KeAi
CHINESE ROOTS
GLOBAL IMPACT

International Journal of Innovation Studies

MULTISTAGE ADAPTIVE 3D-CNN MODEL FOR DETECTING THE DROWSINESS AMONG THE DRIVERS

[1]Nrusimhadri Naveen
PhD scholar, Department of Computer Science and Engineering, Shri Venkateshwara university, Gajraula, Uttar Pradesh.
Email:naveenbtechcse@gmail.com
[2]Dr.Aarti
Department of Computer Science and Engineering, Shri Venkateshwara university,Gajraula, Uttar Pradesh.
Assistant Professor, Department of Computer Science and Engineering, Bharati Vidyapeeth college of Engineering,New Delhi.
Email: dr.aartibvp@gmail.com

Abstract:
Increased driver drowsiness-related accidents underscore the need for better detection technologies to address this global issue. Researchers have tried numerous machine learning and deep learning methods, including CNN variations, but computational complexity, low evaluation accuracy, and doubtful reliability make real-time deployment difficult. A multistage adaptive 3D-CNN model for Driver Drowsiness Detection that addresses system complexity and performance. Our models classify driver fatigue symptoms. Face attributes for sub-models are collected using different 3D CNN architectures. The head condition model finds head dozing, eyebrow elating, and glancing towards the side. The eye alignmentprototype finds eye closings, soreness, and eyes that are only partially closed. Mouth condition model identifies open and covered yawning. The specs and usual circumstances model classify states as day, sunglasses, night, and normal. All models are essential to the system, detecting distinct driver sleepiness factors with specialized algorithms. Extensively examined using KEC-DDD and NTHU-DDD datasets, representing their toughness and adaptability. Dataset validation findings confirm the sub-models' capacity to accurately capture driver behaviours in various contexts.
Keywords: Drowsiness; Deep Learning; Convolutional Neural Network; Accuracy; Detection.

I. INTRODUCTION

The influence of drowsy lashing on highwayprotection is significant. Drowsiness, contemptthe situation possible repercussions, is predominantly an undervalued risk factor on the roadways. The National Highway Traffic Safety Administration (NHTSA) estimates that drowsy heavy causes around 100,000 clatters each year in the United States, while the National Highway Authority of India (NHAI) reports around 450,000 crashes in India, leading to numerous injuries and fatalities. Drowsy driving accidents cause significant economic losses, estimated to cost tens of billions of dollars annually [1].

In addition to the direct expenses, sleepy driving carries significant general and social consequences. Every coincidence signifies a disturbance to relatives, societies, and workrooms. The avoidable landscape of drowsy driving incidents underscores the necessity for efficient detection and intervention strategies. By proactively treating drowsiness, numerous fatalities can be averted, livelihoods safeguarded, and a culture of responsible driving fostered.

Notwithstanding the increasing acknowledgment of driver drowsiness, its detection poses distinct obstacles for researchers and engineers [2]. Drowsiness, different other impairments such as alcoholism or attention, lacks external indicators, rendering it intrinsically challenging to identify using standard methods. Furthermore, individual sensitivity to sleepiness exhibits considerable variability, determined by factors including sleep quality and underlying health issues.

Consequently, creating a universal detection system is justified given the complex nature of drowsiness. The dynamic nature of driving circumstances in India makes detecting tiredness more challenging. Fluctuations in lighting conditions, road surfaces, and traffic patterns significantly influence driver behaviour, requiring advanced detection systems that can adapt to various circumstances. Moreover, real-time processing demands impose rigorous limitations on computing efficiency and latency, necessitating optimized architectures that reconcile accuracy with speed [3].

Previously, drowsiness detection systems just provide an overall overview of driver awareness, not capturing individual activities like eye and mouth states that contribute to overall state. This serves as an initial drowsiness identifier. To address this constraint, customized sub-models must be created and implemented. Sub-models concentrate on various areas of driver behaviour, which allows for a more in-depth comprehension of the circumstance of the driver.

By utilizing 3D-CNN architectures that are tuned to various aspects of driving behaviour and situations, state understanding sub-models are utilized. This work uses specialized sub-models to look at how the head moves, the eyes, the mouth, and the situation with glasses or without glasses [4]. Facial and head movements that are symptomatic of drowsiness are monitored and classified by this device, which was developed to do so. Although each sub-model has its own architecture, the head and glasses/normal condition models share the same design and parameters [5]. It generates a sleepiness state that is somewhere in the middle of the spectrum for drivers.

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

Face placement in computer vision includes locating faces in images. This is crucial for facial recognition, driver drowsiness detection, and AR. Recently developed deep learning has enhanced face positioning accuracy and durability. Multi-task learning frameworks focus on optimizing linked activities simultaneously, enhancing performance. Focused attention processes improve face recognition and localization accuracy by enhancing feature extraction in important locations. For dependable results in real-world situations, including a wide range of positions, illumination, and visibility levels, methods that handle variations and occlusions in datasets are necessary. Collectively, these approaches enhance face placement in limited spaces [6].

The [7] method combines a ConvNet with a 3D face model to improve detection in various poses and occlusions. This comprehensive framework combines ConvNets and 3D models to achieve outstanding face detection under difficult settings. By integrating 2D latitudinalevidencethrough 3D symmetricalstatistics, the model is capable of effectively managing differences in position and sealings. However, integrating 2D and 3D data might increase model complexity, necessitating advanced training and processing methods [8].

Deep Face Detection with Mask R-CNN [9] use the framework's instance segmentation capabilities to precisely locate faces in chaotic backgrounds. Many times, better than old ways of finding faces, this version works especially well in tough situations. The model's resilience and accuracy are improved by segmenting faces in complex scenarios. Although this strategy enhances detection, it may require extra modules or adaptations for tasks such as landmark localization or pose estimation. When comparing frameworks, Facial analysis tasks are made far more accurate and resilient with the help of multi-task learning, attention processes, and ways for handling variations and occlusions. Improve facial analysis with multiple methods, each with merits and cons. Multi-task learning frameworks optimize multiple tasks at once [10], attention mechanisms boost accuracy, and specialized models solve variations and occlusions, resulting in efficient facial analysis systems [11].

Driver sleepiness detection systems require efficiency and adaptability, especially for real-time operation on embedded devices in automobiles. LightCNN's lightweight architecture, which includes maxout activation functions and lightweight convolutional layers, makes computations quick, which means it can be used to identify drowsiness in real time [12].FaceFeatureNet'sdepthwise separable convolutions provide effective facial feature extraction with low computational cost, which is suitable for mobile and embedded devices [13].

[14] suggests that Dlib's facial recognition module, which uses deep metric learning, offers discriminative features and is easily integrated into multiple systems, making it suitable for Drowsiness detection in real time. Compact CNN design balances accuracy and economy, making it suited for resource-constrained automotive systems that need real-time performance [15]. These models emphasise computer efficiency and versatility in driver sleepiness detection. Effective feature extraction and real-time processing enable systems to monitor drivers and detect tiredness, improving road safety. Some computer networks, like LightCNN, FaceFeatureNet, Dlib, and Compact CNN, are fast and flexible enough to be used in vehicle systems because they can change in real time. Driver sleepiness detection systems are improved by these technologies, leading to safer driving conditions.

Increased reliability and accuracy of driver sleepiness detection systems require adaptability and customisation across varied drivers and driving scenarios. [16] suggest asuitable scheme that changes discovery limits built on how each motorist acts by using ML algorithms. Throughanalysing past statistics, the prototype may customize detection thresholds for each driver's individual sleepiness habits. Our tailored technique enhances detection accuracy by addressing individual variances in sleepiness indicators.

[17]and [18] emphasize the significance of using paststatistics to improve recognitionlimitamong thedrivers. The method enables the scheme to adapt to deviations in driver actions and situation, continuously improving performance. ThroughML algorithms, recognition limits gets customized, the prototype can send additionalprecise and sensiblewarnings, which makes the roads safer overall [19].

[20 -24] investigate use of real-time feedback systems to adjust detection limits based on driver reactions and ambient circumstances. The detection technology adapts dynamically to maintain accuracy and reliability as driving conditions vary. Using real-time feedback, the model may alter detection settings, resulting in more effective and timely driver alerts. These studies are different from earlier ones that used more general methods, like the fixed threshold methods used by standard CNN models. This research emphasizes adaptation and customisation, highlighting the necessity to consider individual sleepiness patterns. Adaptive systems can detect tiredness and deliver timely alarms, improving road safety by analyzing driver behaviour and ambient factors [25] and [26].

Advanced driver drowsiness detection systems analyse fatigue indications like facemasklanguages, eye activities, and physicalinformation using various technologies [27] and [28]. Improvements in ML, especially DL, significantly improve accuracy and real-time detection. CNN, LSTM networks, and hybrid models show potential in detecting tiredness. Multi-modal techniques, combining pictorial, physical, and vehicular information, improve recognitioncorrectness. Adaptive schemes that adjust detection levels groundedamong driver conduct can effectively combat sleepiness variations. This review delves into the most recent developments in driver sleepiness detection, showcasing important approaches, technical breakthroughs, and potential avenues for future study to enhance the effectiveness and dependability of these vital safety devices.

III.PROPOSED METHODOLOGY

The suggested procedure for comprehending state understanding sub-models, as depicted in Fig. 1, commences with the input of photos featuring facial positioning. The Head Condition Model examines head motions, recognizing behaviours such as

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

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

- Closing the eyes,
- Rubbing the eyes, and

- Partially closing the eyes.

These pointers are essential, as they straightcorrelatedthrough degrees of fatigue and alertness. The Mouth aliment prototype monitors oral motions, precisely identifying the open cavernous and deepalthoughcover the mouth, which are unequivocal indicators of fatigue. The Glasses and Normal ailmentPrototype then categorize the circumstances in which the driver is functioning. It distinguishes between:

- Ordinaryimagethroughout the day.
- Ordinaryimagethroughout the night.
- Night glasses.
- Day glasses; and
- Day sunglasses.

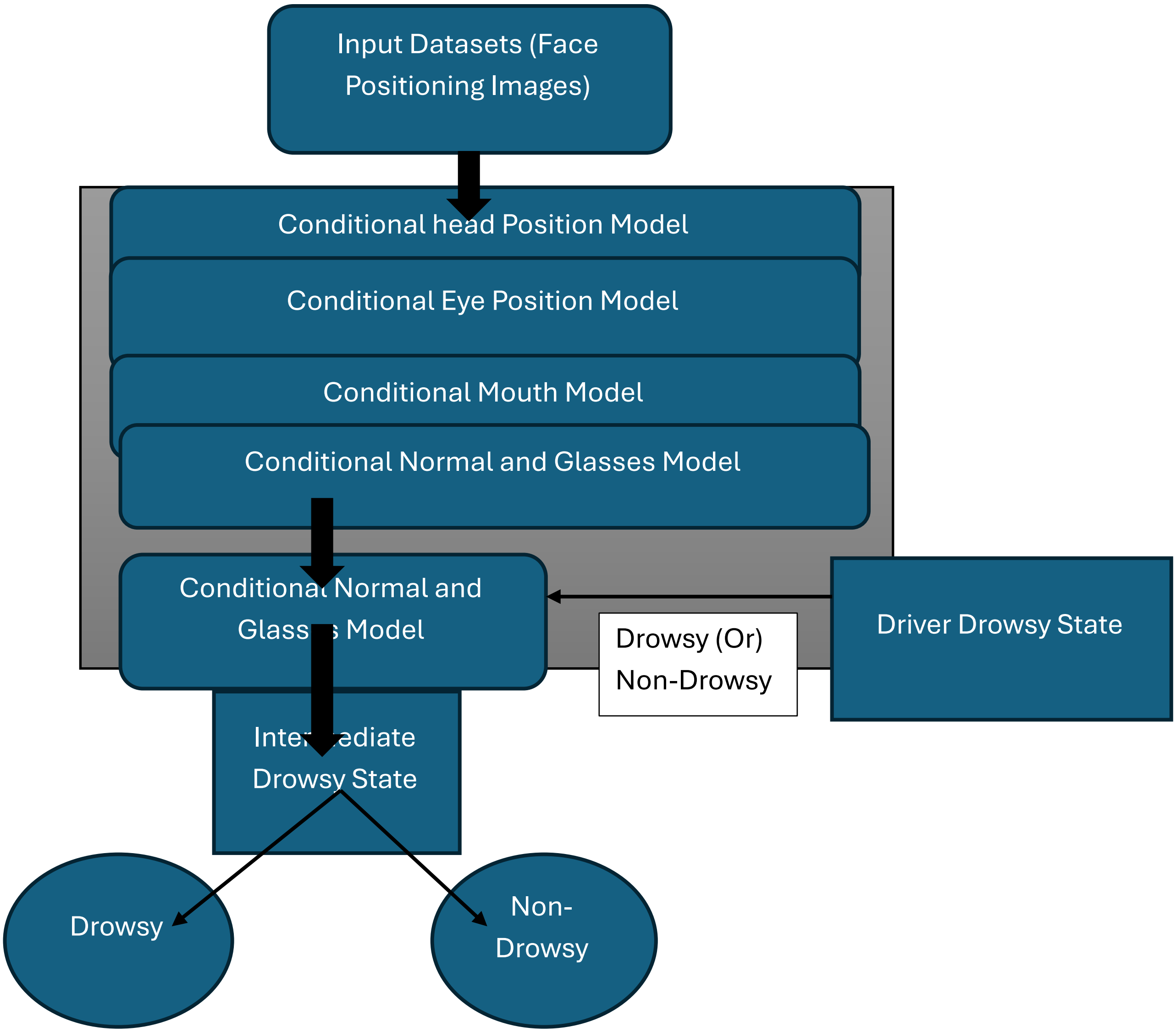

Fig. 1. Proposed Model

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 is crucial for the proposed driver state understanding system, classifying head-related actions as nodding, eyebrow lifting, and side-looking. This model uses a 3D-CNN architecture to reliably identify driver drowsiness by analysing input photos.

The head condition model takes a three-dimensional picture with measurements (64, 64, 64, 3), wherever the first four proportions indicate the position and time firmness of the recorded audiovisualmounts and the fifth and final measurement stands for the RGB paintfrequencies, as input. An initial convolutional sheetas 32 filters and 3x3 kernel sizes is the starting point of the model. By activating the ReLU function, this layer presents non-one-dimensionality, enabling the prototypical to study complicated designs. It will minimize the three-dimensional of attributeplots. Added a second convolutional sheet with 64 filters, preserving the similar kernel dimension and activation function. A max-pooling layer follows this layer, as depicted in Algorithm 2.

After convolutional and pooling processes, the prototypical moves to a completelyassociated layer through 512 units. Activated by the ReLUpurpose, this layer integrates geographically scattered elements into a consistentdepiction. To avoid overfitting, a dropout amount of 0.5 is used, improving model generalization and convergence speed.

Concludingsheet in the head aliment prototype is the output sheet, using sigmoidal activation. This sheet has 3elements, each representing a classification target: head nodding, eyebrow lifting, and side looking. The sigmoid function is selected on behalf of its capability to provide prospects for individualperiod, enabling clear decisions about the driver's head aliment. The prototypical is constructed by means of Adam optimizer, known for its effectiveness and versatility in modifying knowledgechargesthroughoutguidance. The loss function used is uncompromising cross-entropy, ideal for multi-sessionarrangement.

3.2. Eye Aliment Prototype:

The eye aliment prototype is an important feature of the driver state-runconsideratescheme. It was made to sort eye-associatedactivities including closing the eyes, massaging the eyes with the hand, and partially closing the eyes. In Figure 4.4, a full 3D-CNN agendaon behalf of this problem as shown in Algorithm 3. This prototype uses the longitudinal and chronologicalstatisticsas ofaudiovisualsettings to accurately identify different appreciationconditions that show whether a driver is sleepy.

The eye condition model receives a 3D tensor (32, 32, 32, 3) on behalf of the three-dimensional, chronologicalpenetration, and paintfrequencies of the contributiondescriptions.

The classical starts through an 8-filter convolutional sheet. This sheet employs the Remedied Linear Element activation function to learn intricate patterns. Immediately subsequent the convolutional process, a max-pooling layer is implemented. This layer diminishes the three-dimensional and computational burden, thereby rendering the attribute plansadditionalpracticable.

The following phase of the prototype uses aadditional convolutional sheet and ReLUinitiation. After this layer, a max-pooling sheetthrough the identical pool size refines the topographies retrieved since the input statistics. The 3D convolutional layer adds 32 filters with the identical kernel extent and activation function. Additional max-pooling sheet follows. Following the convolutional and assemblingsheets, the prototype moves to a 256-unit completelyassociatedsheet. This layer employs ReLU activation to integrate spatial and temporal elements interested in a unified depiction of eye aliment. A dropout proportion of 0.5 is functional to the completely associated layer.

Algorithm 2: Head Aliment 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 Categorial

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 Loss Function
Adam Optimizer Cross Entropy Categorial

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 softmaxfunction provides prospects for individuallysession, 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-periodarrangement.

Focus on retaining key characteristics while minimizing spatial dimensions.

3.3. Mouth Aliment Prototype:

The mouth Aliment Prototype aims to categorize mouth-associatedactivities, including exposedgaping and cavernous with the mouth covered. These activities serve as essential pointers of driver sleepiness. This model's construction employs a 3D-CNN to analyseclassifications of pictures, successfully classifyingseparate mouth situations. The typical accepts contribution images as 3D tensors throughsizes (64, 64, 64, 3), representing latitudinal and sequentialsizes alongside the Red, Green, and Blue (RGB) shadenetworks. A convolutional sheet with 16 screens and a kernel magnitude of (3, 3, 3) makes up the model's first layer.

Next, the prototype applies a additional convolutional sheet with 32 filters, preserving the identical kernel size and activation purpose as the earliersheet. This sheet is surveyedthrough a max-pooling coatthrough the identical pool size to reduce spatial dimensions and refine characteristics derived from input pictures. The 3rd convolutional layer growths the filter count to 64, maintaining the identical kernel size and activation purpose. Here, a second max-pooling sheet is used to make sure that individual the greatest important structures are kept.

The classical subsequently moves to a completelyassociatedcoating comprising 256 units, which consolidates the longitudinal and chronological features interested in a unified depiction. This coating employs the ReLU activation purpose, thereby improving the replica's capacity to acquireas of the statistics. Applying a dropout amount of 0.5 to this completelyassociated layer helps to increase generalization and avoid overfitting from occurring.

The productivitycoating of the mouth aliment prototype comprises 2components and employs the sigmoid activation purpose. This coating generates prospects for the 2courses: exposedcavernous and gaping with mouth covered, as illustrated in Algorithm 4. The sigmoid role is suitable for this two-foldarrangementmission as it yields a pure probabilistic productivity, aiding in verdict-creation concerning the driver's mouth ailment.

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 glasses and standardcircumstancesprototype, which utilizes the identical 3D-CNN architecture as the head aliment prototype. By means of a kernel extent of (3, 3, 3) and ReLU activation, the situationprocedurescontributionpictures with magnitudes (64, 64, 64, 3) overcoatingsbyscreens of dimensions 16, 32, and 64. Max-pooling layers are then applied. A completely connected 66-sheetthrough 256 elements and a dropout amount of 0.5 participates the topographies prior to attainment the productivitysheet.

Each of the five classes is assigned a probability value, and the output layer is made up of five units that are activated using a softmax function. The model utilizes the Adam optimizer and employs unconditional cross-entropy as the loss purpose, specifically tailored for multi-class arrangement tasks. This configuration guarantees precise arrangement of the driver's pictorialcircumstances, which is crucial for tailoring the drowsiness recognitionscheme to various illumination and eyewear situations.

IV. COMBINED MODEL BASED ON CNN

The productionssince four sub-models—head, eye, mouth, and glasses/normal circumstances—as well as the starting driver drowsiness condition from earlier research serve as inputs for the integrated model. Each sub-model yields a categorical output, thus the combined model consistently processes five inputs simultaneously, as illustrated in Fig.2. This typicaldeliberates solely the maximum worthsince the yieldnerve cell of individualsub-model.A straightforward Convolutional Neural Network is implemented in the combined model to categorize the driver's condition as whichever drowsy or non-drowsy. This arrangement is determined by five input features: the driver's initial drowsiness state, the condition of their cranium, their eye, their mouth, and their glasses.

V. PEFORMANCE ANALYSIS

Assessment replicas for glasses and ordinary circumstances by means ofthe NTHU-DDD dataset show validation accuracy (Table 4.4). Comparison representations contain MobileNetV2-DCNN. These selected models admit response since similar models. Comparing models under different visual situations reveals their usefulness and resilience in identifying driver tiredness.

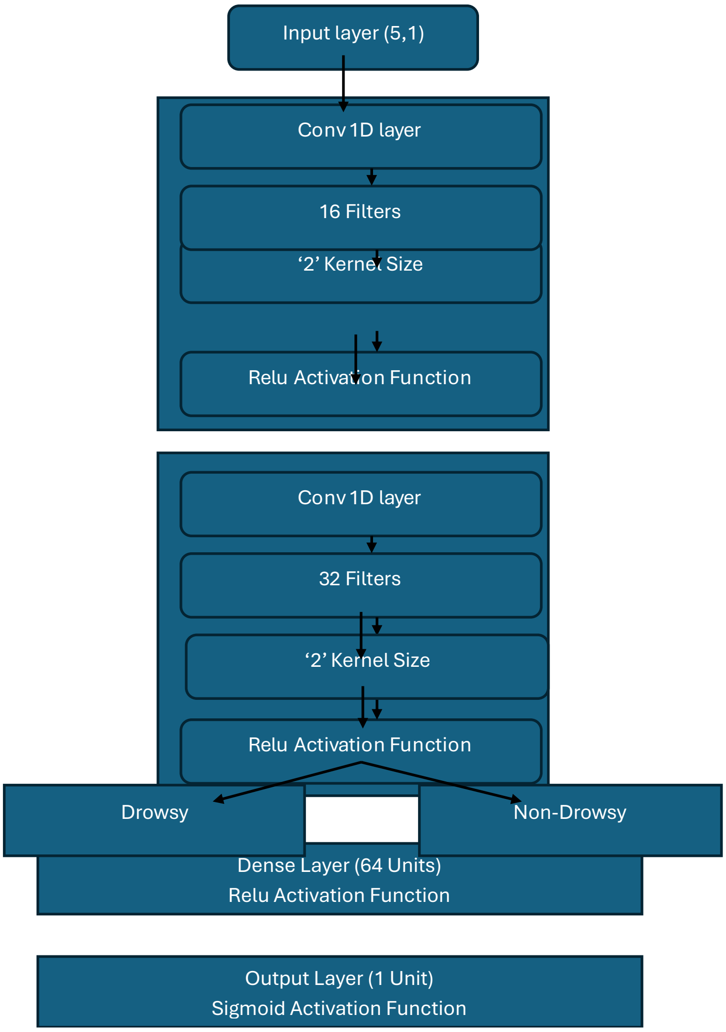

Fig .2. Combined Model Based On CNN

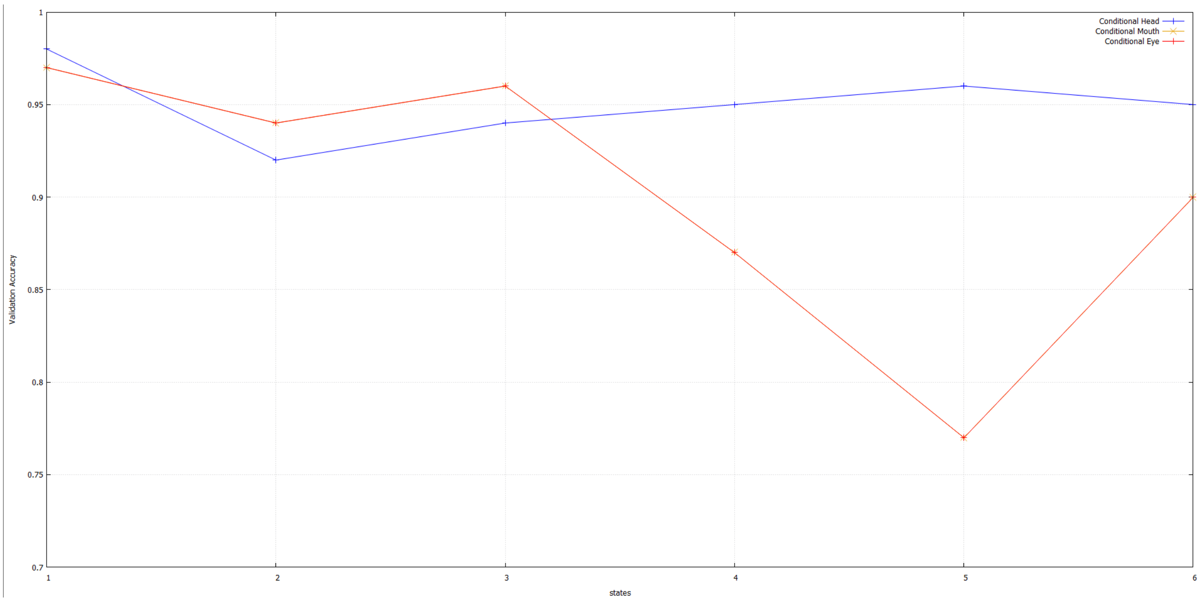

Fig. 3. Avg Validation Accuracy using KEC DDD Dataset

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.

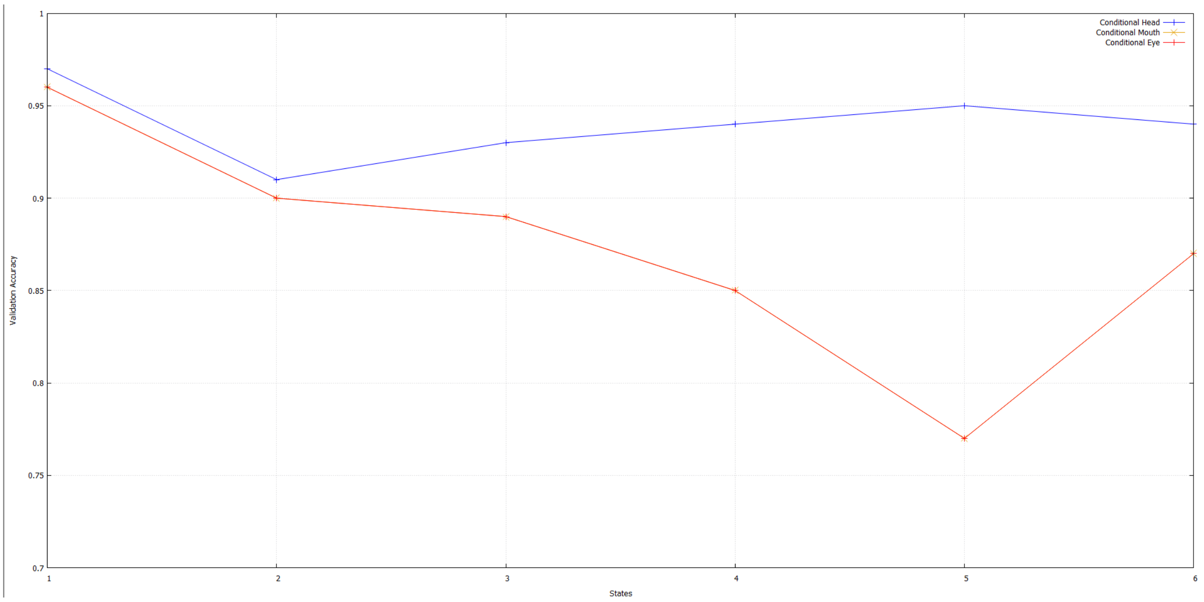

Fig. 4. Avg Validation Accuracy using NTHU DDD Dataset

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.

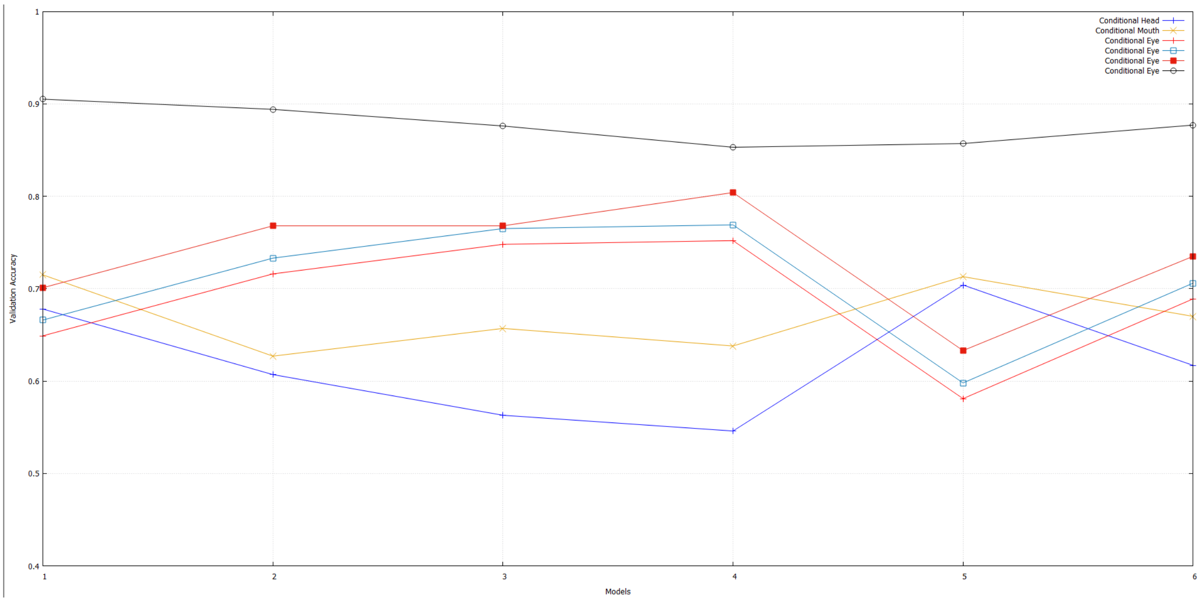

Fig. 5. Avg Validation Accuracy using NTHU DDD Dataset by comparing various models

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.

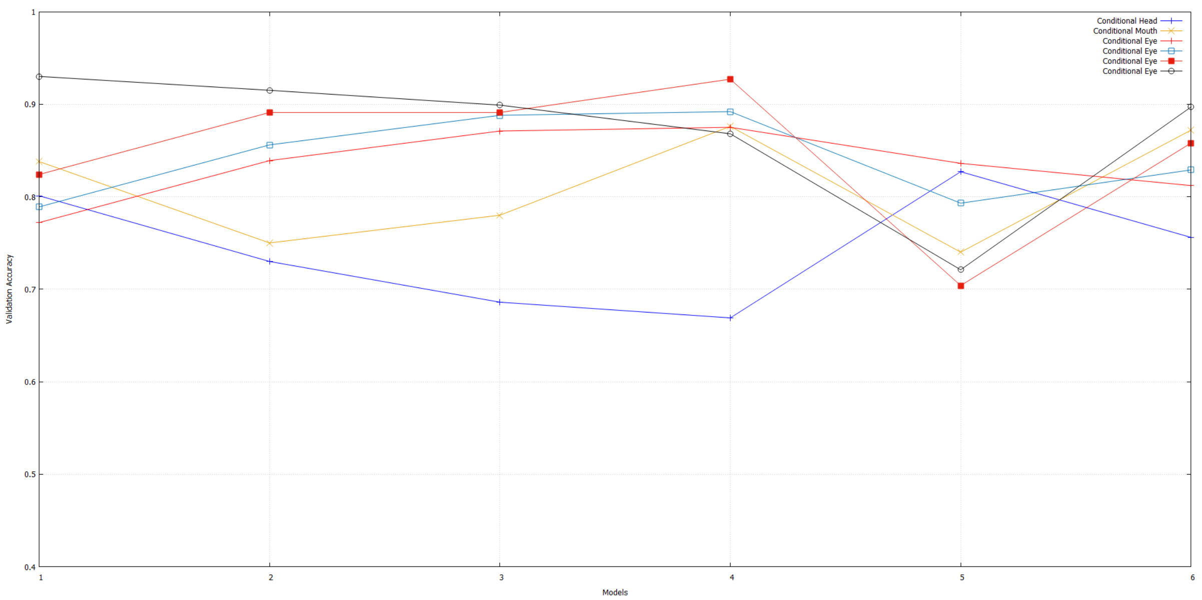

Fig. 6. Avg Validation Accuracy using KEC DDD Dataset by comparing various models

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.

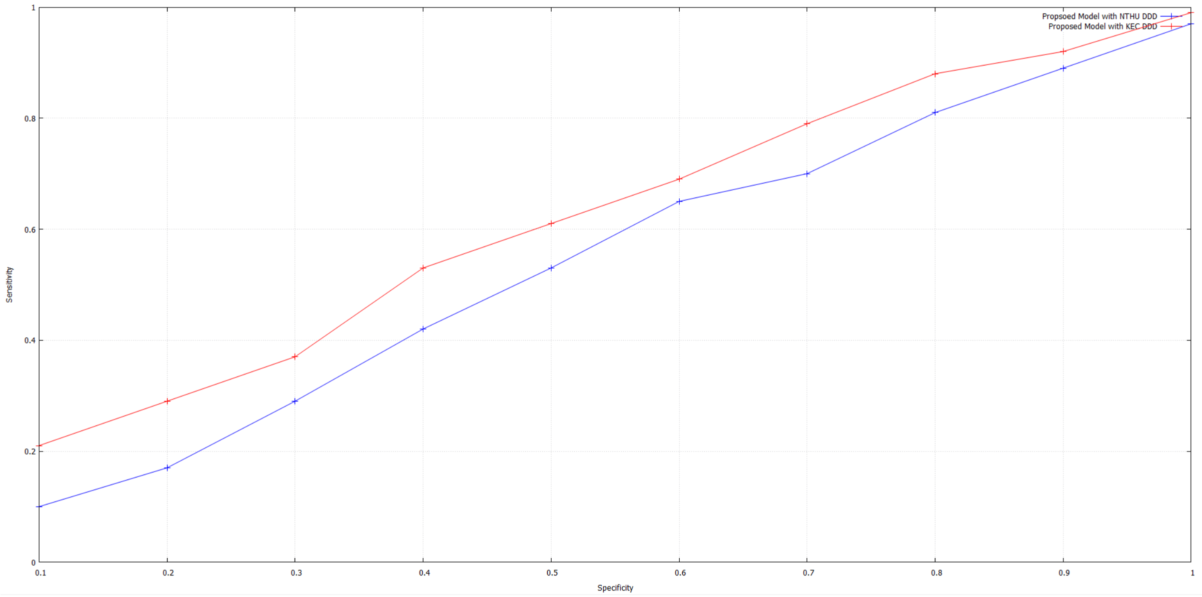

Fig. 7. AUC ROC using KEC and NTHU & DDD Dataset for the proposed model

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 explored sub-models to identify and classify driver sleepiness indicators. Different 3D CNN architectures are used to collect facial characteristics for each sub-model: head, eye, mouth, and glasses/normal. To detect head nodding, eyebrow lifting, and sideways staring. Mouth condition model detects open and covered yawning. All models are crucial to the system, addressing different aspects of driver drowsiness through specialized detection algorithms.

The integration of four sub-model outputs into a combined model, considerably improving sleepiness detection accuracy and reliability. Utilizing each model's strengths, the system analyses the driver's state, integrating drowsy and non-drowsy results. The sub-models' performance was extensively examined, analysing their robustness and adaptability to various real-world settings. Dataset validation findings confirm the sub-models' capacity to accurately capture driver behaviours in various contexts. This stage generates the mid-level driver drowsiness state from the KEC-DDD dataset, giving an important intermediate assessment for further research.

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.