KeAi
CHINESE ROOTS
GLOBAL IMPACT

International Journal of Innovation Studies

TRIPLE STAGE CNN MODEL TO DETECT THE DROWSINESS AMONG THE DRIVER'S BASED ON FACE POSITIONING TECHNIQUE

Nrusimhadri Naveen
PhD scholar, Department of Computer Science and Engineering, Shri Venkateshwara university, Gajraula,Uttar Pradesh.
Email:naveenbtechcse@gmail.com

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:
Generally, we need strong detection tools to combat driver drowsiness, which is an increasing problem around the world. Deep Learning (DL) and Machine Learning (ML) methods, such as different types of Convolutional Neural Networks (CNNs), have been tried by researchers, but they haven't been able to be used in real-time situations because they are too hard to implement because they are so complicated to compute and don't work very well when tested. Regarding these restrictions, a multistage adaptive multi-level -CNN model has been developed specifically for Driver Drowsiness Detection (DDD), emphasizing the resolution of system complexity and performance challenges. The suggested system employs a multiple tier methodology for driver sleepiness detection, incorporating breakthroughs from multiple separate research stages. The findings illustrate the efficacy of the suggested model across various datasets. The proposed model outperforms previous models, such as YOLOv7-Face, MTCNN, AAMS, and Hyperface-Resnet, in terms of bounding box regression and facial point localization accuracy. The results demonstrate the proposed model's efficacy in facial positioning and the detection of initial driver drowsiness, rendering it helpful for improving road safety.
Keywords: Convolutional Neural Networks; Accuracy; Detection; Deep Learning; Drowsiness.

I. INTRODUCTION

Driver face positioning pertains to the precise detection and localization of a motorist'sexpression within an image sample or video sample setting taken from real-time applications. Thus, this entails pinpointing the precise organization of the face and its critical milestones such as the eyes, mouth, etc., within the pictorial input, notwithstanding fluctuations in illumination, viewpoints, and hindrances.

Facial posture is crucial for various explanations. Initially, it constitutes the essential prerequisite for somewhat ensuing study, including the identification of facial expressions, ocular movements, head orientation, and lip motions. The credibility of these studies would be substantially undermined without precise facial positioning. Furthermore, accurate facial alignment facilitates the proficient management of various environmental factors (e.g., meteorological conditions impacting image quality) and driver behaviours, guaranteeing that the system functions reliably in real-world scenarios.

Upon accurate facial detection, the system is capable of tracking critical markers of tiredness, including things like closing the eyes, blinking often, looking in a certain direction, and nodding her head. Prolonged eyelid closure and frequent yawning are indications of weariness, which can be diagnosed through the examination of the eyes and mouth, respectively. Likewise, alterations in head position may signify a decline in concentration or vigilance. By precisely tracking these face traits, the system can deliver dependable evaluations on the health. This evidence is necessary for exposure the algorithms, which combine data from numerous face markers to determine the driver's alertness or weariness.

The suggested system's novelty lies in its utilization of an adaptive multi-level CNN for accurate facial location and preliminary detection of drowsiness among the drivers'. It uses an image pyramid methodology and three lightweight networks (Input-Network, Process-Network, Output-Network) to progressively enhance face detection, bounding box prediction, and facial landmark localization, guaranteeing precise alignment using the Non-Intersection Over Union Suppression (NIOUS) technique. The results show that facial recognition and the early detection of driver fatigue are now more robust and precise.

II. RELATED WORKS

Face placement is a critical task in CPUvisualisation that requires identifying and localizing looks within acopy image. It is required for several requests, including facemaskacknowledgement, drowsydiscovery, and Augmented Reality (AR). Current advances in DL knowingly increased expressiveplacement precision and durability. To improve overall performance, multi-task learning frameworks optimize numerous related tasks at the same time. Attention processes improve feature extraction by focusing on important areas of the face, which makes recognition and localization more accurate. Methods that tackle variances and occlusions in datasets guarantee dependable performance in real-world situations, accommodating varied stances, lighting conditions, and partial visibility. These methodologies jointly enhance the efficacy of facial positioning in restricted settings.

Zhang et al. (2017) advance multi-task learning by concurrently addressing facial detection, landmark localization, and head pose estimation. The efficiency of each individual task is significantly enhanced by the collaborative learning method, which makes use of connections between tasks. Several benchmark datasets support this approach, showing that it works well for a wide range of facial analysis tasks. The primary problem lies in the intricate balance of optimizing numerous activities, necessitating meticulous adjustment and oversight of common representations in order to avoid task obstruction.

Likewise, the methodology employed by Zhang et al. (2014) for facial landmark detection utilizes deep multi-task learning to enhance precision and resilience. Multi-task erudition improves performance in tough settings, highlighting its promise for facial analysis. The method's resilience and efficacy in practical scenarios underscore the benefits the learning process that involves performing jobs simultaneously. However, like other multiple tasking methods, which adds more activities could make the optimizationmethod more difficult and need more extensive administration of collectivetopographies.

Zafeiriou et al. (2017) use a convolutional network (ConvNet) with a 3D face model to improve detection across diverse poses and occlusions. This outline architect effectively influences the combined assets of ConvNets and 3D models, attaining enhanced efficacy in face detection under difficult circumstances. The model's ability to handle posture and occlusion fluctuations is enhanced by merging 2D and 3D data. The amalgamation of 2D and 3D data can elevate the model's complexity, necessitating more advanced training and processing methodologies.

Deep Face Detection utilizing Mask R-CNN (He et al. 2017) modifies the Mask R-CNN framework for facial recognition, leveraging its instance segmentation features to accurately identify faces inside complex backdrops. This adaptation leads to substantial enhancements in superiority to conventional Face detection methodologies, especially in stimulating conditions. The replica's resilience and precision are improved by its capability to isolate specificexpressions from composite backgrounds. This strategy enhances discovery but might require supplementary modules or modifications to tackle jobs such as Milestone localization and posture approximation.

By contrasting these outlines, it is obvious that a diversity of techniques, such as multitask learning, attention regulation, and methods for dealing with occlusions and variations, greatly improve face analysis tasks' accuracy and resilience. Each method possesses distinct advantages and drawbacks, providing varied approaches for improving technologies for facemaskinvestigation.

III. PROPOSED MODEL BASED ON FACE POSITIONING

A system for detecting driver drowsiness, initiated by processing rarecontributiondescriptions. Images obtained from a camera directed at the driver's expression are converted into an image pyramid using bilinear interpolation, resulting in a fourfold reduction in size. This multi-scale methodology guarantees that the ensuing steps can proficiently manage faces of varying dimensions and orientations.

The system is fundamentally built on a triple stage CNN model, which is critical for regression and classification tasks. The CNN is designed for managing the w structured grid information, usually used in duplicateacknowledgment and sortingjobsengineered to analyze the image pyramid, extracting pertinent information essential for assessing the driver's condition. This model functions in three phases, enhancing feature extraction and assuring the acquisition of essential features for precise classification and regression.

The multiple phase CNN model splits into two separate components: the organisation head and the worsening head. The sorting head assigns to one of two categories: drowsy or non-drowsy. This two-foldarrangement employs attributes obtained by the CNN to ascertain outcomes according to defined sleepiness criteria. The regression head generates 14 values representing facial landmark locations, including X-coordinates for each eye, nose, mouth corner, and Y-coordinates for each eye, nose, mouth corner, and mouth corner.

The deteriorationproductions are exploited to produce a bounding image surrounding the spotted face. The concerned area inside the image can be accurately identified and delineated with the help of these values. The system is capable of construct a bounding box that encloses the face by using the regression data to identify the locations of the facial landmarks. This bounded image is formerlyloaded and ready for further work, making sure that the noticed face is segregated and prepared for the state considerate sub models. to look more closely. Simultaneously, driver tiredness is discovered by passing situations and integrated into the more advanced detection system.

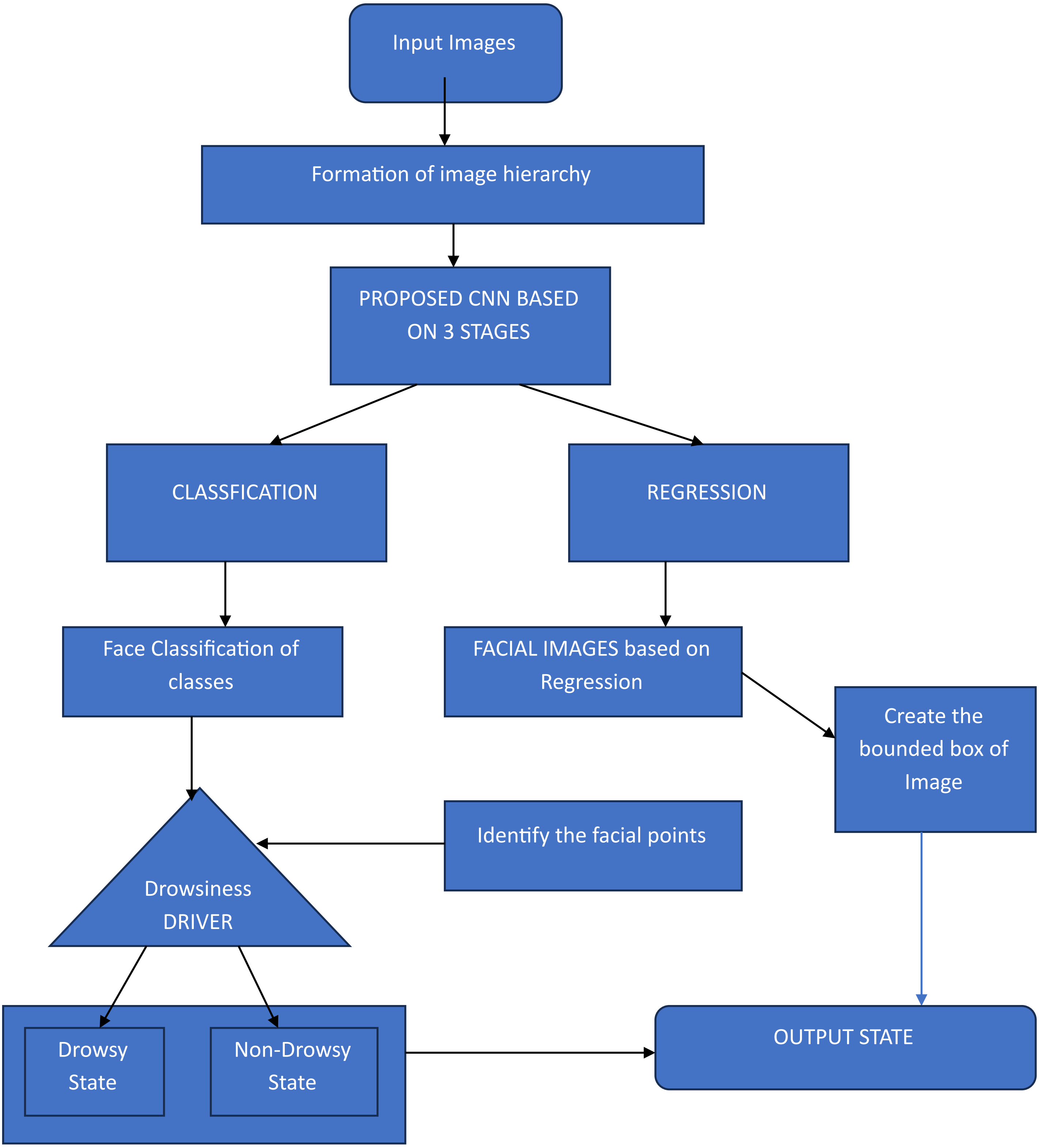

Figure 1. Architecture on Face Positioning for Drowsiness Detection

Algorithm: Detecting the Driver Drowsiness based on face position
Input: Input Images
Output: Drowsy (OR) Non-Drowsy State
Initialize the Input Images
Perform the data pre-processing
    Based on the input images,
        Generate and obtain Image hierarchy
Perform Initial Image detection on [Input-Network]
    Apply detection on Image hierarchy
        To identify the bounded images
Perform Image Refinement on [Process-Network]
    Apply P-Network in bounded images
        To refine the images
Perform Output on [Output-Network]
    Based on Final bounded images and landmark coordinates,
        To Obtain bounded images
Perform Detection on Driver Drowsiness
    Classify based on Drowsy (or) Non-Drowsy State
Obtain Drowsy State

A. Triple Stage CNN Mo

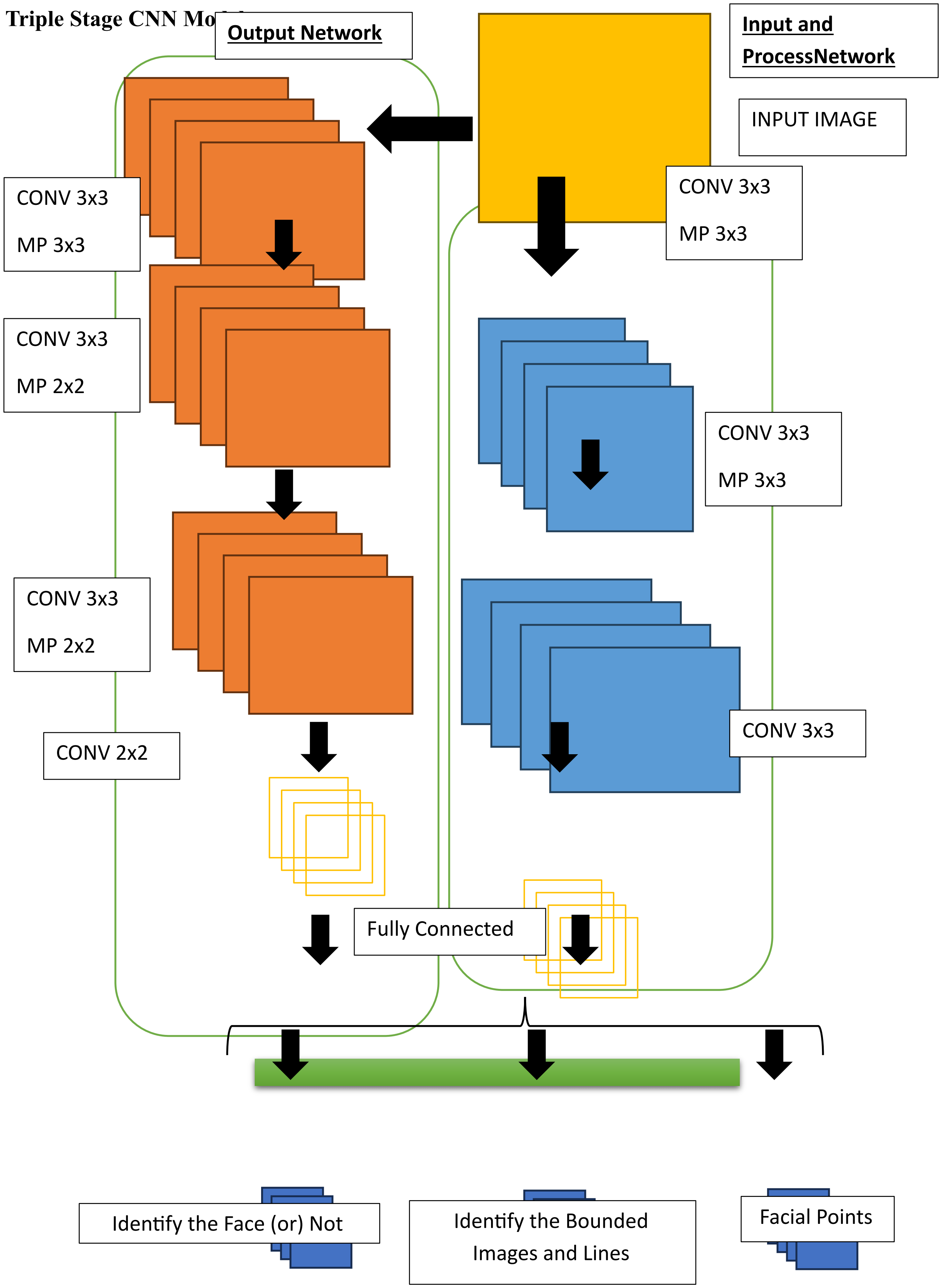

Fig. 2. Triple Stage CNN Model

The construction planning diagram depicts the design of Input-Net, Process-Net, and Output - Net, which are critical mechanisms of a multiple-stage CNN model utilized for expressiondiscovery and localization. Lightweight Input-Network and Process-Network have analikeconstruction and are intended to analysecontribution photos measuring 24x24x3. The systemstarts with a convolutional sheet that employs a 3x3 filter, trailedthrough a max-pooling sheetthat diminishes spatial dimensions while preserving essential properties. This is followed by an additional series of convolutional and max-pooling sheets that additional enhance feature maps. The last convolutional layer employs a 2x2 filter, reducing the dimensionality of the structures.

This creates a fully associated layer of 128 neurons, as seen in Fig. 3.2. Input-Network and Process-Network generate three separate sorts of productions, individually fulfilling a specific function. The first output is a 1x1x2 tensor indicating whether the identified sectioncomprises a face or not. This two-fold categorization aids in finding prospective facial sections within the image. The second output is a 1x1x4 tensor that denotes the leapingimagemanages for the identified face. These coordinates are essential for identifying the face inside the image. The third output is a 1x1x10 tensor that delineates the organizes of facemaskbreakthroughs for the eyes, nose, and corners of the mouth.

Output Network, procedureshigher input image descriptions of dimensions 48x48x3 and aims to improve the discoveries executed by Input-Network and Process-Network. The Output-Networkconstructioninitiates with a convolutional layer trailedthrough a max-pooling sheet, akin to the structures of Input-Network and Process-Network. Output-Networkcomprises an extra 36 layers to address the increased complexity of the input images. The network comprises numerous convolutional and max-pooling sheets that systematically decreasethree-dimensionalalthoughobtaining higher-level landscapes. Finally, the learned information is combined to create the final output via a fully linked layer consisting of 256 neurons.

O-Net produces three distinct outputs, akin to those of I-net and P-net. The initial output is a 1x1x2 tensor that indicates whether the identified section corresponds to a face. The second output is a 1x1x4 tensor that specifies the image bounded containerorganises, offeringdetailed localization of the face within the greater input image. A 1x1x10 tensor containing the coordinates of the facial landmarks is the result. O-Net's more intricate architecture enables it to offerextrapreciseimagebounded and facemaskbreakthrough locations, as well as better initial detections. This hierarchical process, in which O-Net relies on the discoveries of I-Net and P-Net, guarantees a healthy and accurate face discovery and localization.

IV. RESULT ANALYSIS AND DISCUSSION

This investigation compares the preparation and justificationcorrectness of different replicas for facemask point localization, image bounddeterioration, and expressionorganisationjobsutilizing two datasets: NTHU-DDD and KEC-DDD. In the Following graph Figures 3, 4, 5 and 6. As in the above graphs, models indicated as 1 to 7 are mentioned as,

1. YOLOv7-Face

2. MTCNN
3. AAMS
4. LOTR
5. Hyperface-Resnet
6. JFD-CCNN
7. Proposed Model

The endorsementprecision for image boundregressiondesignates the model's ability to accurately forecast the organises defining the image bound everywhere a face gets detected. The suggested prototypical performs better than the others in both datasets, attaining precision of 0.92 with KEC-DDD and 0.93 with NTHU-DDD.

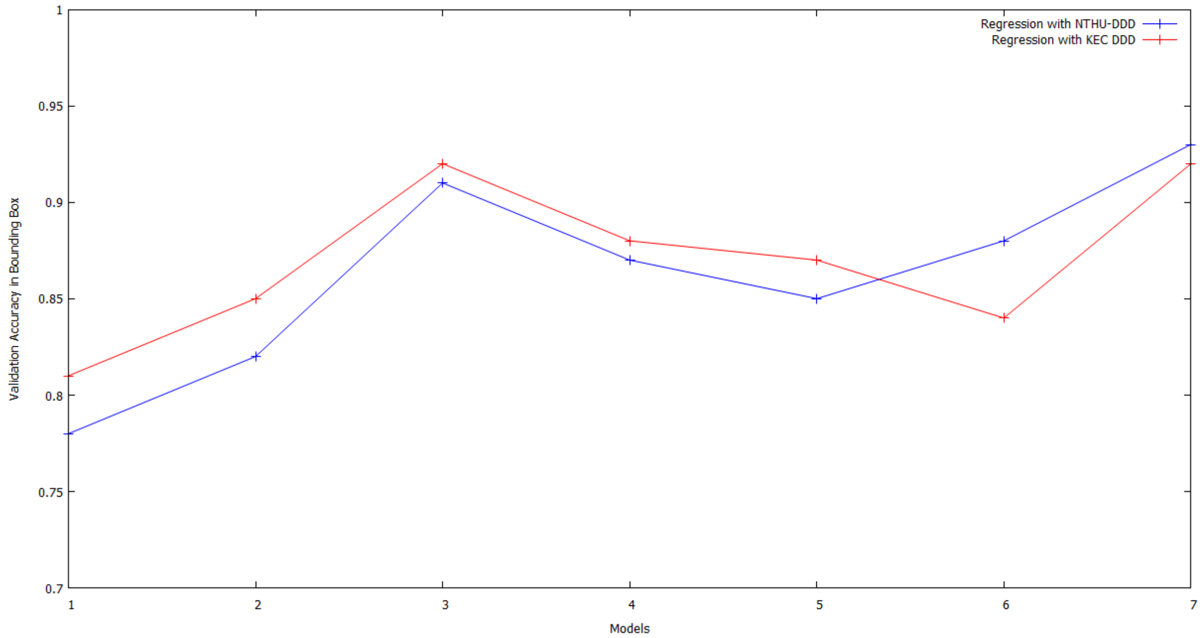

Fig. 3. Validation Accuracy

This constancy underscores its resilience and capacity to simplify effectively transversely many datasets. AAMS demonstrates commendable performance, throughprecisions of 0.91 and 0.92, indicating its status as a formidable competitor. MTCNN and YOLOv7-Face offer reasonabledevelopments in performance. Hyperface-ResNet and LOTR showmarginally diminished presentation, yet they preservemoderatelyextremeprecision levels. JFD-CCNN does well on NTHU-DDD, however it does not do as well on KEC-DDD.

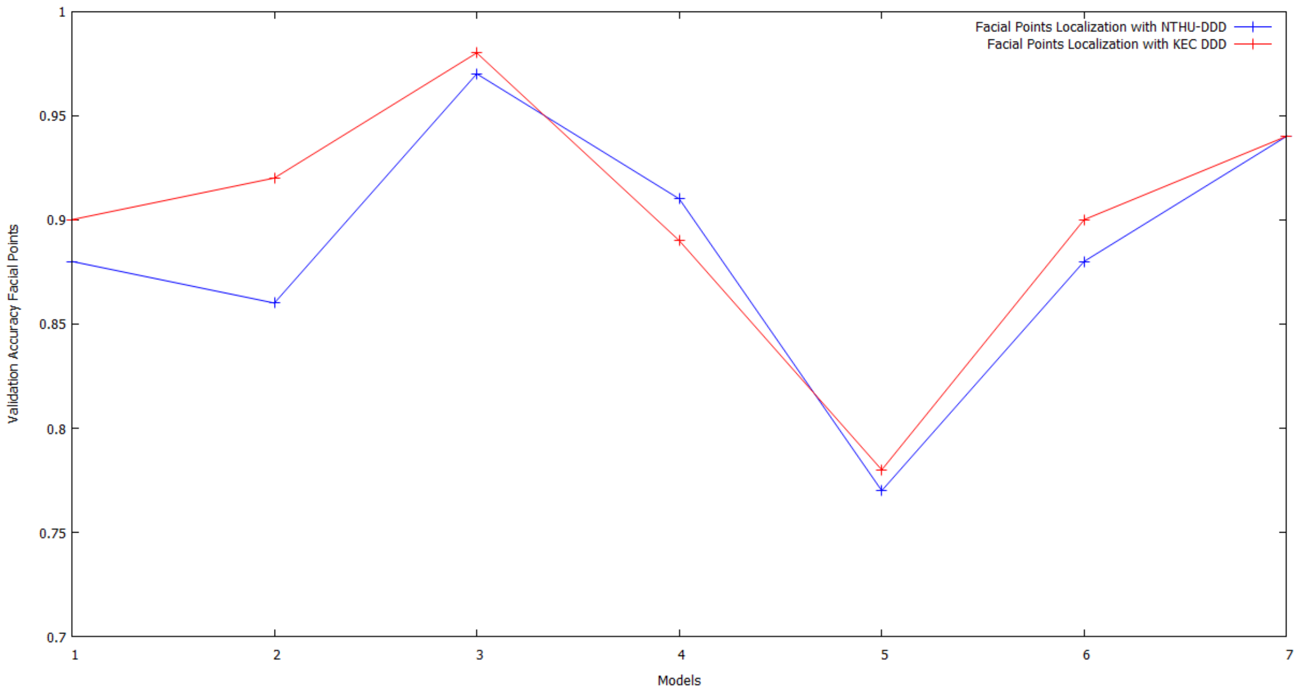

Fig. 4. Validation Accuracy

Face point localization accuracy measures how well the models can detect important face landmarks like the eyes, nose, and mouth corners.

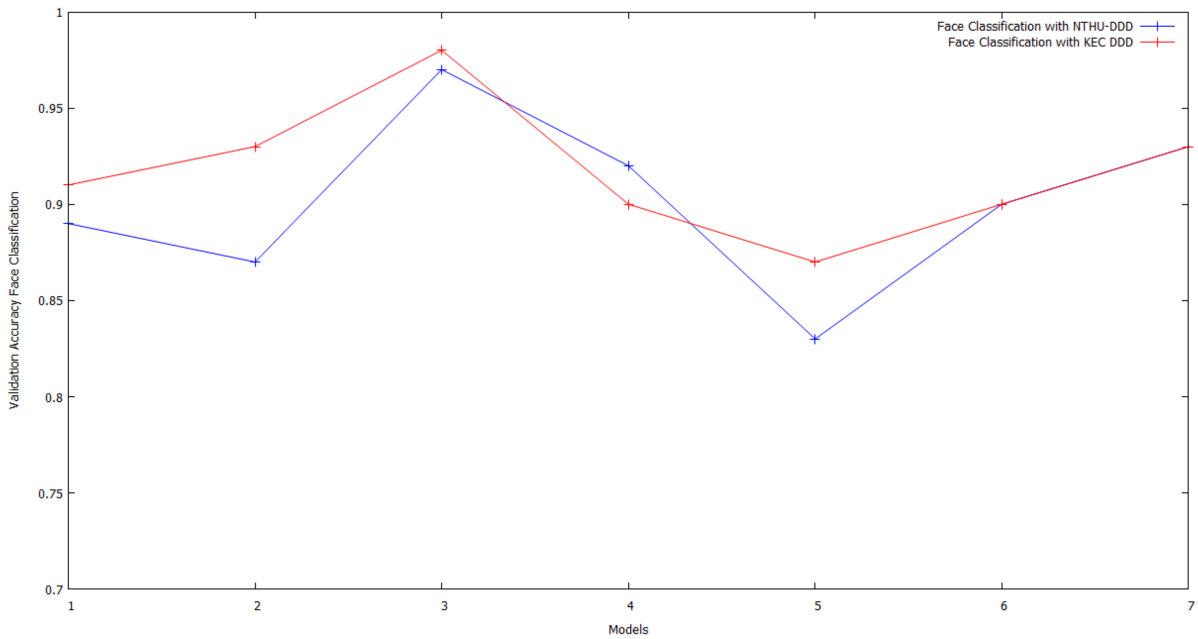

Fig. 5. Validation Accuracy

The suggested model attains an accuracy of The model achieved a score of 0.94 on both datasets, indicating its high proficiency in this task. AAMS demonstrates exceptional precision, attaining scores of 0.97 on NTHU-DDD and 0.98 on KEC-DDD, thereby positioning itself as the leading performer in facial point localization. MTCNN shows aimportantdevelopment from 0.86 to 0.92, demonstrating its flexibility to the KEC-DDD dataset. YOLOv7-Face and JFD CCNN exhibit strong performance, achieving accuracies ranging from 0.88 to 0.90, thereby indicating their reliability. LOTR demonstrates constant albeit marginally inferior performance, whereas Hyperface-Resnet exhibits the lowest accuracy, indicating its potential ineffectiveness for this task.

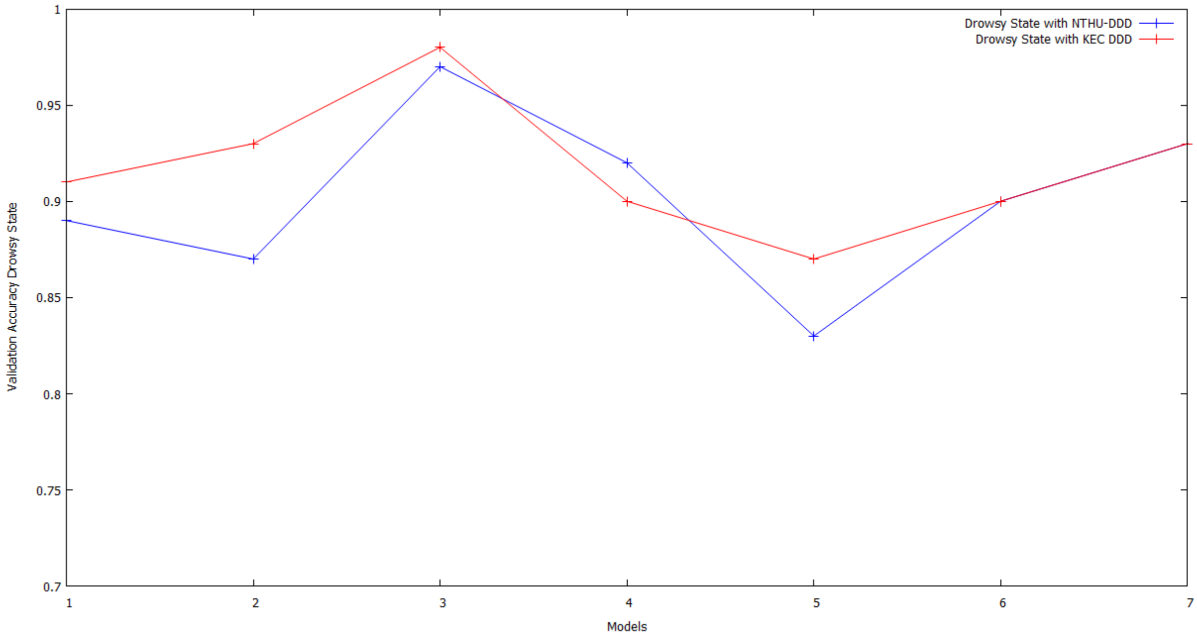

Fig. 6. Validation Accuracy

The proposed model demonstrates improved authenticationprecision, attaining 0.86 on NTHU-DDD and 0.87 on KEC-DDD compared to current models. AAMS exhibits robustroutine, achieving aendorsementprecision of 0.88 on both datasets. MTCNN and YOLOv7-Face demonstrate reliableroutine, upholdingcorrectness around 0.83. LOTR and 50 Hyperface-Resnet demonstrate consistent performance, though endorsementprecision is slightly reduced. JFD-CCNN shows aimportantenhancement from NTHU-DDD (0.77) to KEC-DDD (0.84), representingimprovedvariation to the KEC-DDD dataset. The exactness and recall standards are 0.88 and 0.86, respectively, when the prototypical is skilled with KEC-DDD, as indicated by the confusion matrix. The F1 metric of the model is 0.8686. This score signifies an elevated performance level, properly balancing precision and recall.

The efficiency of the suggested Three Stage CNN model in different face positioning and classification features is shown by the comparison analysis between these tables. The model continuously demonstrates elevated training and validation accuracy, surpassing competing models across multiple tasks.

V. CONCLUSION

Comprehensive methodology for identifying driver drowsiness via a three-tier CNN model. The organisationstarts with the dispensation of raw input photos to produce an image. Pyramid structure enables the extraction of features across multiple scales. The CNN model incorporates organisation and reversion components. The organisation head classifies the occurrence of a face, while the reversion head generates 14 standards that correspond to the organises of importantfacemaskbreakthroughs and image bound. The reversion outputs are critical for producing preciseimagebound around noticed faces and for

analyticaldetailedfacemasktopographies, such as the eyes, nose, and corners of the mouth. The constrained driving surfaces are then readied for additional examination.
The findings illustrate the efficacy of the suggested model across various datasets. The With validation accuracy of 0.93 and 0.94 on the NTHU-DDD and KEC-DDD datasets, respectively, the proposed model performed better than the current models, such as YOLOv7-Face, MTCNN, AAMS, and Hyperface-Resnet. The proposed model exhibited improved presentation in face categorization tasks, achieving an accuracy of 0.93 across both datasets. The endorsementprecision for detecting the original sleepiness condition of drivers demonstrated the replica'sstrength, achieving scores of 0.86 and 0.87 on the NTHU-DDD and KEC-DDD datasets, correspondingly. These outcomes confirm that the suggested model is effective at facial placement and the detection of the initial condition of driver drowsiness, rendering it helpful for improving road safety.

References:

[1] Ahmed, M, Masood, S, Ahmad, M & El-Latif, AAA 2021, 'Intelligent Driver Drowsiness Detection for Traffic Safety Based on Multi CNN Deep Model and Facial Subsampling', IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 10, pp. 19743 - 19752.

[2] Albadawi, Y, Takruri, M & Awad, M 2022, 'A Review of Recent Developments in Driver Drowsiness Detection Systems', Sensors,vol. 22, no. 2069, pp. 1-42.

[3] Chen, J & Chen, Q 2021, 'LOTR: Face Landmark Localization Using Localization Transformer', pp. 16530-16543. IEEE Access, vol. 10, no. 1,

[4] He, K, Zhang, X, Ren, S, & Sun, J 2016, 'Design of a Deep Face Detector by Mask R-CNN', In Proceedings of 27th Signal Processing and Communications Applications Conference, Sivas, Turkey, pp. 427-441.

[5] Hu, S, Zheng, G & Peters, B 2013, 'Driver fatigue detection from electroencephalogram spectrum after electrooculography artefact removal', IET Intelligent Transport Systems, vol. 7, no. 1, pp. 105-113.

[6] Chen, S, Wang, Z & Chen, W 2020, 'Driver Drowsiness Estimation Based on Factorized Bilinear Feature Fusion and a Long-Short-Term Recurrent Convolutional Network', Information, vol. 12, no. 3, pp. 1-15.

[7] Deng, J, Guo, J, Xue, N & Zafeiriou, S 2019, 'ArcFace: Additive Angular Margin Loss for Deep Face Recognition', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 10, pp. 5962 – 5979

[8] Doe, J & Smith, J 2023, 'Driver Drowsiness Detection Using Deep Learning', In Proceedings of 2nd International Conference on Smart Electronics and Communication, pp. 567-578.

[9] Zhang, K, Zhang, Z, Li, Z & Qiao, Y 2016, 'Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks', IEEE Signal Processing Letters, vol. 23, no. 5, pp. 1499-1503.

[10] Zhang, K, Zhang, Z, Li, Z & Qiao, Y 2019, 'AAN-Face: Attention Augmented Networks for Face Recognition', IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2179-2189.

[11] Dua, M, Shakshi, Singla, R, Raj, S &Jangra, A 2020, 'Deep CNN models-based ensemble approach to driver drowsiness detection', Neural Computing and Applications, vol. 33, no. 8, pp. 3155-3168.

[12] Ed-Doughmi, Y, Idrissi, N &Hbali, Y 2020, 'Real-Time System for Driver Fatigue Detection Based on a Recurrent Neuronal Network', Journal of Imaging, vol. 6, no. 8, pp. 1-15.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] Fernandez, S & Gomez, R 2020, 'Adaptive Drowsiness Detection System Using Machine Learning', Expert Systems with Applications, vol. 146, no. 5, pp. 1186-1197.

[18] Garcia, F & Martinez, J 2021, 'Real-Time Driver Drowsiness Detection on Mobile Platforms Using 3D Neural Networks', Neural Computing and Applications, vol. 32, no. 3, pp. 9731–9743.

[19] Guo, JM &Markoni, H 2018, 'Driver drowsiness detection using hybrid convolutional neural network and long short-term memory', Multimedia Tools and Applications, vol. 78, no. 4, pp. 29059-29087.

[20] Hashemi, M, Mirrashid, A & Beheshti Shirazi, A 2020, 'Driver Safety Development: Real-Time Driver Drowsiness Detection System Based on Convolutional Neural Network', SN Computer Science, vol. 1, no. 289, pp. 1-22.

[21] 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. 22.

[22] 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.

[23] Yu, J, Park, S, Lee, S & Jeon, M 2019, 'Driver Drowsiness Detection Using Condition-Adaptive Representation Learning Framework', IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 11, pp. 4206-4218.

[24] Zhang, T & Zhao, Y 2021, 'Drowsiness Detection for Drivers Using Convolutional Neural Networks and Neurocomputing, vol. 42, no. 3, pp. 157-166.