

A Triple-Stage CNN Framework For Early Driver Drowsiness Detection Using Facial Positioning

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:

In general, we need powerful detection technologies to fight driver fatigue, which is becoming a global issue. Researchers have experimented with Deep Learning (DL) and Machine Learning (ML) techniques, such as various Convolutional Neural Networks (CNNs), but they have not been able to be applied in real-time scenarios because they are too difficult to implement due to their high computational complexity and poor test results. In light of these limitations, a multistage adaptive multi-level CNN model has been created especially for Driver Drowsiness Detection (DDD), with a focus on resolving system complexity and performance issues. For the purpose of detecting driver sleepiness, the proposed system uses a multi-tier methodology that incorporates innovations from several distinct research phases. The results show how well the proposed model works on different datasets. When it comes to bounding box regression and facial point localisation accuracy, the suggested model performs better than earlier models like YOLOv7-Face, MTCNN, AAMS, and Hyperface-Resnet. The outcomes show how effective the suggested model is at identifying early driver drowsiness and face positioning, which makes it useful for enhancing traffic safety.

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

I. INTRODUCTION

The accurate identification and localisation of a driver's expression inside an image or video sample collected from real-time applications is known as driver face placement. Therefore, this means identifying the exact arrangement of the face and its important landmarks, like the lips, eyes, etc., within the visual input, despite variations in lighting, perspectives, and obstacles.

For a number of explanations, facial posture is important. First of all, it is a necessary precondition for certain further research, such as the recognition of facial emotions, eye movements, head orientation, and lip motions. Without accurate facial positioning, these studies' credibility would be seriously compromised. Additionally, precise face alignment makes it easier to effectively control a variety of environmental elements (such as weather that affects image quality) and driver behaviours, ensuring that the system operates dependably in real-world situations.

The device can follow important indicators of fatigue, such as closing her eyes, blinking frequently, staring in a certain direction, and nodding her head, once it has detected her face accurately. Examining the eyes and mouth, respectively, might reveal signs of fatigue, such as prolonged eyelid closure and frequent yawning. Changes in head posture might also indicate a loss of focus or alertness. The technology can provide trustworthy health assessments by accurately monitoring these facial characteristics. The algorithms that use information from multiple face markers to assess a driver's level of awareness or fatigue require this proof.

The innovation of the proposed system is its use of an adaptive multi-level CNN for precise facial positioning and early detection of driver fatigue. It gradually improves face detection, bounding box prediction, and facial landmark localisation using an image pyramid methodology and three lightweight networks (Input-Network, Process-Network, and Output-Network). The Non-Intersection Over Union Suppression (NIOUS) technique ensures accurate alignment. The findings demonstrate that early driver fatigue detection and facial recognition are now more reliable and accurate.

II. RELATED WORKS

Face placement, which involves locating and identifying looks inside a copied image, is a crucial CPU visualization task. Facemask acknowledgement, sleepy detection, and augmented reality (AR) are among the requirements that necessitate it. The precision and durability of expressive placement have been enhanced by recent developments in DL. Multi-task learning frameworks optimise multiple related tasks simultaneously to enhance overall performance. By concentrating on crucial facial regions, attention processes enhance feature extraction, improving recognition and localisation. Techniques that address dataset variations and occlusions ensure reliable performance in real-world scenarios, taking into account different postures, illumination, and partial visibility. When combined, these techniques improve the effectiveness of facial positioning in constrained environments.

By simultaneously tackling head posture estimation, landmark localisation, and facial detection, Zhang et al. (2017) improve multi-task learning. By utilising links between activities, the collaborative learning approach greatly increases the effectiveness of each individual activity. This method is supported by a number of benchmark datasets, demonstrating its effectiveness for a variety of facial analysis tasks. The main issue is the delicate balancing act of optimising many tasks, which requires careful modification and supervision of shared representations to prevent task blockage.

Similarly, deep multi-task learning is used in Zhang et al. (2014)'s face landmark detection technology to improve accuracy and robustness. The potential of multitask erudition for facial analysis is demonstrated by its ability to enhance performance in challenging environments. The method's adaptability and effectiveness in real-world situations highlight the advantages of learning while working concurrently. However, adding more tasks could make the optimisation process more challenging and need more thorough management of collective topographies, just like other multiple tasking techniques.

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 specific expressions 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 facemask investigation.

III. PROPOSED MODEL BASED ON FACE POSITIONING

A system that processes infrequent contribution descriptions in order to identify driver fatigue. Using bilinear interpolation, images captured by a camera aimed at the driver's expression are transformed into an image pyramid, resulting in a fourfold reduction in size. This multi-scale approach ensures that the subsequent stages can effectively handle faces of different sizes and orientations.

A triple stage CNN model, which is essential for regression and classification tasks, forms the foundation of the system. The CNN is intended to handle the structured grid data, which is typically utilised in duplication acknowledgement and sorting tasks designed to examine the image pyramid and extract relevant data necessary for evaluating the driver's condition. This model operates in three stages, improving feature extraction and guaranteeing the capture of crucial features for accurate regression and classification.

The organization head and the deteriorating head are the two distinct parts of the multiple phase CNN model. One of two categories—drowsy or not—is assigned by the sorting head. This two-fold design uses CNN-obtained attributes to determine results based on predetermined sleepiness criteria. X-coordinates for each eye, nose, and mouth corner, as well as Y-coordinates for each eye, nose, mouth corner, and mouth corner, are among the 14 values that the regression head produces to indicate the locations of facial landmarks.

The spotty face is surrounded by a boundary image created by taking use of the degrading productions. These values allow for the precise identification and delineation of the affected area within the image. By using the regression data to determine the locations of the facial landmarks, the system can create a bounding box that encloses the face. In order to ensure that the identified face is separated and ready for the state-conscious sub models to examine more deeply, this bounded image is previously imported and prepped for additional processing. Concurrently, passing scenarios identify driver fatigue and incorporate it into the more sophisticated detection system.

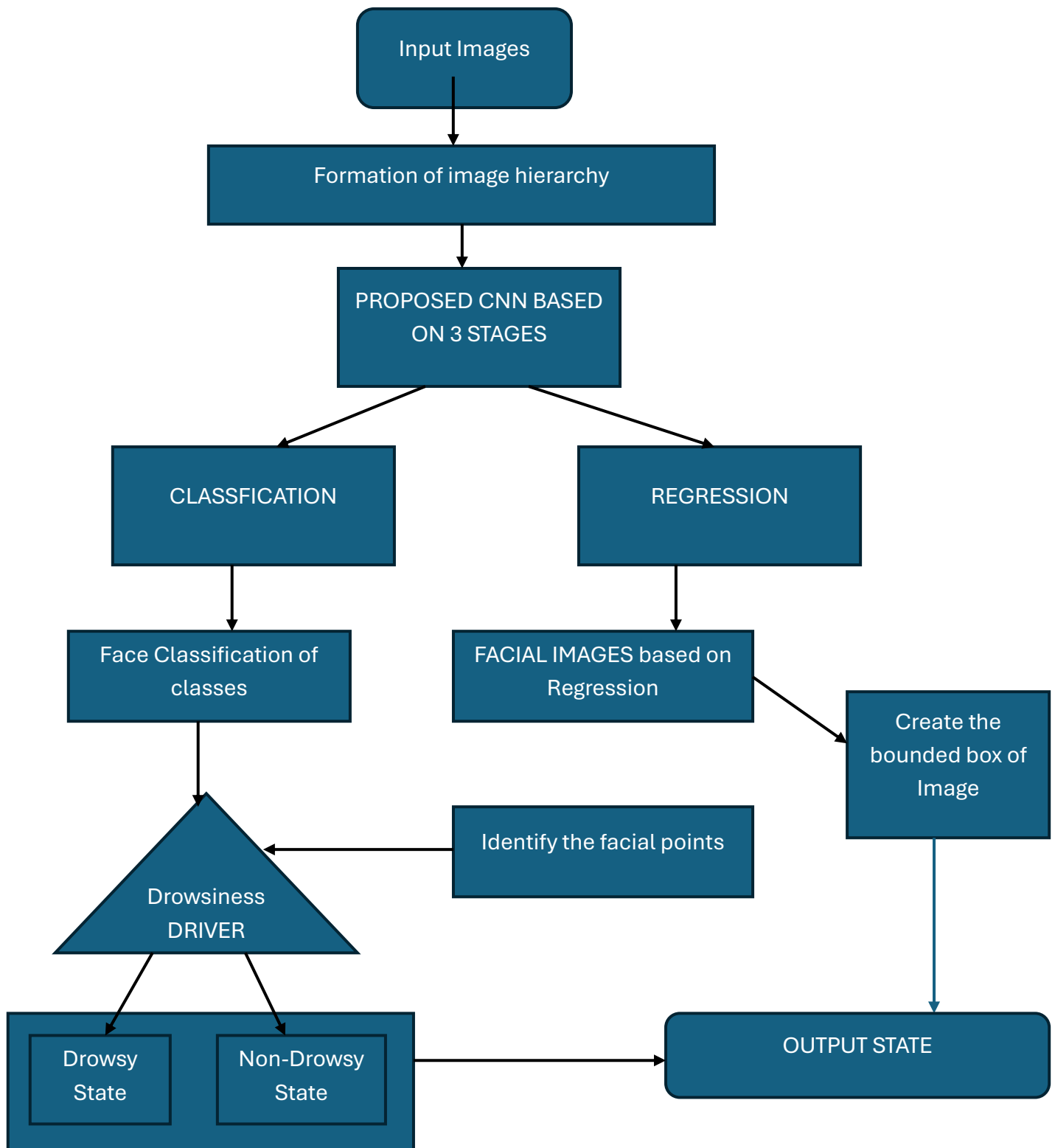


Figure 1. Architecture on Face Positioning for Drowsiness Detection

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 Model:

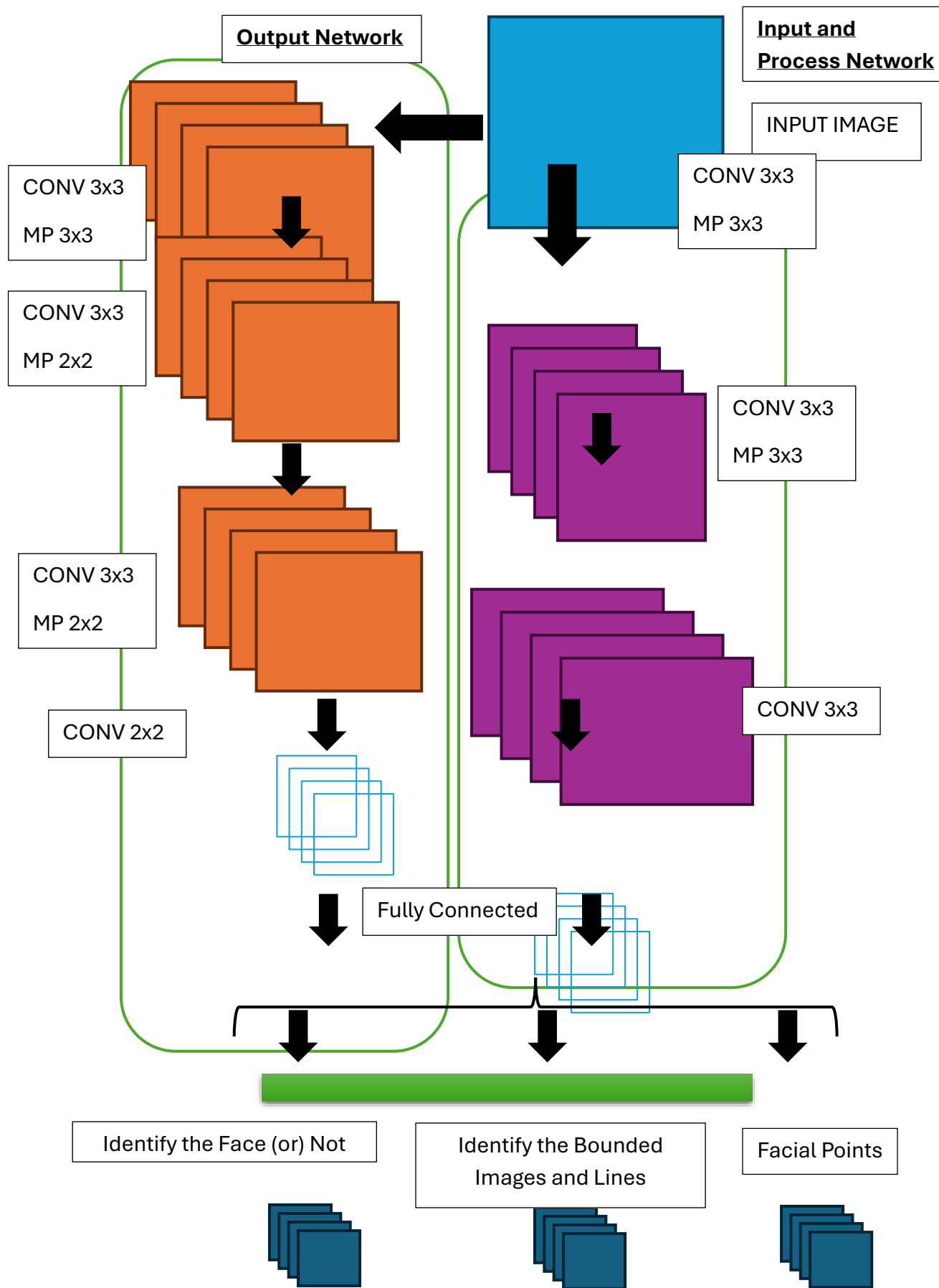


Fig. 2. Triple Stage CNN Model

The design of Input-Net, Process-Net, and Output-Net—three crucial components of a multi-stage CNN model used for expression finding and localization—is shown in the building planning diagram. Lightweight Input-Network and Process-Network are designed to evaluate 24x24x3 contribution pictures and share a similar construction. The system begins with a convolutional sheet that uses a 3x3 filter, followed by a max-pooling sheet that reduces spatial dimensions while maintaining key characteristics. Convolutional and max-pooling sheets that further improve feature maps come next. A 2x2 filter is used in the last convolutional layer to lower the structures' dimensionality. As seen in Fig. 3.2, this results in a fully coupled layer of 128 neurones. Three distinct types of productions are produced by Input-Network and Process-Network, each of which serves a particular purpose. A 1x1x2 tensor that indicates whether or not the specified area contains a face is the first output. This dual classification makes it easier to locate potential facial regions in the picture. A 1x1x4 tensor representing the leaping image manages for the recognised face is the second output. To identify the face inside the picture, these coordinates are crucial. The third output is a 1x1x10 tensor that shows how the eyes, nose, and mouth corners are arranged in facemask designs.

Higher input picture descriptions with dimensions of 48x48x3 are processed by the Output Network, which seeks to enhance the findings made by the Input-Network and Process-Network. Similar to the Input-Network and Process-Network designs, the Output-Network development starts with a convolutional layer followed by a max-pooling sheet. In order to handle the increased complexity of the input images, the output network has an additional 36 layers. Many convolutional and max-pooling sheets make up the network, which consistently reduces three dimensions while producing higher-level landscapes. Ultimately, a completely linked layer made up of 256 neurones combines the learnt information to produce the final output. Similar to I-net and P-net, O-Net generates three different outputs. A 1x1x2 tensor that shows whether the recognised portion matches a face is the first output. The second output is a 1x1x4 tensor that provides precise localisation of the face within the larger input image and describes the image bounded container organisations. The outcome is a 1x1x10 tensor with the coordinates of the facial landmarks. Because of its more complex design, O-Net can provide better initial detections and extremely accurate picture bounds and facemask breakthrough locations. A robust and precise face detection and localisation is ensured by this hierarchical approach, where O-Net depends on the findings of I-Net and P-Net.

IV. RESULT ANALYSIS AND DISCUSSION

This study uses two datasets, NTHU-DDD and KEC-DDD, to assess the preparation and justification correctness of several replicas for facemask point localisation, picture bound degradation, and expression organization tasks. Figures 3, 4, 5, and 6 in the graph that follows. Similar to the graphs above, models numbered 1 through 7 are described as

1. YOLOv7-Face
2. MTCNN
3. AAMS
4. LOTR
5. Hyperface-Resnet
6. JFD-CCNN
7. Proposed Model

The endorsement precision for image bound regression designates 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.

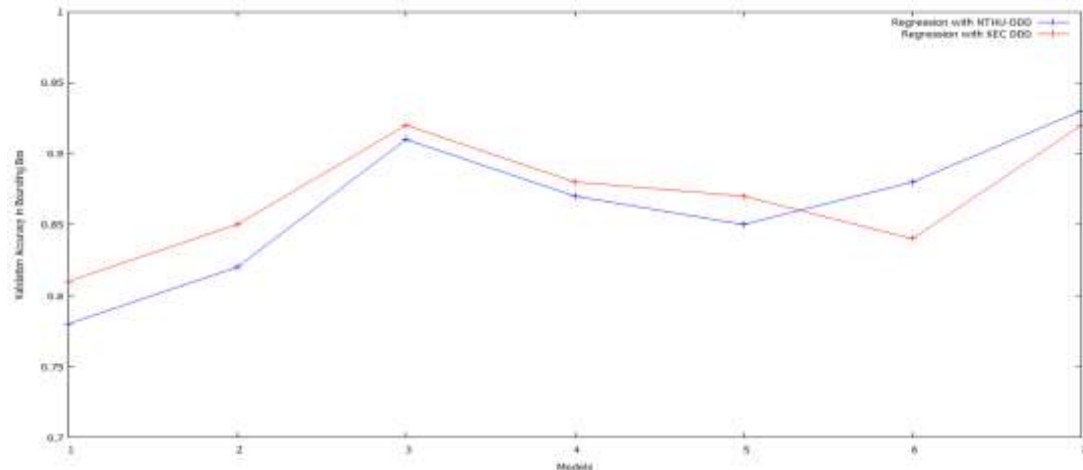


Fig. 3. Validation Accuracy

This consistency highlights its robustness and ability to efficiently simplify several datasets. With precision scores of 0.91 and 0.92, AAMS has excellent performance, demonstrating its position as a strong rival. MTCNN and YOLOv7-Face provide respectable performance improvements. Hyperface-ResNet and LOTR maintain somewhat severe precision levels but exhibit slightly reduced presentation. JFD-CCNN performs poorly on KEC-DDD, but it does well on NTHU-DDD.

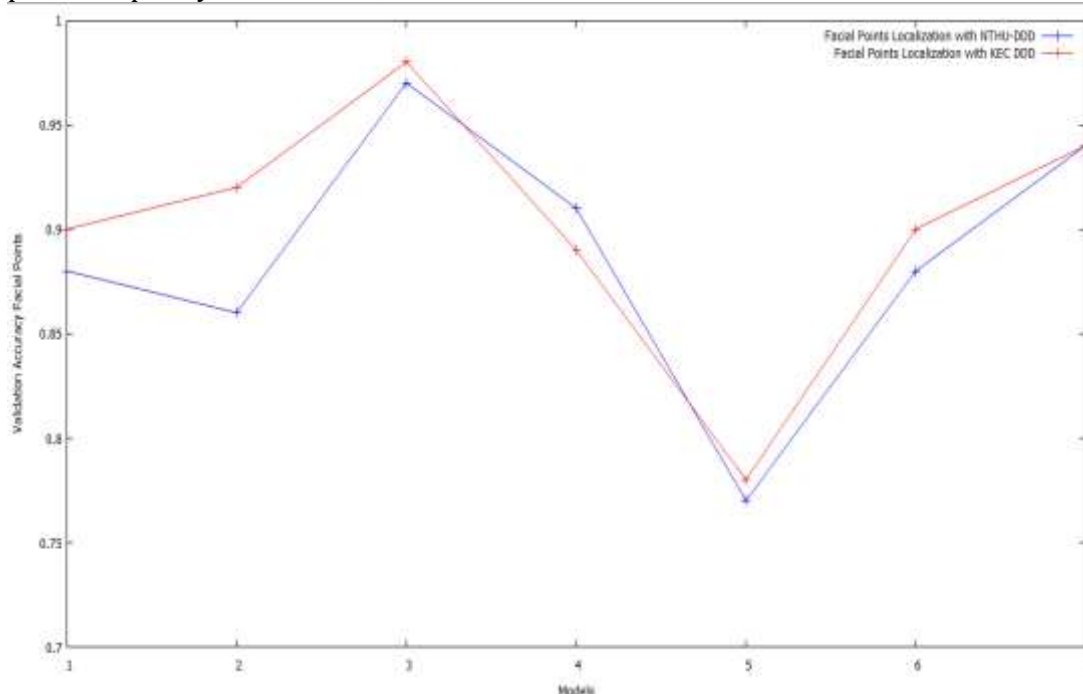


Fig. 4. Validation Accuracy

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

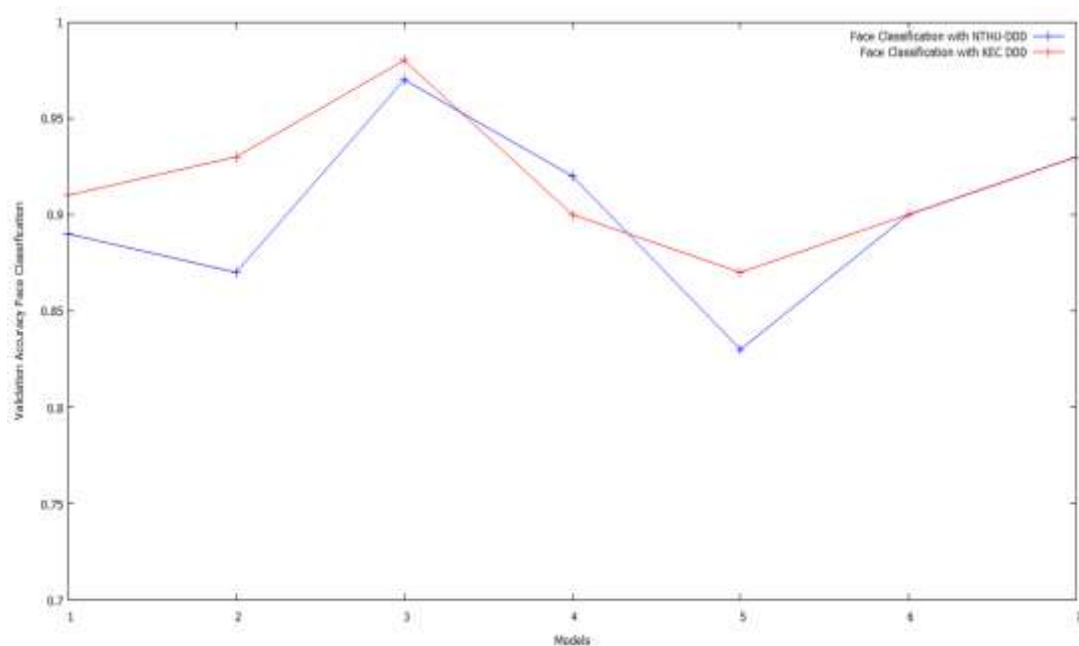


Fig. 5. Validation Accuracy

The recommended model achieves an accuracy of 0.94 on both datasets, demonstrating its excellent level of competence in this task. With scores of 0.97 on NTHU-DDD and 0.98 on KEC-DDD, AAMS exhibits remarkable precision and establishes itself as the top performer in facial point localisation. MTCNN demonstrates its adaptability to the KEC-DDD dataset with a significant improvement from 0.86 to 0.92. The reliability of YOLOv7-Face and JFD CCNN is demonstrated by its strong performance, with accuracies ranging from 0.88 to 0.90. Hyperface-Resnet has the lowest accuracy, suggesting that it may not be useful for this task, whereas LOTR consistently performs somewhat worse.

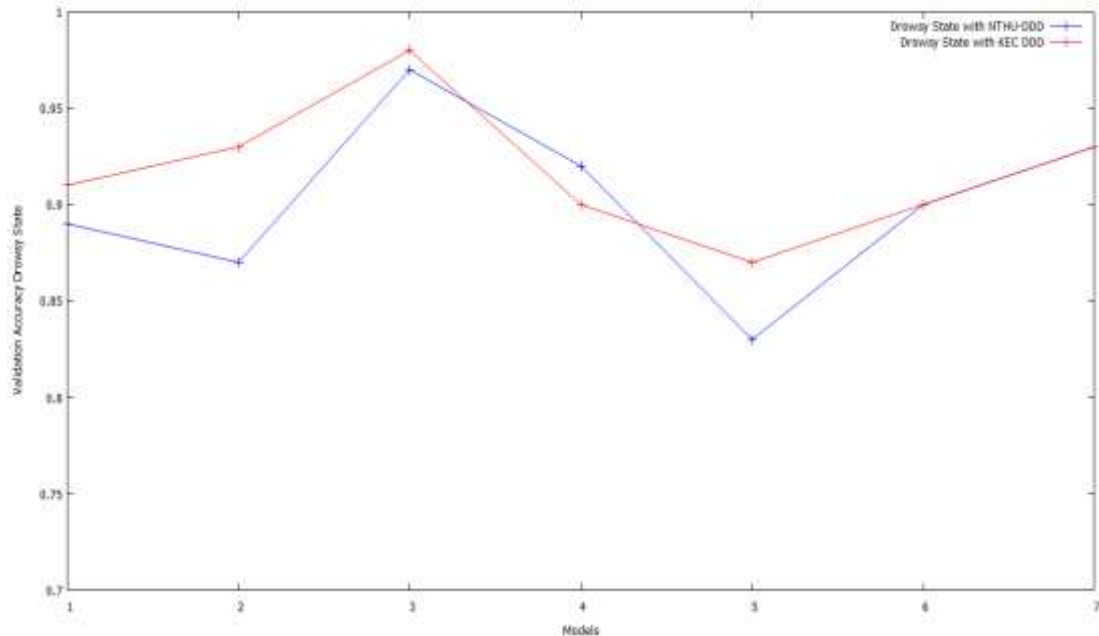


Fig. 6. Validation Accuracy

The suggested approach outperforms existing methods in terms of authentication precision, achieving 0.86 on NTHU-DDD and 0.87 on KEC-DDD. AAMS achieves an

endorsement precision of 0.88 on both datasets, demonstrating robustness. Reliability is demonstrated using MTCNN and YOLOv7-Face, maintaining accuracy around 0.83. Although endorsement precision is slightly lower, LOTR and 50 Hyperface-Resnet show consistent performance. From NTHU-DDD (0.77) to KEC-DDD (0.84), JFD-CCNN exhibits a significant improvement, indicating better variation to the KEC-DDD dataset. The confusion matrix shows that when the student is proficient with KEC-DDD, the exactness and recall criteria are 0.88 and 0.86, respectively. The model's F1 metric is 0.8686. This score indicates a high level of performance that appropriately balances recall and precision.

The comparison analysis between these tables demonstrates the effectiveness of the proposed Three Stage CNN model in various face positioning and categorisation aspects. The model consistently outperforms rival models on a variety of tasks, exhibiting high training and validation accuracy.

V. CONCLUSION

A three-tier CNN model provides a comprehensive mechanism for detecting driver drowsiness. The organization begins with the distribution of raw input photographs to create an image. The pyramid structure facilitates the extraction of features at many scales. The CNN model includes organisational and reversal components. The organization head categorises the occurrence of a face, whilst the reversion head creates 14 criteria that correspond to the organization of essential face mask breakthroughs and image binding. The reversion outputs are crucial for producing exact images bound around recognised faces as well as analytically detailed face mask topographies, such as the eyes, nose, and corners of the mouth. The confined driving surfaces are then prepared for further investigation.

The findings demonstrate the effectiveness of the proposed model across a variety of datasets. With validation accuracy of 0.93 and 0.94 on the NTHU-DDD and KEC-DDD datasets, respectively, the proposed model outperformed existing models such as YOLOv7-Face, MTCNN, AAMS, and Hyperface-Resnet. The suggested approach enhanced presentation in face categorisation tasks, with an overall accuracy of 0.93 across both datasets. The endorsement precision for detecting the original sleepy condition of drivers displayed replica strength, with scores of 0.86 and 0.87 on the NTHU-DDD and KEC-DDD datasets, respectively. These findings demonstrate that the proposed model is effective at face placement and detecting the early stages of driver tiredness, making it useful for increasing 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.