separable convolutions, which are more efficient than standard convolutions, to reduce the number of parameters and computation needed. Additionally, it uses a technique called "inverted residuals" which allows for the expansion of the feature depth while keeping the number of parameters small.
MobileNetV2 is trained on the ImageNet dataset and is able to achieve high accuracy on this dataset while having a small number of parameters and low computational cost. This makes it a good choice for applications such as object detection, image classification, and facial recognition on mobile and embedded devices with limited computational resources.

3.2 Hardware requirements

3.2.1 Camera-

It can be seen that the camera is used for monitoring the driver’s face continuously and upon detection of drowsiness or fatigue, the system in the dashboard generates a voice alert type warning to the driver.

3.2.2 Speaker-

We are using a speaker that emits an alarm sound or vibration to wake a person up if they begin to fall asleep.

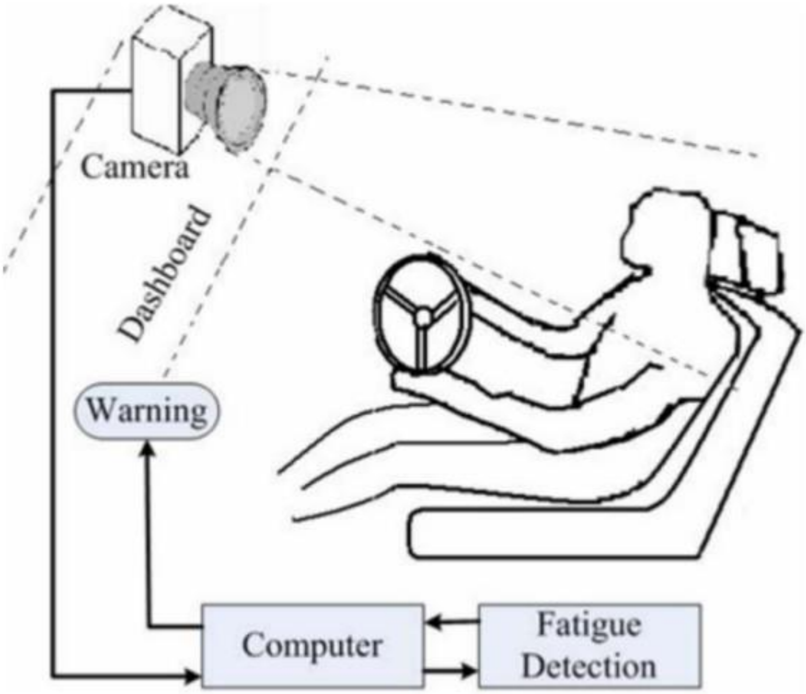

Fig - 3.1 System Design

3.3 Flow chart

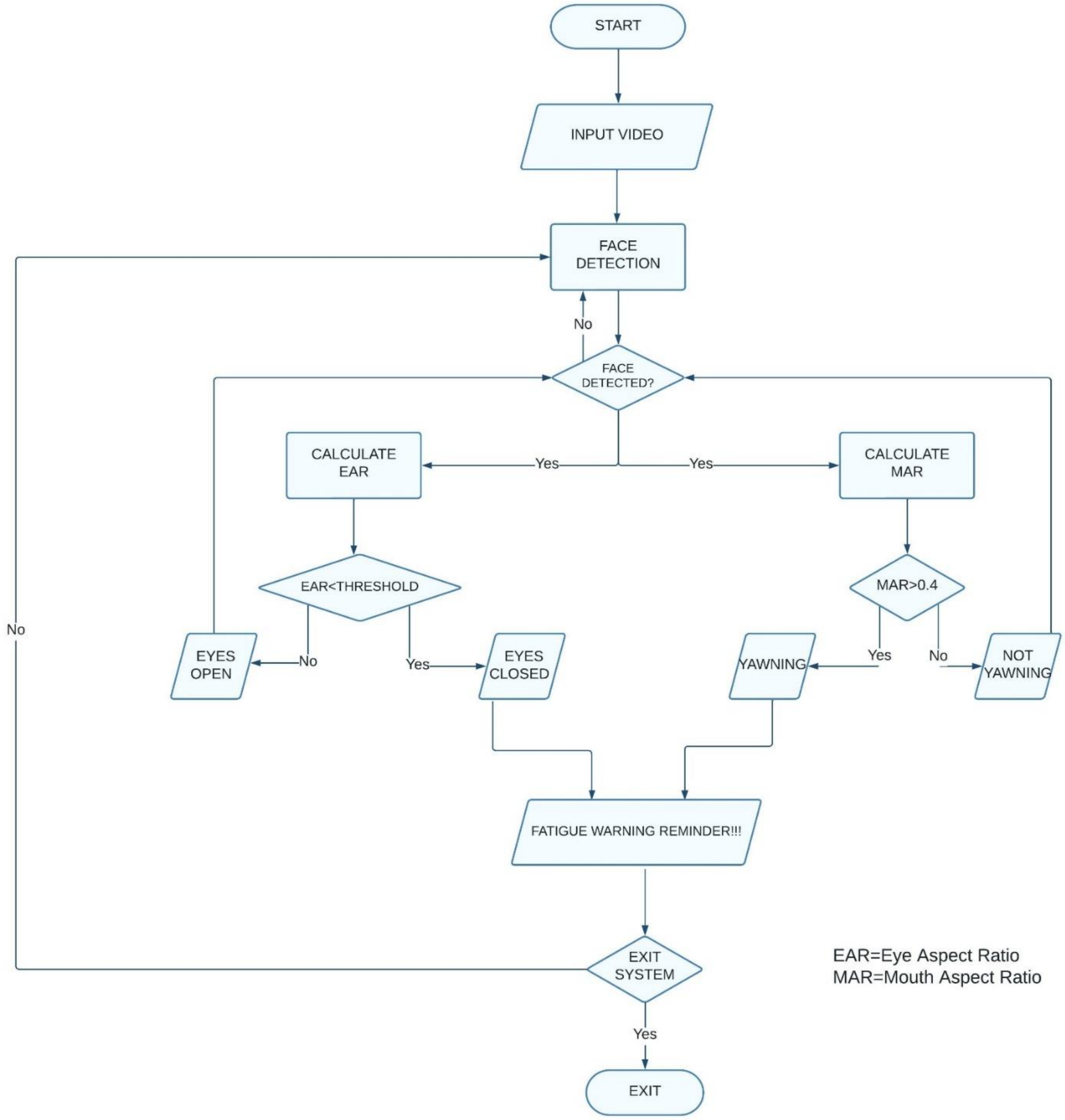

Fig – 3.2 Flow Chart

3.4 Methodology

3.4.1 Input video-

The system has been designed such that the face and hence the eyes and mouth of the driver are always monitored and if the predefined levels of alertness are observed to be defaulted and compromised, then an appropriate alarm is set off, and accordingly, action is taken to prevent any fatalities. Depicts the System Design of Driver Drowsiness and Yawn Detection System. It can be seen that the camera is used for monitoring the driver’s face continuously and upon detection of drowsiness or fatigue, the system in the dashboard generates a voice alert type warning to the driver.

3.4.2 Preprocessing-

1. Normalization: Resizing (each image is resized to 256 x 256) is done to normalize the image data so that it has a consistent scale and can be more easily processed by the machine learning algorithm.
2. Dimensionality reduction: Preprocessing is done to reduce the dimensionality of the image data, which helped us to speed up the training process of our model.

3.4.3 Face detection-

We employ MobilenetV2 as the backbone. In order to reduce the resolution loss caused by pooling or convolution with stride larger than 1, dilation convolution is adopted in the fifth residual block, therefore the resolution of the output is 1/16 rather than 1/32 of the input. In order to leverage the global texture information, pyramid spatial pooling with different scales are used before the classifier learning. Then, the output feature maps with different resolutions are concatenated with high-resolution feature maps generated by the last residual block after interpolation to the same size. The integrated feature maps are used to predict the semantic label for each pixel. Boundary pixels are important but challenging to be predicted. In order to improve the segmentation performance for boundary pixels, a boundary aware branch is added to learn boundary features by a boundary detection task. First, it extracts shared features from different layers of MobilenetV2 in the semantic branch, and then projects them into a new space where boundary details are well preserved. The output of this branch is a boundary map in which each value refers to the confidence score that pixel is located on the boundary without considering semantics. We then use this boundary map and landmarks to train our model.

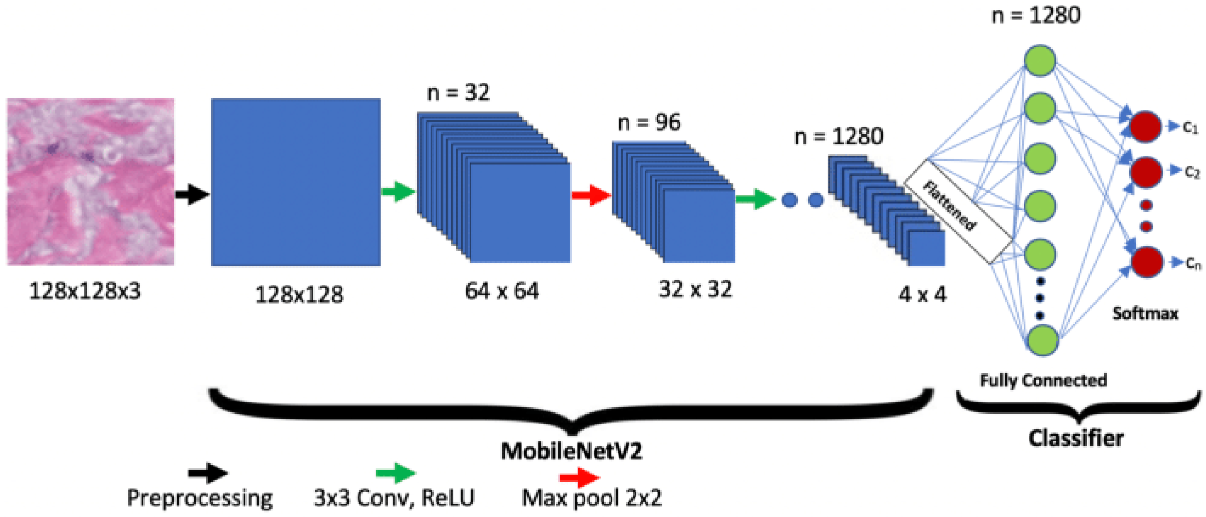

Fig - 3.3 MobileNet Architecture

3.4.4 Calculate MAR and EAR –

The mouth aspect ratio (MAR) is a metric used in facial analysis to determine the degree of mouth openness in an image. It is calculated by dividing the distance between the upper and lower lip by the distance between the two outer corners of the mouth. A higher MAR indicates a greater degree of mouth openness. This metric is commonly used in facial recognition systems, such as yawn detection, and it is used to detect the drowsiness of a driver or the level of interest of a person in video conferencing.

$$MAR = \frac{|(P_{99} - P_{103})| + |(P_{100} - P_{102})| + |(P_{98} - P_{104})|}{3|(P_{101} - P_{97})|}$$

(Refer Fig - 3.4 Facial Landmarks)

The eye aspect ratio (EAR) is a metric used in facial analysis to determine the degree of eye closure in an image. It is calculated by dividing the distance between the inner corners of the eyes (i.e. the distance between the two pupils) by the distance between the outer corners of the eyes (i.e. the distance between the two outermost points of the eyes). A lower EAR indicates a greater degree of eye closure. This metric is commonly used in facial recognition systems, such as blink detection, and it is used to detect the drowsiness of a driver or the level of attention of a person in video conferencing. It's also used to track the eyes movement and gaze estimation. Both MAR and EAR are commonly used in facial recognition systems, such as blink detection and yawn detection.

$$\text{EAR FOR LEFT EYE} = \frac{|(P_{69}-P_{73})| + |(P_{70}-P_{72})| + |(P_{68}-P_{74})|}{3|(P_{67}-P_{71})|}$$

$$\text{EAR FOR RIGHT EYE} = \frac{|(P_{78}-P_{82})| + |(P_{79}-P_{81})| + |(P_{77}-P_{83})|}{3|(P_{80}-P_{76})|}$$

(Refer Fig - 3.4 Facial Landmarks)

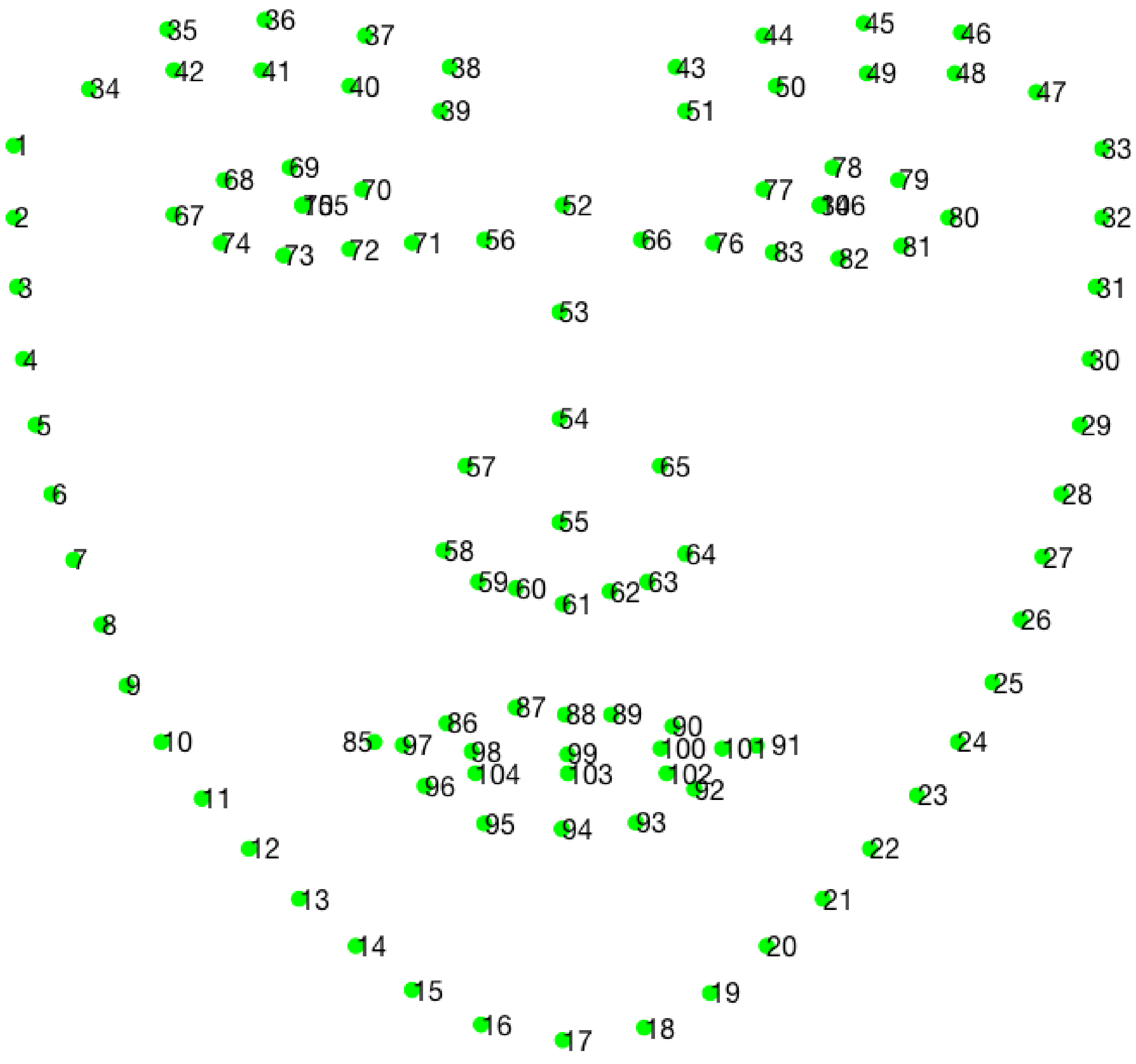

Fig - 3.4 Facial Landmarks

3.4.5 Face warning reminder-

Our project deals with the EAR and MAR function which computes the ratio of distances between the horizontal and vertical eye landmarks. A voice module is also deployed which is used for giving appropriate voice alerts when the driver is feeling drowsy or is yawning.

3.5 Dataset

The Landmark-Guided Face Parsing dataset (LFPD) is a dataset for evaluating face parsing algorithms. It includes a set of images of faces, along with corresponding annotation data that labels the facial regions (e.g., eyes, nose, mouth, etc.) and facial landmarks (e.g., corners of the eyes, tip of the nose, etc.). The dataset also includes information about the poses and expressions of the faces in the images.

The dataset is designed to be challenging, with images taken under various conditions such as different lighting, facial expressions, and occlusions, making it suitable for testing the robustness of face parsing algorithms. The annotation includes both pixel-level and landmark-level information, which can be used to evaluate different types of face parsing algorithms. The dataset is intended to be used as a benchmark for comparing the performance of different face parsing algorithms, and to help in the development of new algorithms.

Landmark guided face Parsing (LaPa) dataset consists of more than 22,000 images, covering large variations in facial expression, pose and occlusion. Each image coordinates of 106-point landmarks.

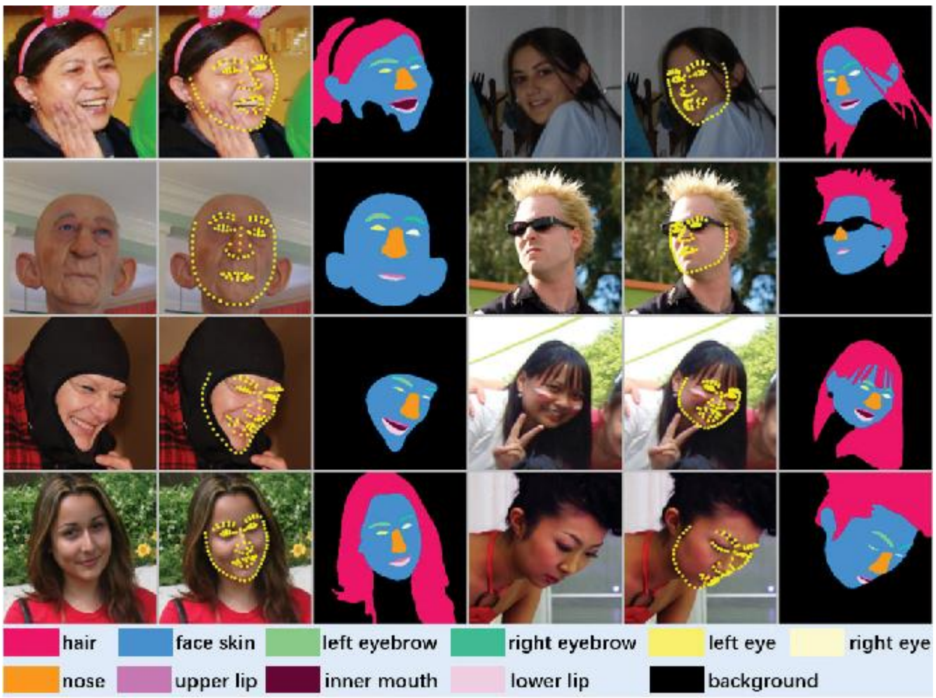

Fig – 3.5 Lapa Dataset

Chapter-4

RESULT

In our model we have accomplished yawning detection by using MAR (Mouth Aspect Ratio) as our threshold but we have failed to accomplished eyes blink pattern detection.
As we know dataset plays an important role in building a model, because of insufficient variation in movements of eyes in our dataset images, the boundary lines of eyes are constant, due to which there is no movements in coordinates of eyes when we give real-time input to our model. Now the EAR (Eyes Aspect Ratio) threshold cannot be calculated as it depends on the coordinates of eyes.

MAR Threshold calculation-

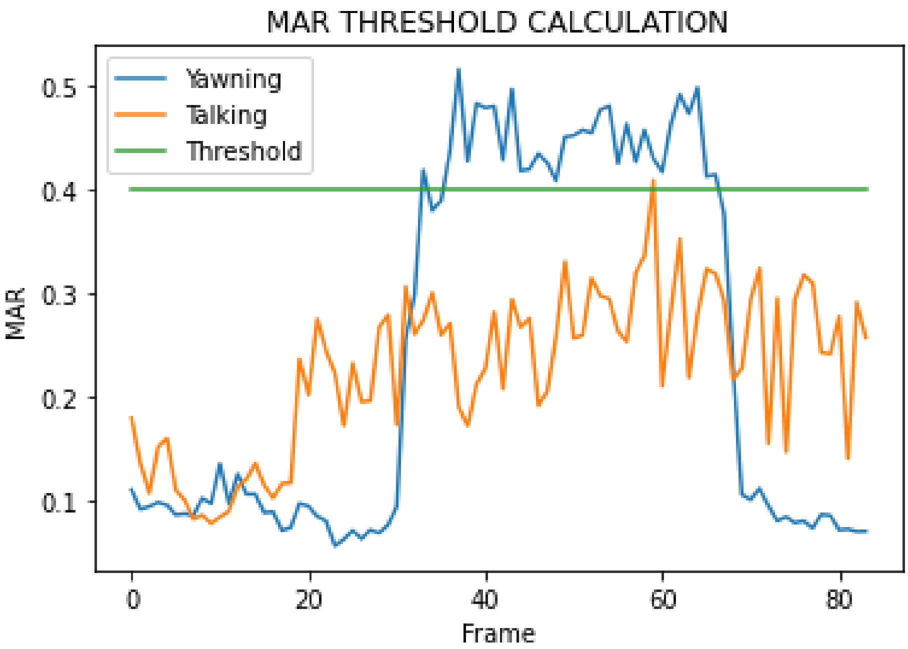

Fig – 4.1 MAR Calculation

MAR threshold could be different for different for models. For our model we have plotted a multiline graph
between image frame(x-axis) vs MAR(y-axis). We have plotted graph for two cases –
Case 1-calculating threshold while talking-
We have done this case so that our model does not get confuse between talking and yawning. We can observe that the peak MAR when we talk is 0.4.
Case 2-calculating threshold while yawning-
We have done this case to get optimal MAR threshold for our model. We can observe that there is constant MAR rise after 0.4.
Hence, we have taken 0.4 as our MAR threshold.

Final result-

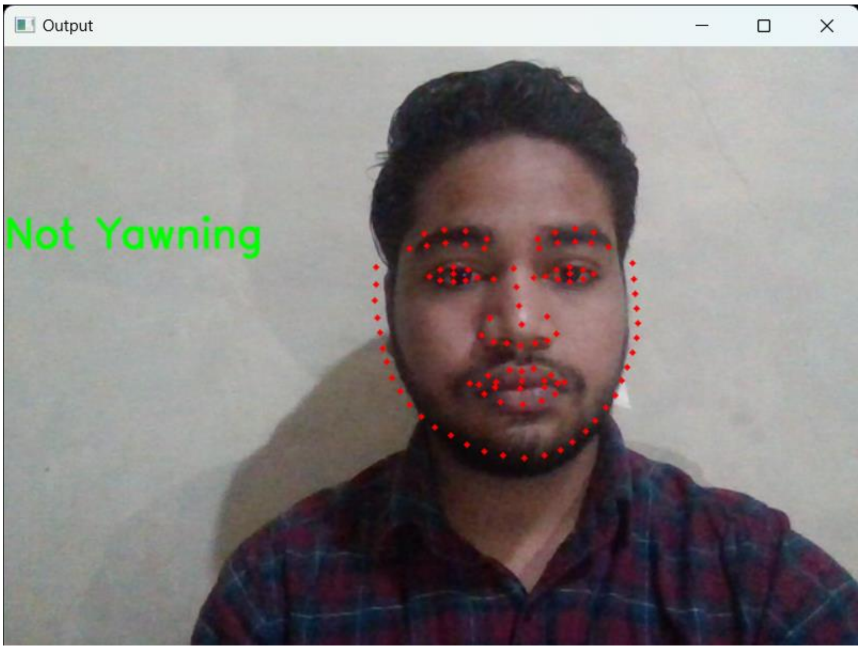

Fig - 4.2 Not Yawning

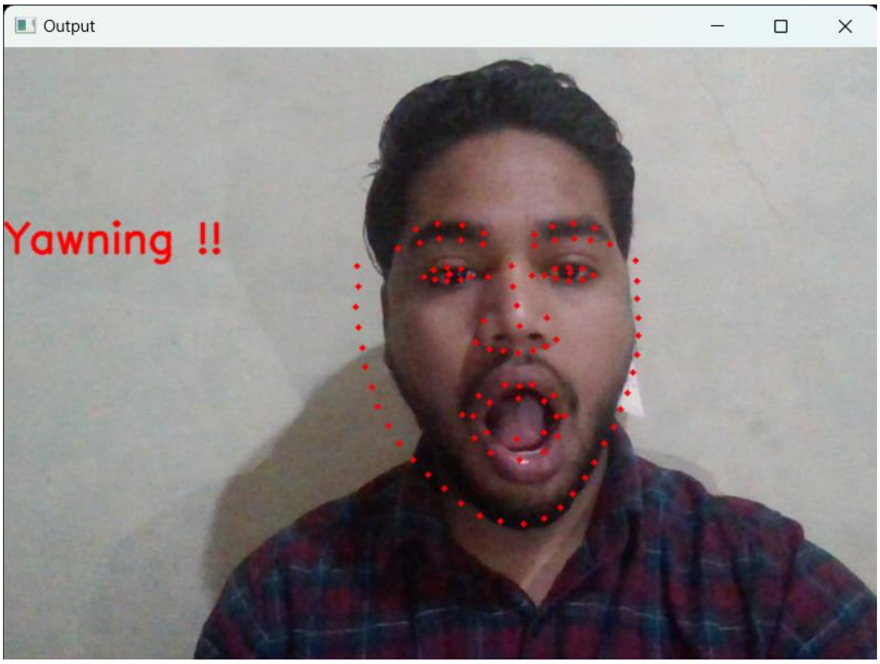

Fig - 4.3 Yawning

Chapter-5

CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

In conclusion, a drowsiness detection system is a tool that can be used to detect signs of drowsiness in individuals and alert them or take appropriate actions to prevent accidents or errors caused by fatigue. These systems can be based on various methods such as monitoring eye movement and blink rate, analyzing facial expressions, and measuring physiological signals. However, it is important to note that the detection methods are not always accurate and further research is needed to improve their reliability. Additionally, it's important to consider the ethical and privacy implications of using these detection methods in different settings.

5.2 Future scope

The future scope of drowsiness detection systems is promising as there is a growing need for such systems in various industries to improve safety and productivity. Some potential areas of application include:

1. Healthcare: Drowsiness detection systems can be used in hospitals to monitor patients who are at risk of falling asleep during procedures or treatments.

2. Industrial: Drowsiness detection systems can be used in industrial settings such as factories, where workers may be at risk of falling asleep while operating heavy machinery.

3. Sports: Drowsiness detection systems can be used to monitor athletes during training and competition to detect signs of fatigue and prevent injuries.

With the growing use of technology, the advancements in AI and Machine learning techniques, the future of drowsiness detection systems is expected to improve in terms of accuracy and reliability.

Chapter-6

REFERENCES

1. Willem B. Verweya, David M. Zaidelb: "***Predicting drowsiness accidents from personal attributes, eye blinks and ongoing driving behavior***", Volume 28, Issue 1, 1 January 2000, Pages 123-142

2. Arun Sahayadhas *, Kenneth Sundaraj and Murugappan Murugappan: "**Detecting Driver Drowsiness Based on Sensors: A Review** ", Sensors 2012, 12, 16937-16953; doi:10.3390/s121216937

3. Christine Dewi, Rung-Ching Chen, Chun-Wei Chang, Shin-Hung Wu, Xiaoyi Jiang Hui Yu: "**Eye Aspect Ratio for Real-Time Drowsiness Detection to Improve Driver Safety**" , Volume 11, Issue 19, 30 September 2022

4. Burcu Kir Savas, Yasar Becerikli: "**Real Time Driver Fatigue Detection Based on SVM Algorithm**", 2018 6th International Conference on Control Engineering & Information Technology (CEIT)

5. Yue Wu, Qiang Ji: "**Facial Landmark Detection: A Literature Survey**", International Journal on Computer Vision, 2017