

A Multistage Adaptive 3D-CNN Framework for Driver Drowsiness Detection Using Multi-Condition Facial Behavior Analysis

Nrusimhadri Naveen¹, Dr. Aarti²

¹PhD scholar, Department of Computer Science and Engineering, Shri Venkateshwara university, Gajraula, UP.

²Department of Computer Science and Engineering, Shri Venkateshwara university, Gajraula, UP.

²Assistant Professor, Department of Computer Science and Engineering, Bharati Vidyapeeth college of Engineering, New Delhi.

Abstract:

The need for improved detection technology to address this worldwide problem is highlighted by an increase in driver fatigue-related incidents. Many machine learning and deep learning techniques, such as CNN variations, have been tested by researchers; nevertheless, real-time deployment is challenging due to computational complexity, limited evaluation accuracy, and questionable reliability. Driver drowsiness detection using a multistage adaptive 3D-CNN model that takes system performance and complexity into account. Driver fatigue symptoms are categorised using our models. Several 3D CNN architectures are used to gather face attributes for sub-models. The head condition model detects sideways glancing, eyebrow elevation, and head snoozing. Eye closures, pain, and partially closed eyes are all detected by the eye alignment prototype. Open and covered yawning are recognised using the mouth condition model. States are categorised as day, sunglasses, night, and normal by the specs and typical conditions model. Each model, which uses specialised algorithms to identify different driver sleepiness variables, is crucial to the system. thoroughly investigated utilising the KEC-DDD and NTHU-DDD datasets, which demonstrate their resilience and flexibility. The sub-models' ability to effectively capture driver behaviours in a variety of settings is confirmed by dataset validation results.

Keywords: Drowsiness; Deep Learning; Convolutional Neural Network; Accuracy; Detection.

I. INTRODUCTION

Highway protection is significantly impacted by sleepy lashing. Drowsiness is mostly an underappreciated danger factor on the roads, regardless of the situation's potential consequences. While the National Highway Authority of India (NHAI) reports about 450,000 crashes in India that result in countless injuries and fatalities, the National Highway Traffic Safety Administration (NHTSA) believes that sleepy heavy causes about 100,000 collisions annually in the United States. Tens of billions of dollars are thought to be lost each year as a result of drowsy driving accidents [1].

Sleepy driving has serious social and general repercussions in addition to the immediate costs. Every coincidence is a disruption to families, communities, and workplaces. The need for effective detection and intervention techniques is highlighted by the preventable landscape of sleepy driving events. Proactively treating sleepiness can prevent many deaths, protect livelihoods, and promote a responsible driving culture.

Despite the growing recognition of driver fatigue, researchers and engineers face unique challenges in identifying it [2]. Like other impairments like drunkenness or attention, drowsiness lacks external symptoms, making it difficult to diagnose using conventional techniques. Additionally, there is a great deal of variation in each person's sensitivity to tiredness, which is influenced by things like sleep quality and underlying medical conditions.

Therefore, considering the complexity of drowsiness, developing a universal detection method is justified. It is more difficult to recognise fatigue in India due to the dynamic nature of driving conditions. Road surfaces, traffic patterns, and lighting variations all have a big impact on how drivers behave, necessitating sophisticated detecting systems that can adjust to different situations. Furthermore, real-time processing needs place strict constraints on latency and computational efficiency, requiring optimised architectures that balance speed and accuracy [3].

In the past, drowsiness detection systems only gave a general picture of the driver's alertness; they did not record specific behaviours like eye and mouth states that affect the overall state. This is a first indicator of tiredness. Customised sub-models must be developed and put into use in order to overcome this limitation. Sub-models focus on different aspects of driver behaviour, enabling a deeper understanding of the driver's situation. State understanding sub-models are employed by employing 3D-CNN architectures that are tailored to different features of driving behaviour and scenarios. In order to examine how the head, eyes, mouth, and circumstances with or without glasses move, this work use specialised sub-models [4]. This device was designed to monitor and categorise head and facial movements indicative of sleepiness. The head and glasses/normal condition models have the same design and characteristics, despite the fact that each sub-model has its own architecture [5]. For drivers, it produces a drowsiness condition that falls midway in the middle of the range.

The objective of the paper is listed as,

- Our models categorize indications of driver weariness. Facial traits for the head, eyes, mouth, and glasses/normal sub-models are acquired via various 3D CNN architectures.
- Models are integral to the system, identifying specific elements of driver drowsiness through specialized algorithms.
- The presentation of these sub-models was thoroughly assessed and illustrated towards robustness and adaptability to diverse real-world contexts.

The paper is organized as follows: the section 2 is discussed with literature survey and the proposed models based on the sub models.

II. RELATED WORKS

Finding faces in pictures is part of face placement in computer vision. This is essential for AR, driver drowsiness detection, and facial recognition. The accuracy and durability of face location have been improved by recently discovered deep learning. The goal of multi-task learning frameworks is to improve performance by optimising related tasks concurrently. By improving feature extraction in crucial areas, focused attention procedures increase the accuracy of face recognition and localisation. Methods that handle variances and occlusions in datasets are required for reliable results in real-world scenarios, including a wide range of positions, illumination, and visibility levels. When combined, these strategies improve face positioning in constrained areas [6].

To enhance detection in different positions and occlusions, the [7] approach combines a ConvNet with a 3D face model. This all-inclusive framework achieves exceptional face detection in challenging environments by combining ConvNets and 3D models. The model can efficiently handle variations in position and sealings by incorporating 2D latitudinal evidence through 3D symmetrical statistics. Nevertheless, combining 2D and

3D data could make the model more complex, requiring sophisticated processing and training techniques [8].

Deep Face Detection with Mask R-CNN [9] accurately locates faces in chaotic backgrounds by utilising the framework's instance segmentation capabilities. This version is frequently more effective than traditional methods of identifying faces, and it performs particularly well in challenging circumstances. Segmenting faces in complex settings increases the accuracy and durability of the model. While this approach improves detection, it might need additional modules or modifications for tasks like posture estimation or landmark localisation. When comparing frameworks, multi-task learning, attention processes, and methods for managing variances and occlusions make facial analysis tasks far more accurate and robust. Enhance facial analysis using a variety of techniques, each with advantages and disadvantages. Attention mechanisms improve accuracy, multi-task learning frameworks optimise many tasks simultaneously [10], and specialised models resolve occlusions and variances to produce effective facial analysis systems [11].

Efficiency and flexibility are essential for driver tiredness detection systems, particularly for real-time operation on embedded sensors in cars. LightCNN may be used to detect tiredness in real time because of its lightweight architecture, which incorporates maxout activation functions and lightweight convolutional layers that speed up calculations [12]. For mobile and embedded devices,

Face Feature Net's depthwise separable convolutions offer efficient face feature extraction at a low computational cost [13]. According to [14], Dlib's facial recognition module, which employs deep metric learning, provides discriminative features and is simple to integrate into many systems, making it appropriate for real-time drowsiness detection. For automotive systems with limited resources that require real-time performance, the compact CNN architecture strikes a balance between accuracy and economy [15]. These models highlight the effectiveness and adaptability of computers in detecting driver fatigue. Road safety is increased by technologies that can monitor drivers and identify fatigue through efficient feature extraction and real-time processing. Because they can adapt in real time, some computer networks—such as Light CNN, Face Feature Net, Dlib, and Compact CNN—are quick and adaptable enough to be utilised in automotive systems. These technologies enhance driver tiredness detection systems, resulting in safer driving situations.

Driver sleepiness detection systems must be flexible and customisable to accommodate different drivers and driving situations in order to increase its accuracy and dependability. [16] provide an appropriate approach that uses machine learning techniques to modify discovery limitations based on each driver's behaviour. The prototype may adjust detection thresholds for each driver's unique drowsiness patterns by assessing historical data. By addressing individual differences in drowsiness markers, our customised method improves detection accuracy.

[17] and [18] highlight how important it is to use historical data to increase drivers' recognition limits. The technique makes it possible for the scheme to adjust to variations in driver behaviour and circumstances, hence enhancing performance over time. The prototype can provide more precise and reasonable warnings by using machine learning algorithms to customise recognition limitations, making the roads safer overall [19]. [20–24] look at using real-time feedback systems to modify detection limits according to driver responses and surrounding conditions. As driving conditions change, the detection system constantly adjusts to preserve accuracy and dependability. The model may change detection parameters based on real-time feedback, producing driver alerts that are more efficient and timely. These investigations differ from previous ones that

employed more broad techniques, such as the fixed threshold techniques employed by conventional CNN models. This study highlights the need to take individual drowsiness patterns into account by emphasising personalisation and adaptation. By examining driver conduct and environmental conditions, adaptive systems can identify fatigue and provide timely alerts, enhancing road safety [25] and [26].

Advanced driver drowsiness detection systems use a variety of technologies to assess signs of exhaustion, such as facial expressions, eye movements, and bodily data [27] and [28]. Accuracy and real-time detection are greatly enhanced by advancements in machine learning, particularly deep learning. CNN, LSTM networks, and hybrid models exhibit promise for fatigue detection. Pictorial, physical, and vehicular information are combined in multi-modal strategies to increase recognition accuracy. Sleepiness differences can be successfully addressed by adaptive techniques that modify detection thresholds based on driver behaviour. This review explores the latest advancements in driver tiredness detection, highlighting significant strategies, technological innovations, and possible directions for further research to improve the efficiency and reliability of these crucial safety devices.

III. PROPOSED METHODOLOGY

As illustrated in Fig. 1, the proposed process for knowing state understanding sub-models starts with the input of images with face positioning. The Head Condition Model analyses head movements to identify behaviours like

- Nodding,
- Elevating eyebrows, and
- Lateral gazing.

These movements may signify disturbance or exhaustion, together which remain crucial for evaluating driver awareness. The Eye alimnet prototype looks at behaviours that have to do with the eyes, like

- Closing the eyes,
- Rubbing the eyes, and
- Partially closing the eyes.

These guidelines are crucial since they directly connect with levels of alertness and weariness. The Mouth alimnet prototype tracks oral movements and accurately detects the open cavernous and deep although cover of the mouth, which are clear signs of exhaustion. The conditions in which the driver is operating are then classified by the Glasses and Normal Illness Prototype. It makes a distinction between:

- Ordinary image throughout the day.
- Ordinary image throughout the night.
- Night glasses.
- Day glasses; and
- Day sunglasses.

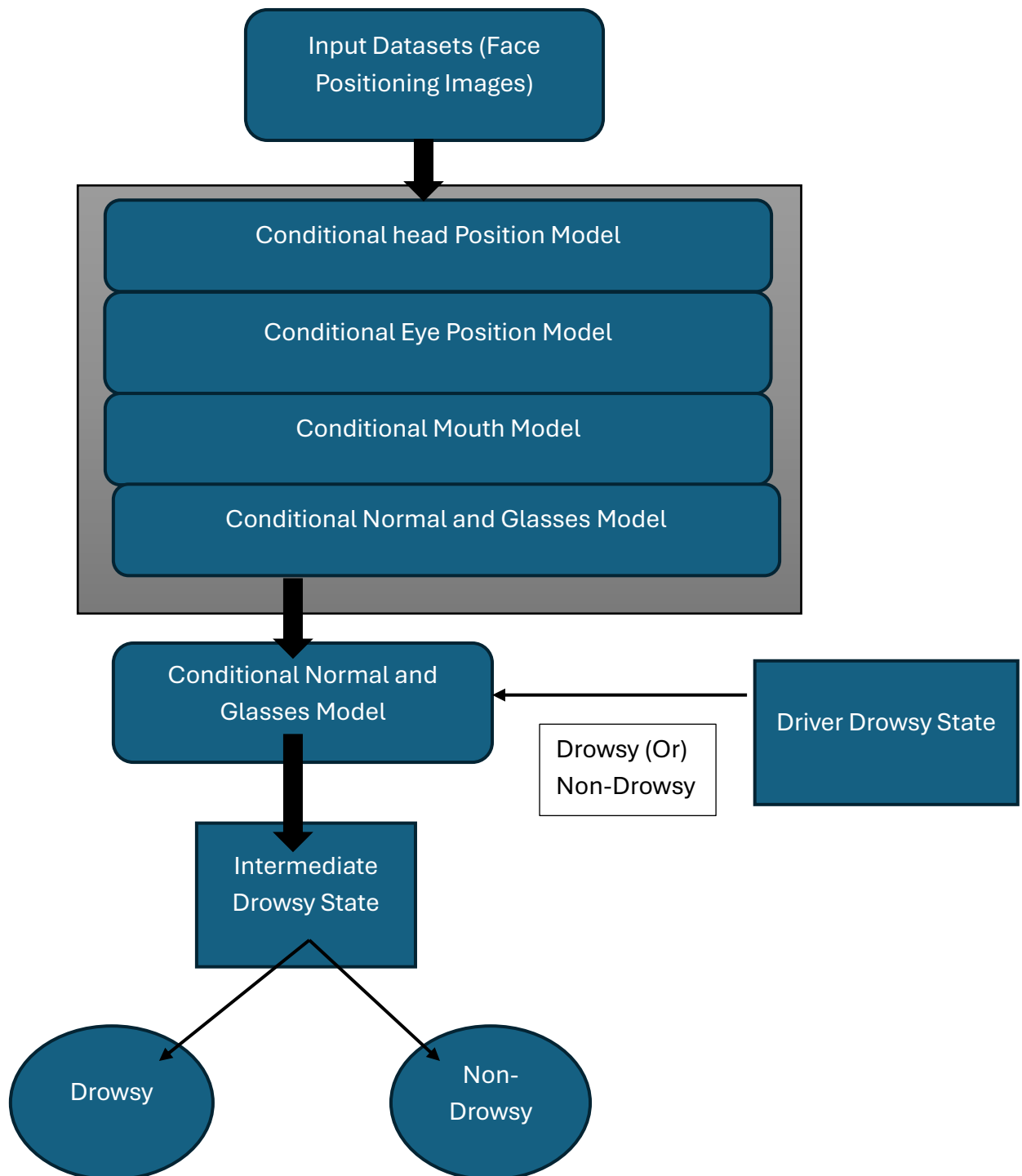


Fig. 1. Proposed Model

Sample locations of KEC-DDD The dataset considered for numerous sub-replicas is illustrated in Figure 4.2

- With the eyes shut,
- Soreness them with the palm of the hand,
- Eyes partially closed,
- Spectacles falling from the eyes,
- Head nodding,

- Eyebrows lifted,
- Head turned to one side,
- Mouth half-open and laughing,
- Mouth half-open and holding a hand, and
- Head, eyes, and mouth set to normal.

To regulate whether the driver is drowsy or not, the Combined Model aggregates and analyses the outputs from several separate sub-models. This grade represents the intermediate level of assurance for the comprehensive driver drowsiness detection system as based on Algorithm 1.

Algorithm 1: Positioning the Face Images and Drowsiness

Input: Face Positioning Datasets

Output: Drowsy Detection

Initialize the Conditional Models

```
{
    Conditional Head Model
        If (Condition is head Nodding || Eyebrows lifting || Eye Closure)
            Conditional Head State (Drowsy)
        Else
            Non-Drowsy State
    Conditional Eye Closure
        If (Eye Closure || Rubbing Eye || Eye Closure)
            Conditional Eye State (Drowsy)
        Else
            Non-Drowsy State
    Conditional Mouth
        If (Day glasses || Sunglass | Night Glass || Normal Day || Night Normal)
            Glasses and Normal Condition
        Else
            Non-Drowsy State
    Combined Model
        If (Output = Head || Eye || Mouth || Glasses || Initial Drowsy Detection)
            Drowsy State
        Else
            Non-Drowsy State
    Detection of Drowsy
        Apply (Drowsiness Condition)
            Classify (Drowsy Level)
            Drowsy State
        Else
            Non-Drowsy State
}
```

3.1. Conditional Head Model:

The head condition model, which categorises head-related behaviours including nodding, eyebrow lifting, and side-looking, is essential to the suggested driver state comprehension system. This algorithm analyses input photographs using a 3D-CNN architecture to accurately detect driver tiredness.

The head condition model uses a three-dimensional image with measures (64, 64, 64, 3), where the fifth and final measurement represents the RGB paint frequencies, while the first four proportions show the position and time hardness of the recorded audiovisual

mounts. The model begins with an initial convolutional sheet with 32 filters and 3x3 kernel sizes. This layer provides non-one-dimensionality by turning on the ReLU function, allowing the prototype to examine complex designs. It will reduce the attributeplots' three-dimensionality. added a second convolutional sheet with 64 filters while keeping the activation function and kernel dimension the same. As seen in Algorithm 2, a max-pooling layer comes after this layer.

The prototype proceeds via 512 units to a fully coupled layer after convolutional and pooling procedures. This layer, which is triggered by the ReLU Purpose, combines geographically dispersed elements into a coherent representation. A dropout level of 0.5 improves model generalisation and convergence time by preventing overfitting.

The output sheet of the head alimnt prototype uses sigmoidal activation. The three elements on this sheet—head nodding, eyebrow rising, and side-looking—each constitute a classification objective. The sigmoid function was used because it may offer predictions for particular periods, allowing for precise selections regarding the driver's head alimnt. The Adam optimiser, which is renowned for its efficiency and adaptability in changing knowledge charges through coaching, is used to build the prototype. Uncompromising cross-entropy is the chosen loss function, which is perfect for multi-session arrangements.

3.2. Eye Alimnt Prototype:

An essential component of the driver state-run consideration method is the eye alimnt prototype. It was designed to classify eye-related activities such as partially closing the eyelids, massaging the eyes with the hand, and closing them completely. A complete 3D-CNN agenda for this problem is displayed in Algorithm 3 in Figure 4.4. This prototype accurately identifies several appreciation situations that indicate whether a driver is drowsy by using the longitudinal and chronological statistics of audiovisual settings.

For the three-dimensional, chronological penetration, and paint frequencies of the contribution descriptions, the eye condition model is given a 3D tensor (32, 32, 32, 3). An 8-filter convolutional sheet is where the classical begins. This sheet learns complex patterns by using the Remedied Linear Element activation function. A max-pooling layer is introduced right after the convolutional procedure. This layer makes the attribute designs more feasible by reducing the computational and three-dimensional load.

The prototype's next stage makes use of ReLU initiation and an extra convolutional sheet. Following this layer, the topographies acquired from the input statistics are refined using a max-pooling sheet with the same pool size. Thirty-two filters with the same kernel extent and activation function are added by the 3D convolutional layer. This is an additional max-pooling sheet. The prototype proceeds to a 256-unit totally related sheet after the convolutional and assembly sheets. In order to combine spatial and temporal components interested in a cohesive representation of eye alimnt, this layer uses ReLU activation. For the fully associated layer, a dropout ratio of 0.5 is useful.

Algorithm 2: Head Alimnt Prototype

Input: Image Shape (64, 64, 64, 3)

Output: Compiled Model (head)

Initialize and Compilation Process

Image Filter (32);

(3, 3, 3) Kernel Size;

Relu Activation Model;

Max Pooling Layer;

(2, 2, 2) Pooling Size;

// Iteration

```

    Image Filter (64);
    (3, 3, 3) Kernel Size;
    Relu Activation Model;
    Max Pooling Layer;
    (2, 2, 2) Pooling Size;
// Iteration
    Image Filter (128);
    (3, 3, 3) Kernel Size;
    Relu Activation Model;
    Max Pooling Layer;
    (2, 2, 2) Pooling Size;
// Fully Connected Layer
    512 Units.
    Relu Activation.
    Dropout Rate: 0.5
// Output Layer
    3 Units
    Sigmoid Activation
// Compile Model
    Adam Optimizer
    Loss Function
    Cross Entropy Categorical
Algorithm 3: Eye Aliment Prototype
Input: Shape (32, 32, 32, 3)
Output: Compile Model (Eye)
Initialize and Compilation Process
    Image Filter (8);
    (3, 3, 3) Kernel Size;
    Relu Activation Model;
    Max Pooling Layer;
    (2, 2, 2) Pooling Size;
// Iteration
    Image Filter (16);
    (3, 3, 3) Kernel Size;
    Relu Activation Model;
    Max Pooling Layer;
    (2, 2, 2) Pooling Size;
// Iteration
    Image Filter (32);
    (3, 3, 3) Kernel Size;
    Relu Activation Model;
    Max Pooling Layer;
    (2, 2, 2) Pooling Size;
// Fully Connected Layer
    256 Units.
    Relu Activation.
    Dropout Rate: 0.5
// Output Layer
    3 Units
    Softmax Activation
// Compile Model
    Adam Optimizer
    Loss Function
    Cross Entropy Categorical

```

Closing layer in eye aliment prototype is yield sheet, utilizing softmax activation purpose. This sheet has 3 units for categorization targets: eye closed, hand-rubbed eye, and half-closed eye. The softmax function provides prospects for individually session, aiding in result-creation concerning the driver's eye aliment. Model compilation uses Adam optimizer. The loss function used is definite cross-entropy, ideal for multi-period arrangement.

Focus on retaining key characteristics while minimizing spatial dimensions.

3.3. Mouth Aliment Prototype:

The goal of the mouth Aliment Prototype is to classify mouth-related behaviours, such as cavernous with the mouth covered and exposed gaping. These behaviours are crucial indicators of driver fatigue. The design of this model uses a 3D-CNN to assess photo classifications, effectively classifying different mouth scenarios. In addition to the Red, Green, and Blue (RGB) shade networks, the standard allows contribution images as 3D tensors through sizes (64, 64, 64, 3), indicating latitudinal and sequential sizes. The first layer of the model consists of a convolutional sheet with 16 screens and a kernel magnitude of (3, 3, 3). The prototype then applies a second convolutional sheet with 32 filters while maintaining the same kernel size and activation function as the previous sheet. In order to reduce spatial dimensions and refine attributes generated from input photographs, this sheet is surveyed using a max-pooling coat through the same pool size. With the same kernel size and activation function, the third convolutional layer increases

the filter count to 64. To ensure that each of the most significant structures is retained, a second max-pooling sheet is utilised.

The classical then switches to a fully associated coating with 256 units that combines the chronological and longitudinal properties into a single representation. By using the ReLU activation function, this coating enhances the replica's ability to acquire statistics. To improve generalisation and prevent overfitting, use a dropout level of 0.5 to this fully related layer.

The mouth alimnet prototype's productivity coating consists of two parts and uses sigmoid activation. As seen in Algorithm 4, this covering creates prospects for the two courses: gaping with mouth covered and exposed cavernous. Because it produces a pure probabilistic productivity, the sigmoid role is appropriate for this two-fold arrangement, helping to create a verdict regarding the driver's mouth condition.

Algorithm 4: Eye Model

Input: Shape (64, 64, 64, 3)

Output: Compile Model (Eye)

Initialize and Compilation Process

```

    Image Filter (16);
    (3, 3, 3) Kernel Size;
    Relu Activation Model;
    Max Pooling Layer;
    (2, 2, 2) Pooling Size;
// Iteration
    Image Filter (32);
    (3, 3, 3) Kernel Size;
    Relu Activation Model;
    Max Pooling Layer;
    (2, 2, 2) Pooling Size;
// Iteration
    Image Filter (64);
    (3, 3, 3) Kernel Size;
    Relu Activation Model;
    Max Pooling Layer;
    (2, 2, 2) Pooling Size;
// Fully Connected Layer
    256 Units.
    Relu Activation.
    Dropout Rate: 0.5
// Output Layer
    2 Units
    Sgmoid Activation
// Compile Model
    Adam Optimizer
    Loss Function
    Cross Entropy Binary

```

3.4. Glass and Normal Condition:

The head alimnet prototype and the glasses and standard conditions prototype both use the same 3D-CNN architecture. Contribution images with magnitudes (64, 64, 64, 3) are overcoated by screens with dimensions of 16, 32, and 64 using a kernel extent of (3, 3, 3) and ReLU activation. Next, max-pooling layers are used. Before obtaining the productivity

sheet, a fully linked 66-sheet through 256 elements and a dropout amount of 0.5 engage in the topographies.

The output layer consists of five units that are triggered by a softmax function, and each of the five classes is given a probability value. Specifically designed for multi-class arrangement tasks, the model makes use of the Adam optimiser with unconditional cross-entropy as the loss purpose. This setup ensures that the driver's visual circumstances are precisely arranged, which is essential for adjusting the sleepiness recognition scheme to different lighting and eyewear conditions.

IV. COMBINED MODEL BASED ON CNN

The integrated model uses the beginning driver sleepiness condition from previous research as well as the outputs from the four sub-models—head, eye, mouth, and glasses/normal situations. As seen in Fig. 2, the integrated model consistently processes five inputs at once since each sub-model produces a categorical output. This typically considers only the maximum value since each sub-model's yieldnervecell. The combined model uses a simple Convolutional Neural Network to classify the driver's state as either sleepy or not. The driver's initial level of drowsiness, the state of their cranium, their eye, their mouth, and their spectacles are the five input features that decide this arrangement.

V. PEFORMANCE ANALYSIS

Validation accuracy is demonstrated by assessment replicates for glasses and typical situations using the NTHU-DDD dataset (Table 4.4). MobileNetV2-DCNN is included in comparison representations. These chosen models allow responses because they are comparable models. Models' utility and robustness in detecting driver fatigue are demonstrated by comparing them under various visual conditions.

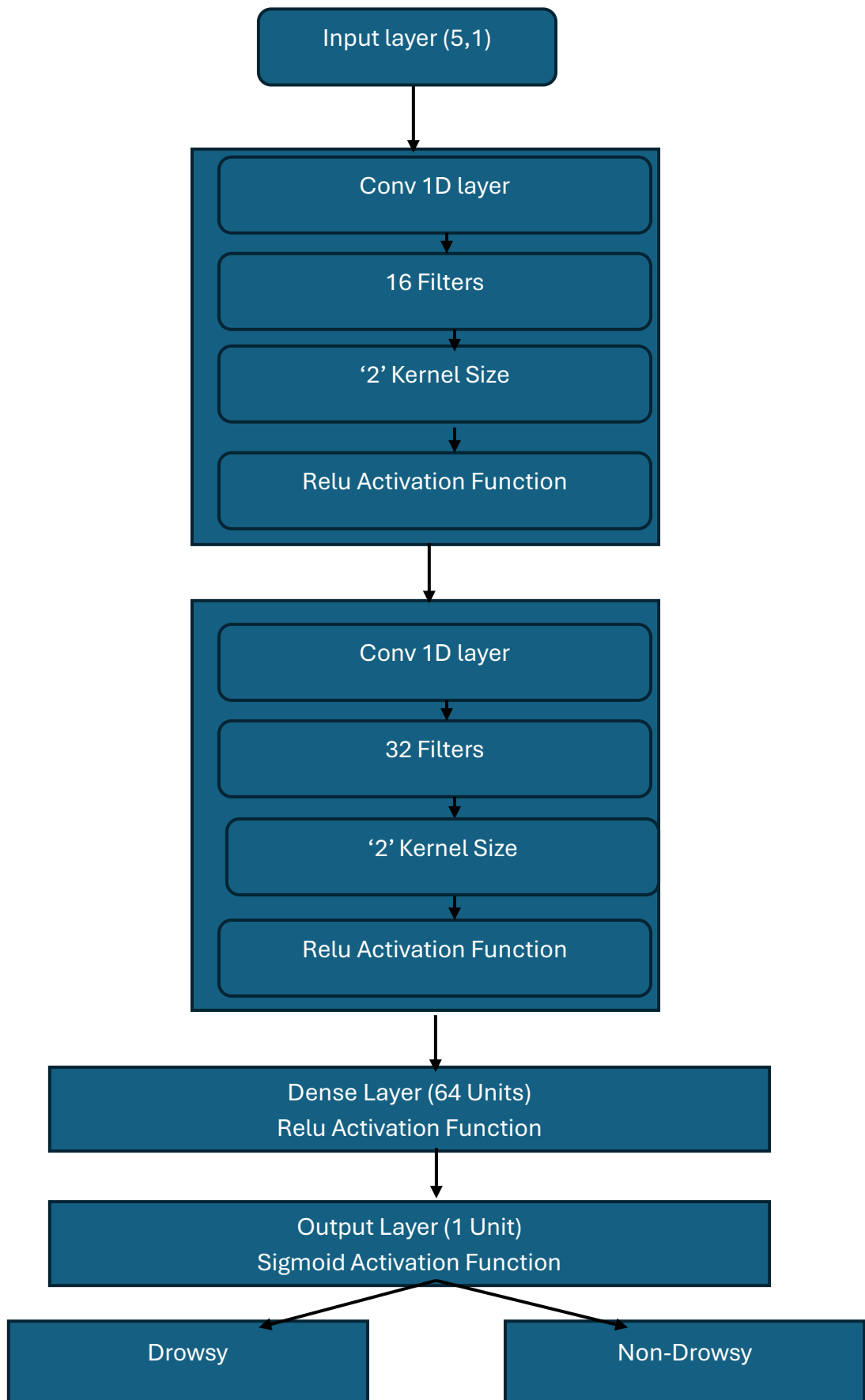


Fig .2. Combined Model Based On CNN

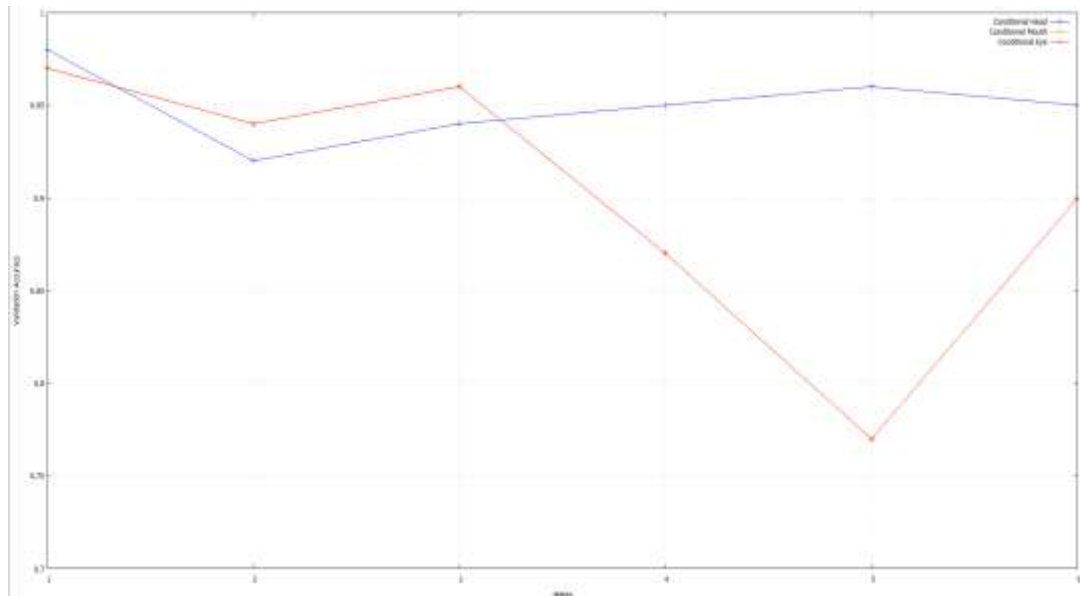


Fig. 3. Avg Validation Accuracy using KEC DDD Dataset

Based on the Fig. 3 & 4, Average accuracy gets calculated based on various states as mentioned below,

1. Day_Normal State
2. Day_Glasses State
3. Night_Normal State
4. Night_Glasses State
5. Day_Sun_Glasses State

For the Fig. 3, Accuracy average is calculated based on the validation process for KEC DDD Datasets for the 3 conditions such as, Head; Mouth and Eye. As based on the graph values are getting distracted based on the variation in the values ranges from 0.7 to 0.9.

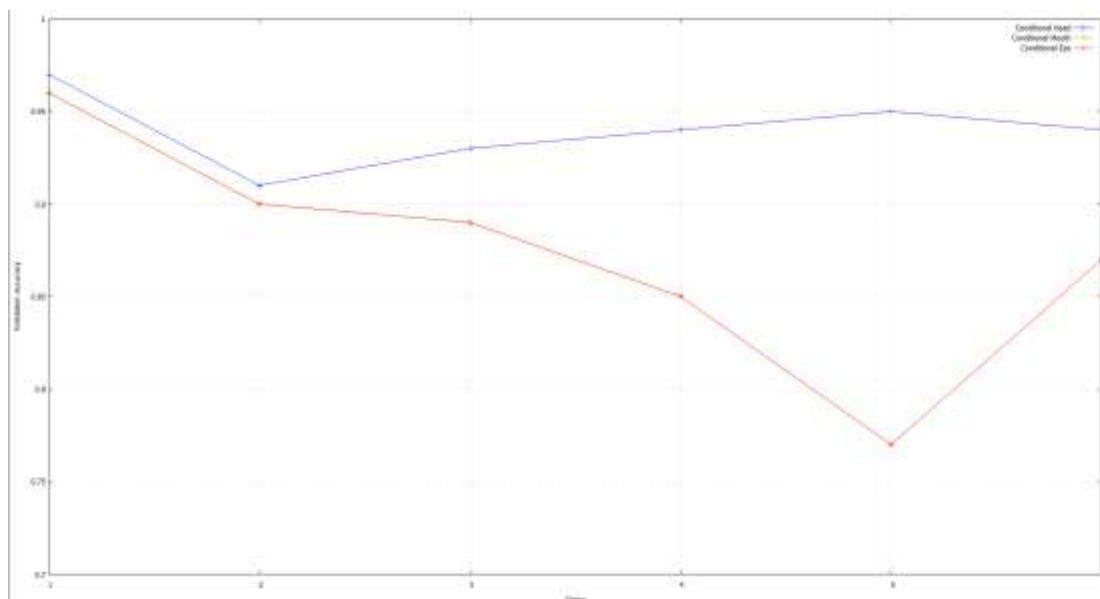


Fig. 4. Avg Validation Accuracy using NTHU DDD Dataset

For the Fig. 4, Accuracy average is calculated based on the validation process for NTHU DDD Datasets for the 3 conditions such as, Head; Mouth and Eye. As based on the graph values are getting distracted based on the variation in the values ranges from 0.7 to 0.9.

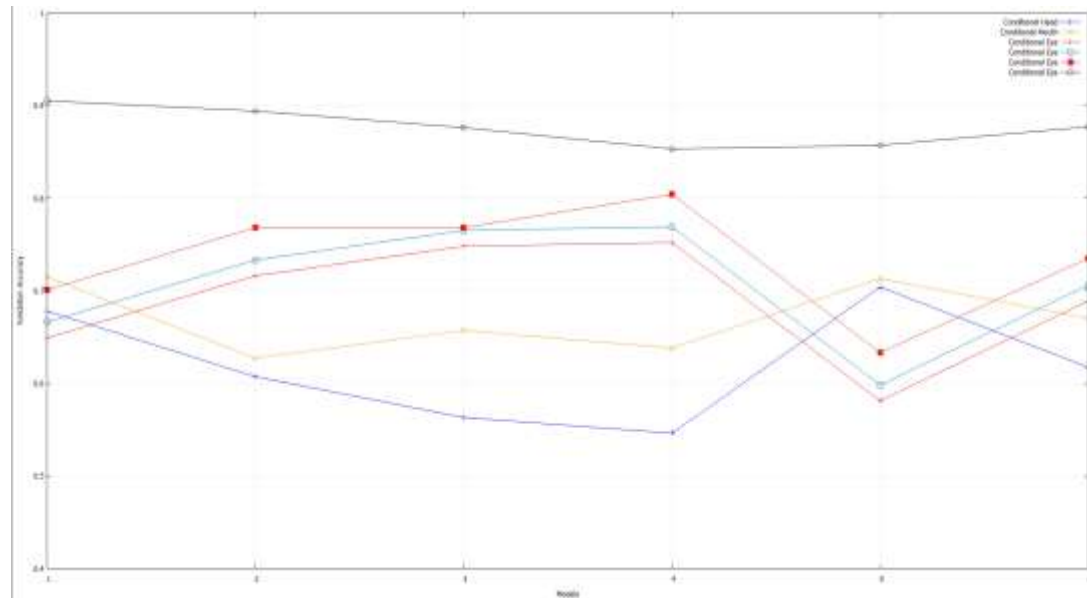


Fig. 5. Avg Validation Accuracy using NTHU DDD Dataset by comparing various models

For the Fig. 5, Accuracy average is calculated based on the validation process for NTHU DDD Datasets for the various models such as, MobileNet V2-DCNN; BisLSTM-DCNN; Multi CNN Deep CNN; 3D CNN; R CNN; Multistage Adaptive 3D CNN. As based on the graph values are getting distracted based on the variation in the values ranges from 0.4 to 1.0.

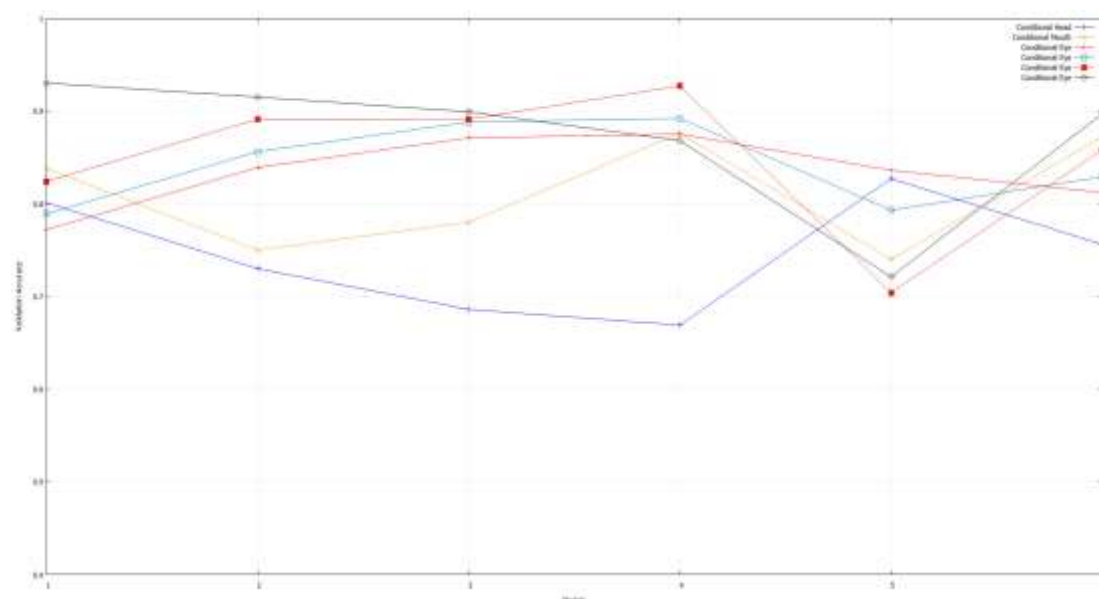
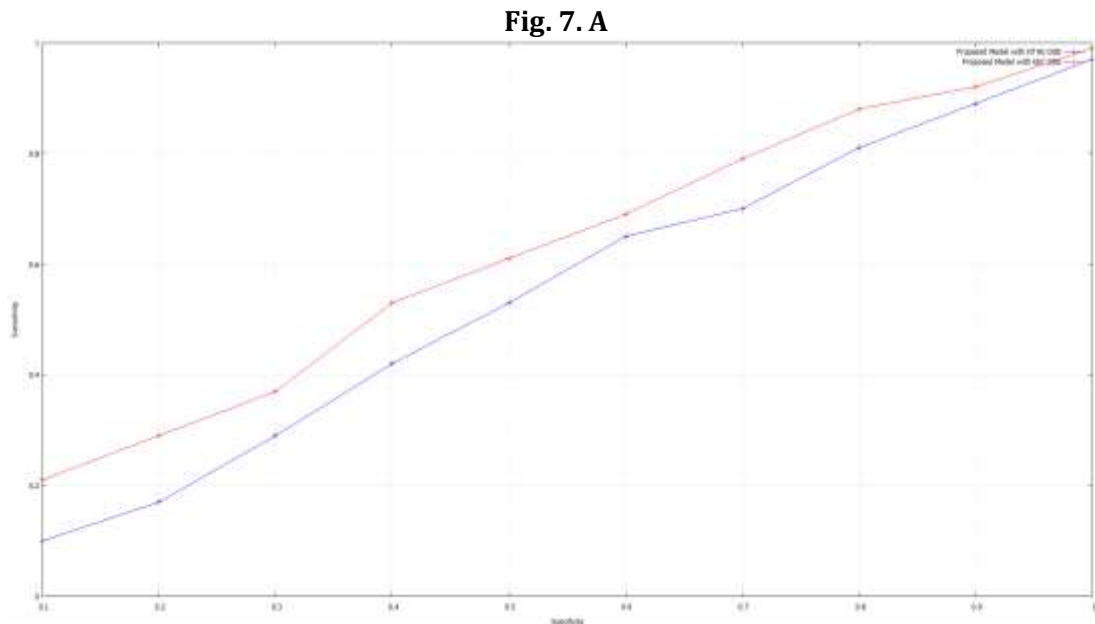


Fig. 6. Avg Validation Accuracy using KEC DDD Dataset by comparing various models

For the Fig. 6, Accuracy average is calculated based on the validation process for KEC DDD Datasets for the various models such as, MobileNet V2-DCNN; BisLSTM-DCNN; Multi CNN Deep CNN; 3D CNN; R CNN; Multistage Adaptive 3D CNN. As based on the graph values are getting distracted based on the variation in the values ranges from 0.4 to 1.0.



UC ROC using KEC and NTHU & DDD Dataset for the proposed model

For the Fig. 7, AUC ROC is calculated based on the validation process for NTHU & KEC DDD Datasets for the various specificity ranges from 0.0 to 1.0. As based on the graph values gets gradually increases from 0.0 to 1.0 for sensitivity as based on the increase in the values ranges from 0.4 to 1.0.

VI. CONCLUSION

We investigated sub-models to find and categorise signs of driver fatigue. Each sub-model's head, eye, mouth, and glasses/normal face features are gathered using distinct 3D CNN architectures. to identify sideways looking, eyebrow raising, and head nodding. The mouth condition model may identify both covered and open yawning. Each model is essential to the system because it uses specific detection algorithms to address various aspects of driver sleepiness. By combining the outputs of four sub-models into a single model, the accuracy and dependability of sleepiness detection are significantly increased. The system assesses the driver's condition by combining findings from both drowsy and non-drowsy models, making use of each model's advantages. The performance of the sub-models was thoroughly investigated, with an analysis of their resilience and flexibility in a variety of real-world scenarios. The results of the dataset validation verify that the sub-models are capable of effectively capturing driver behaviours in a variety of situations. This step creates the mid-level driver drowsiness condition using the KEC-DDD dataset, providing a crucial intermediate evaluation for more study.

References

- [1] Hu, Y, Lu, M, Xie, C & Lu, X 2020, 'Driver Drowsiness Recognition via 3D Conditional GAN and Two-Level Attention Bi-LSTM', *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4755-4768.
- [2] Memisevic, R 2013, 'Learning to relate images', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1829-1846.
- [3] Krizhevsky, A, Sutskever, I & Hinton, GE 2017, 'Imagenet classification with deep convolutional neural networks', *Communications of the ACM*, vol. 60, no. 6, pp. 84-90.
- [4] Lecun, Y, Bottou, L, Bengio, Y & Haffner, P 1998, 'Gradient-based learning applied to document recognition', In *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324.
- [5] Lee, K & Park, M 2021, 'Real-Time Drowsiness Detection Using Eye Closure and Head Pose Estimation', *Sensors*, vol. 8, no. 2, pp. 4071-4082.
- [6] Husain, SS, Mir, J, Anwar, SM, Rafique, W & Ullah, MO 2022, 'Development and validation of a deep learning-based algorithm for drowsiness detection in facial photographs', *Multimedia Tools and Applications*, vol. 81, no. 1, pp. 20425-20441.
- [7] Hwang, J & Kim, N 2021, 'An ensemble learning model for driver drowsiness detection and accident prevention using the behavioral features analysis', *International Journal of Intelligent Computing and Cybernetics*, vol. 15, no. 2, pp. 224-244.
- [8] Ji, S, Xu, W, Yang, M & Yu, K 2013, '3D convolutional neural networks for human action recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 2, pp. 221-231.
- [9] Ji, Z, Cao, Y, & Li, J 2020, 'Self-Supervised Learning for Face Feature Extraction', In *Proceedings of IEEE International Conference on Multimedia and Expo ICME*, pp. 1-6. 28.
- [10] Johnson, M & Davis, E 2022, 'Real-Time Driver Drowsiness Detection Using Convolutional Neural Networks', In *Proceedings of 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, pp. 456-467. 112
- [11] Hu, Z, Duan, Q, Zhang, P & Tao, H, 'An attention-erasing stripe pyramid network for face forgery detection', *Signal Image and Video Processing*, vol. 17, no. 8, pp. 4123-4144.
- [12] Hugar, JG, Naseer, MM, Waris, A & Khan, MA 2020, 'Road Traffic Accident Research in India: A Scientometric Study from 1977 to 2020', *Library Philosophy and Practice*, vol. 14, no. 4325, pp. 1-19. 24.
- [13] Ranjan, R, Patel, VM & Chellappa, R 2016, 'HyperFace: A Deep Multi Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 1, pp. 121-135.
- [14] Kim, S & Lee, H 2020, 'AUTOMOTIVE: A Case Study on Automatic multimodal Drowsiness detection for smart vehicles', *IEEE Access*, vol. 9, no. 1, pp. 153678-153700.
- [15] King, DE 2009, 'Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research*', vol. 10, no. 3, pp. 1755-1758.
- [16] Liu, W & Chen, X 2020, 'Deep Belief Networks for Driver Drowsiness Detection', *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4505-4514.

- [17] Liu, W, Wen, Y, Yu, Z, Li, M, Raj, B & Song, L 2017, 'SphereFace: Deep Hypersphere Embedding for Face Recognition', In the proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 6738-6746. 113
- [18] Magan, E, Sesmero, MP, Alonso-Weber, JM & Sanchis, A 2022, 'Driver Drowsiness Detection by Applying Deep Learning Techniques to Sequences of Images', *Applied Sciences*, vol. 12, no. 2, pp. 1145-1163.
- [19] Nguyen, T & Bui, H 2020, 'Driver Drowsiness Detection Using Deep Learning and Infrared Sensors', *Applied Sciences*, vol. 10, no. 9, pp. 901-922.
- [20] Parkhi, OM, Vedaldi, A & Zisserman, A 2015, 'Deep Face Recognition', In the proceedings of British Machine Vision Conference (BMVC), Swansea, UK, pp. 786-803.
- [21] Patel, A & Gupta, R 2022, 'Driver Drowsiness Detection Based on Eyelid Movements', *Pattern Recognition Letters*, vol. 14, no. 5, pp. 110-119.
- [22] Li, H, Lin, Z, Shen, X, Brandt, J & Hua, G 2015, 'Face Detection and Alignment in Unconstrained Environments Using Deep Learning', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 382, no. 3, pp. 242-255.
- [23] Li, H, Wang, J, & Fu, Y 2018, 'FaceFeatureNet: A Lightweight and Efficient Face Feature Extraction Model', In the proceedings of 25th IEEE International Conference on Image Processing, Athens, Greece, pp. 2495-2499.
- [24] Li, P, Wang, Z, Wang, B & Yang, Y 2020, 'Adaptive Feature Pyramid Network for Face Detection', In the proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, pp. 11942-11951.
- [25] Schroff, F, Kalenichenko, D, & Philbin, J 2015, 'FaceNet: A Unified Embedding for Face Recognition and Clustering', In the proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, pp. 815-823.
- [26] Shen, W, Wang, Y & Bai, X 2017, 'Context-Aware Face Detection and Landmark Localization', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 10, pp. 2519-2528.
- [27] Shepley, A, Falzon, G & Kwan, P 2021, 'Confluence: A robust non-IoU alternative to non-maxima suppression in object detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 10, pp. 1-15.
- [28] Singh, A & Chawla, N 2020, 'Driver Drowsiness Detection Using a Hybrid Approach of CNN and LSTM', *Pattern Analysis and Applications*, vol. 23, no. 4, pp. 1601-1613.