R4D - Real-time drowsiness and distraction detection system for drivers

Jathurshan Pathmarasa
Department of Computer Engineering
University of Sri Jayewardenepura
Colombo, SriLanka
jathursh11@gmail.com

***Abstract*—Ensuring road safety is essential, and mitigating driver drowsiness and distraction is important to enhance driver safety and attentiveness. In this paper, I propose a real-time Drowsiness and Distraction Detection System using transfer learning with the MobileNetV2 model. This system analyzes the driver's eye state and identifies various distraction behaviors through real-time visual data, enabling timely warnings and interventions. A novelty in my approach is the integration of a combined dataset, including the Media Research Lab (MRL) eye dataset, Closed Eyes In The Wild (CEW) dataset, and Zheijang University (ZJU) eyeblink dataset for drowsiness detection. This combined dataset allows for a more comprehensive representation of different ethnicities, particularly in the context of eyes. For the distraction detection, the model uses the Statefarm driver distraction dataset with manual label error correction and uses data augmentation for increasing precision and accuracy. The drowsiness detection model achieves an overall accuracy of 96.93%, and the distraction detection model achieves an overall accuracy of 97.78%. Beyond just getting better accuracy, the system makes driving safer and shows a dedication to working well for different ethnicities. By combining different datasets, the approach establishes a new benchmark for addressing challenges in detecting drowsiness and distraction in real-world driving scenarios.**

***Index Terms*—Drowsiness detection, Distraction detection, Real-time detection, Transfer learning, Combined dataset**

Abbreviations and Acronyms

MRL	Media Research Lab
ZJU	Zheijang University
CEW	Closed Eyes In The Wild
EAR	Eye Aspect Ratio
MAR	Mouth Aspect Ratio

I. Introduction

In the context of vehicle safety, identifying drowsy drivers is important to prevent accidents. Due to the large number of people using vehicles for their daily travels and convenience, the roads tend to get crowded. However, drowsy driving is a major cause of accidents. To address this, it's important to have effective ways to detect and intervene early when drivers are getting sleepy.

Every year, more than 1.3 million people lose their lives in traffic accidents, and drowsy drivers play a big role in these accidents. To address this issue and reduce accidents, we need advanced technology that can detect drowsy drivers. This involves using cameras and sensors to identify signs of drowsiness early on, helping to prevent dangerous crashes.

Companies like Tesla and Mercedes-Benz have introduced driver assistance systems to enhance safety on the roads [1], but there's still a challenge in creating systems that can accurately detect drowsiness and alert drivers in time. Many studies have shown a connection between driver tiredness and accidents, and while it's hard to know the exact number, fatigue-related accidents are likely underestimated. Previous approaches to drowsiness detection have used machine learning algorithms, including Support Vector Machine (SVM), k-Nearest Neighbors (KNN), and Haar Cascade classifiers [2], [3].

Apart from the loss of life, property damage from traffic accidents is often caused by distracted driving. According to the American National Highway Traffic Safety Administration, about 20% of accidents in the United States happen because drivers are distracted, and 90% of these accidents occur due to human error [4]. Distractions can range from using a mobile phone to eating and drinking, talking to passengers, and more. It's important to monitor and address these distractions to reduce accidents [5], [6].

Several studies have explored different techniques, like using sound to detect mobile phone usage, tracking eye movements to identify distraction, and analyzing images to classify various driver actions. Distractions include activities such as eating, drinking, talking to passengers, using electronic devices, looking at roadside advertisements, and using a mobile phone [7]–[13]. Effectively dealing with distracted driving requires systems that can monitor and adapt to the driver's state.

Given these challenges and advancements, the proposed solution aims to significantly improve driver safety by introducing a real-time Drowsiness and Distraction Detection System. By combining diverse datasets and using transfer learning, the system aims to accurately identify drowsiness and distractions in real-time, providing timely alerts and interventions.

The motivation behind the work is from the rising number of accidents caused by driver inattentiveness. Drowsiness and distraction are major issues, and I believe that effective technological solutions for detection and intervention are important. In this innovative approach, a key feature is the use of a combined dataset, merging the MRL eye dataset, CEW dataset, and ZJU eyeblink dataset for drowsiness detection. Additionally, for distraction detection, I used the Statefarm driver distraction dataset with manual label error correction and data augmentation.

The combination of diverse datasets brings a unique element to the system, potentially improving its performance. Main objectives of this system are to make driving safer by accurately identifying drowsiness and distraction in real-time, to implement a solution that provides timely alerts and interventions based on eye state and distraction behaviors, and to increase adaptability across different ethnicities through the diverse dataset. Through these objectives, the work aims to make an impact on driver assistance systems, encouraging safer and more attentive driving experiences.

II. Related Work

In [15], Pattarapongsin et al. (2020) used a ResNet10-based deep neural networks for face extraction and calculated Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), and pose estimation using 3D location. EAR and MAR are calculated from facial landmarks to determine the eye closure and openness of mouth, respectively. They used the Face Detection Dataset and Benchmark (FDDB) dataset, achieving accuracy, with EAR at 88.31 % and MAR at 92.61%. While their method ensures precision in exceptional cases like occlusion (when drivers wearing masks or sunglasses), and non-frontal face detection, their reliance on single dataset may limit generalizability.

In [16], Jahan et al. (2022) explored VGG16, VGG19, and proposed a convolutional neural network (CNN) model called 4D for drowsiness detection with MRL eye dataset, achieving accuracies of 95.93%, 95.03%, and 97.50% for VGG16, VGG19, and 4D, respectively. While their lightweight approach is commendable, their focus on drowsiness and single dataset usage differs from my multi-modal detection and data diversification strategies.

In [17], Kayadibi et al. (2022) proposed an AlexNet-based deep convolutional neural network model for drowsiness detection using ZJU, CEW datasets and Viola-Jones detector algorithm to detect face and eye regions. They achieved 97.32%, and 97.93% with ZJU, and CEW, respectively. However, their reliance on pre-trained AlexNet potentially misses out on advancements in deeper architectures, and their system is not designed for real-time applications. The use of MobileNetV2 in my system addresses both real-time efficiency and accuracy.

In [18], AlShalfan et al. (2021) proposed a VGG16-based model for distraction detection with Statefarm distraction detection dataset, getting an accuracy of 96.95%. Their focus on driver distraction aligns with our objectives; however, the reliance only on the VGG architecture, with its large number of parameters, could limit its adaptability and introduce overfitting. My multi-modal approach and MobileNetV2 offer advantages in this regard.

Wang et al. (2023) [19] introduced the ResNet50 based multi-scale domain adaptation network (MSDAN), addressing generalization challenges across datasets for distracted driving. The MSDAN model achieved accuracies of 96.82%, and 94.30% with Statefarm, and AUC distracted driver dataset, respectively. But their choice of ResNet50 maybe computationally expensive for real-time applications. My combined dataset strategy prioritizes both accuracy and generalizability within a real-time framework.

My approach introduces a real-time Drowsiness and Distraction Detection System using transfer learning with the MobileNetV2 model. One of the key highlights is the utilization of a combined dataset (MRL + CEW + ZJU) for drowsiness detection and the Statefarm dataset with manual label error correction and data augmentation for distraction detection, demonstrating a novel approach to diversity in dataset integration, and contributing to an increase in precision (through label error correction), while data augmentation significantly increases overall accuracy and reduce overfitting.

III. Methodolgy and Materials

A. Datasets

1) MRL eye dataset: MRL Eye Dataset is the large-scale dataset of human eye images (Closed and Open eye images) which has 84,898 images. This dataset contains infrared images in low and high resolution, all captured in various lightning conditions and by different devices. It contains data from 37 individuals, including 33 men and 4 women, having individuals both with and without glasses and displaying different reflections. It captures the eye images using three different sensors: the Intel RealSense RS 300 sensor with a resolution of 640 x 480, the IDS Imaging sensor with a resolution of 1280 x 1024, and the Aptina sensor with a resolution of 752 x 480 [20].

2) ZJU eyeblink dataset: ZJU eyeblink dataset has 20 different subjects, having individuals both with and without glasses. The dataset has variations in lighting conditions, capturing eye images under normal light settings. This dataset includes 7000 instances of open eyes and 1984 instances of closed eyes, all at a resolution of 24 x 24 dimensions [21].

3) CEW dataset: The CEW dataset includes diverse environmental conditions, including variations in light, blur, and darkness. The dataset has data from 2423 subjects and provides eye images captured at a resolution of 24 x 24. This dataset has 2462 instances of open eyes and 2384 instances of closed eyes [21].

4) Statefarm distraction detection dataset: The StateFarm distraction detection dataset used in this system was published on Kaggle for a competition. This dataset includes nine classes of dashboard images containing 26 subjects. The classes include the following driver behaviors: safe driving, texting - right hand, talking on the phone - right hand, texting - left hand, talking on the phone - left hand, drinking, operating the radio, reaching behind, hair and makeup, and talking to passengers. There are around 2000 RGB images for each class and each image's resolution is 640 * 480 pixels [22].

B. Data Preprocessing

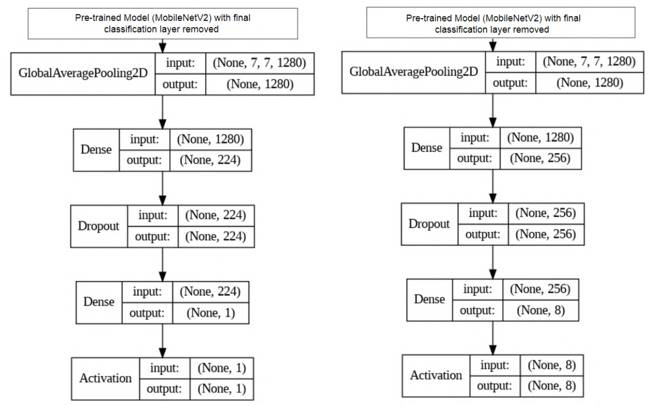
For the drowsiness detection model, a combination of the MRL, ZJU, and CEW datasets was used. Due to resource limitations and memory constraints, only half of the MRL eye dataset was used, while all images from the ZJU and CEW datasets were included, ensuring a balance between open and closed eyes classes. First the images in combined dataset were resized (224*224) to the size needed by the MobileNetV2 model and saved as features and labels. Then features and labels were divided into training and test sets (including a validation set - 50:50), maintaining a ratio of 70:30. The train and test sets were then normalized using MobileNetV2's preprocess input method.

For the distraction detection model Statefarm distraction detection dataset was used with label error correction by manually reviewing each image within every class. Initially, the dataset had nine classes, including specific classes for texting with both the right and left hand, as well as talking on the phone with the right and left hand. I combined texting behaviors into a single class (texting) and similarly combined talking on the phone behaviors into another class (talking). This was done to enhance the model's ability to detect distracted driving behaviors involving both left and right hands (In some countries, driving occurs on the left side of the road, and in other countries, it occurs on the right side). Then, data augmentation is done using random horizontal flipping, brightness adjustments, and contrast changes. The rest of the procedures, such as resizing, train-test split and normalization, are done similarly to the drowsiness detection model.

C. Model training and evaluation

MobileNetV2 is a lightweight deep neural network architecture designed for efficient on-device image classification. We can use it on devices with limited resources, such as embedded systems and mobile phones, for real-time applications due to its high speed and accuracy. MobileNetV2 uses a special type of convolution operation called depthwise separable convolution. This operation helps the model to be computationally efficient by significantly reducing the number of parameters and the amount of computation needed during training and inference [23]. I used MobileNetV2 model as the foundation for drowsiness and distraction detection system. In this network, the weights were initialized with the pre-trained ImageNet weights, followed by the application of the transfer learning concept.

Before training the models (for both drowsiness and distraction), I conducted hyperparameter tuning using the Keras Tuner (random search tuner) to find the best hyperparameters. The hyperparameters considered during tuning were the learning rate, whether to include a dense layer or not, the number of neurons in the dense layer (if the dense layer is included), and the dropout rate. The best hyperparameters found for the drowsiness detection model are 0.001, True, 224, and 0, respectively. For the distraction detection model, the best hyperparameters are 0.001, True, 256, and 0.4, respectively. Proposed model architectures for drowsiness and distraction detection are shown in Figure 1. The pre-trained MobileNetV2 model's final classification layer is replaced based on the required number of classes. Specifically, two classes are used for drowsiness detection, and eight classes are used for distraction detection. To enhance accuracy and reduce overfitting, an additional dense layer and dropout layer are introduced before the final classification layer. Also a global average pooling layer is added to reduce the spatial dimensions of the input feature maps before the final classification layer (Figure 1).



(a) Drowsiness detection (b) Distraction detection

Fig. 1: Model architectures

Both models (drowsiness and distraction) were trained using the best hyperparameters found. During training, an early callback was used to stop the training process at a certain point where the loss continuously starts to increase. Additionally, a learning rate scheduler was used to dynamically adjusts the learning rate during training, contributing to improved model convergence. The final classification layer activation functions were chosen based on the nature of the task. For binary classification tasks, such as drowsiness detection, the sigmoid activation function was used. On the other hand, for multi-class classification tasks like distraction detection, the softmax activation function was used. To train the models, binary crossentropy and categorical crossentropy were chosen as the loss functions for binary and multi-class classification, respectively and a batch size of 32 was used during training. The Adam optimizer is used in both models. After model training, the model is evaluated on the test set, and the confusion matrix and classification report are visualized based on the model's predictions. The classification report contains precision, recall, f1-score, and test accuracy.

D. Real-time detection

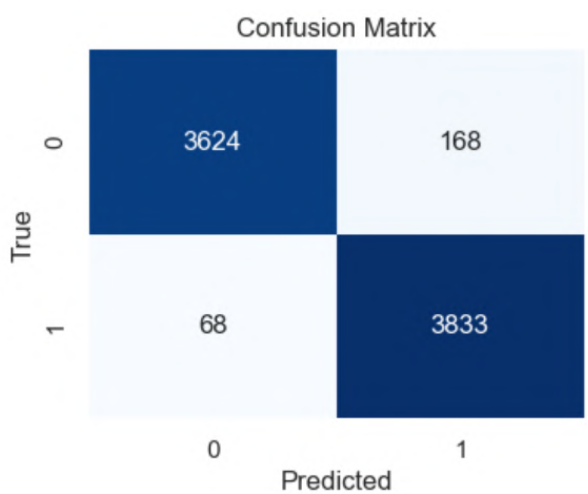
Before starting the real-time detection, the drowsinesss and distraction detction models were converted to tensorflow lite models (lightweight and efficient) with float 16 precision for faster processing.

For drowsiness detection, pre-trained HaarCascade classifiers for face and eye detection built inside OpenCV were used to obtain the region of interest. Then, I detected eyes (using cascade classifiers) inside the detected face, cropped the eye images from each frame captured by the camera in real-time, resized and normalized them, and let the model predict each frame. If the prediction indicates awake (eyes are open), no alarms are triggered. However, if the prediction indicates drowsiness (eyes are closed) and persists for 2 seconds, a warning is displayed, and the alarm is activated.

For distraction detection, no region of interest selection is performed. Frames captured directly from the camera in real-time are resized and normalized, and the model predicts each frame. If the prediction indicates safe driving, no alarms are triggered. However, if the prediction indicates any other distractions and persists for a second, a warning is displayed with the name of the distraction, and the alarm is activated.

IV. Results and Discussion

In this section, I evaluate the drowsiness and distraction models using the test sets created for each model. Confusion matrices and classification reports for the model evaluations are shown in Figures 2 and 3. The drowsiness detection model achieved an accuracy of 96.93%, while the distraction model achieved an accuracy of 97.78%.



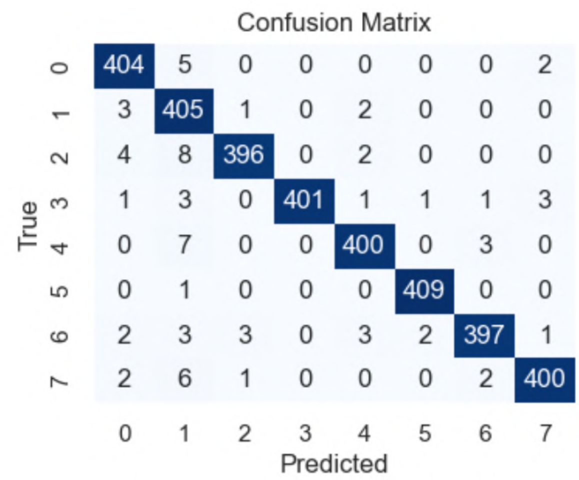
(a) Confusion matrix

	precision	recall	f1-score	support
0	0.9816	0.9557	0.9685	3792
1	0.9580	0.9826	0.9701	3901
accuracy			0.9693	7693
macro avg	0.9698	0.9691	0.9693	7693
weighted avg	0.9696	0.9693	0.9693	7693

(b) Classification report

Fig. 2: Drowsiness detection evaluation results
Class labels: 0 - Closed eye, 1 - Open eye

For the drowsiness detection model, the classification report (Figure 2(b)) indicates strong performance in distinguishing between open and closed eyes. With precision, recall, and F1-score values exceeding 95%, the model effectively identifies instances of drowsiness. The confusion matrix (Figure 2(a)) shows the reliability of the model, with minimal false positives and negatives.



(a) Confusion matrix

	precision	recall	f1-score	support
0	0.9712	0.9830	0.9770	411
1	0.9247	0.9854	0.9541	411
2	0.9875	0.9659	0.9766	410
3	1.0000	0.9757	0.9877	411
4	0.9804	0.9756	0.9780	410
5	0.9927	0.9976	0.9951	410
6	0.9851	0.9659	0.9754	411
7	0.9852	0.9732	0.9792	411
accuracy			0.9778	3285
macro avg	0.9783	0.9778	0.9779	3285
weighted avg	0.9783	0.9778	0.9779	3285

(b) Classification report

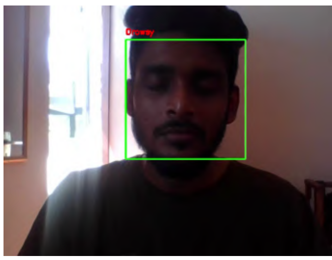
Fig. 3: Distraction detection evaluation results
Class labels: 0 - Normal Driving, 1 - Texting, 2 - Talking on the Phone, 3 - Operating the Radio, 4 - Drinking, 5 - Reaching Behind, 6 - Hair and Makeup, 7 - Talking to Passenger

By looking at the classification report for the distraction detection model (Figure 3(b)), we can observe strong and consistent precision, recall, and F1-score values across all classes. The model shows high accuracy in identifying different distractions, including texting, talking on the phone, and other potential distractions. The confusion matrix (Figure 3(a)) provides a deeper understanding, giving minimal misclassifications and a well-balanced performance across different classes.

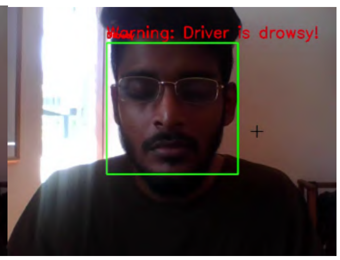
I detected drowsiness and distractions using the models in real-time through cameras, using the mobile phone's front camera for distraction detection and the laptop's front camera for drowsiness detection. Screenshots of real-time detections are shown in Figures 4 and 5.

For drowsiness detection, if the eyes remain closed continuously for a threshold of 2 seconds, it is considered as drowsiness, triggering a warning and alarm. Figure 4 shows how the model detects drowsiness with and without glasses.

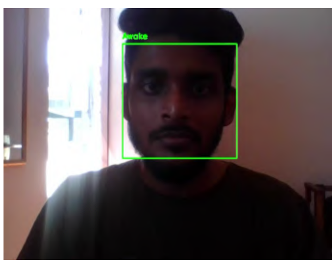
Similarly, for distraction detection, continuous distraction for a threshold of 1 second results in a warning and alarm (Figure 5). Some distractions, like Hair and Makeup and operating the radio, exhibited fluctuations in correct predictions when captured from specific camera positions. For instance, Figure 5 (a)-(h) were taken from the same position and predicted correctly, while Figure 5 (i) was captured from a different position and correctly predicted the Hair and Makeup distraction. However, in the position of Figure 5 (a)-(h), the model did not predict Hair and Makeup correctly. Identifying the correct camera position enables the model to accurately predict distractions. However, if the optimal position is not determined, it may lead to limitations such as a restricted field of view, potentially missing some distractions.



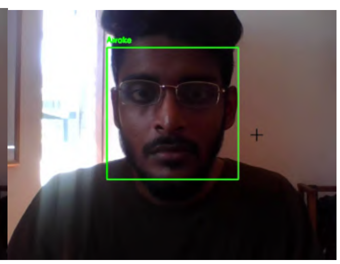
(a) Drowsy without glasses



(b) Drowsy with glasses



(c) Awake without glasses



(d) Awake with glasses

Fig. 4: Real-time detection screenshots for drowsiness detection

V. Conclusion

This paper proposes a real-time Drowsiness and Distraction Detection System using transfer learning with the MobileNetV2 model, aiming to enhance driver safety and attentiveness. The usage of a combined dataset (MRL, CEW, and ZJU) for drowsiness detection, ensures a comprehensive representation of different ethnicities, particularly in the context of eyes. For distraction detection, the Statefarm driver distraction dataset is used with manual label error correction and data augmentation for increasing precision and accuracy.

The drowsiness detection model achieved an accuracy of 96.63% and the distraction detection model achieved an accuracy of 97.7%. Unlike previous approaches, the paper uses a multi-modal detection approach and highlights diversity in dataset integration, contributing to increased precision and accuracy. Real-time experiments conducted through front cameras on a mobile phone and a laptop shows the system's practical applicability. While existing models may excel in certain aspects, the integration of MobileNetV2 ensures a balance between real-time efficiency and accuracy.

(a) Texting

(b) Texting

(c) Talking on phone

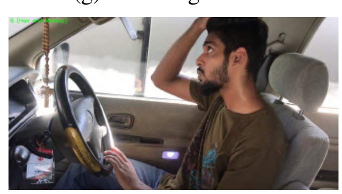
(d) Talking to passenger

(e) Drinking

(f) Operating radio

(g) Reaching behind

(h) Normal driving



(i) Hair and makeup

Fig. 5: Real-time detection screenshots for drowsiness detection with predicted labels

The real-time detection results presented in Figures 4 and 5 shows the system's capability to identify drowsiness and distractions without delay. The threshold-based warning and alarm triggers provide timely interventions, contributing to road safety. However, a few challenges and limitations also exist in this system.

Fisrtly, the distraction detection model has a limited field of view because predicting distractions accurately for varying camera positions is difficult. Therefore, finding the optimal camera position or setup is important.

Secondly, the system may face challenges when the device or system used for detections has slow processing power. This may lead to delays in triggering alerts and warnings. Detecting with a device with good processing power is important.

Thirdly, mispredictions of these models may lead to false warnings or missed interventions. Improving the model's robustness could minimize this limitation.

For future work, we can analyze driving patterns and behaviors. Insights derived from a comprehensive examination of driver actions, reactions, and environmental factors could improve the system's understanding of complex driving scenarios.

REFERENCES

[1] Jabbar, R.; Shinoy, M.; Kharbeche, M.; Al-Khalifa, K.; Krichen, M.; Barkaoui, K. Driver Drowsiness Detection Model Using Convolutional Neural Networks Techniques for Android Application. In Proceedings of the ICIoT 2020, Doha, Qatar, 2–5 February 2020; pp. 237–242

[2] Savaş, B.K.; Becerikli, Y. Real time driver fatigue detection based on SVM algorithm. In Proceedings of the 2018 6th International Conference on Control Engineering & Information Technology (CEIT), Istanbul, Turkey, 25–27 October 2018; pp. 1–4.

[3] Jalilifard, A.; Pizzolato, E.B. An efficient K-NN approach for automatic drowsiness detection using single-channel EEG recording. In Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; pp. 820–824.

[4] Testimony of The Honorable David L. Strickland. (2013). "Administrator national highway traffic safety administrationitle," House Committee on Transportation and Infrastructure Subcommittee on Highways and Transit, vol. 10, no. 6, pp. 1–8.

[5] G. M. Fitch, S. A. Soccolich, F. Guo, J. McClafferty, Y. Fang et al. (2013). , "The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk," National Highway Traffic Safety Administration, vol. 34, no. 3, pp. 23–33.

[6] C. Streiffer, R. Raghavendra, T. Benson and M. Srivatsa. (2017). "DarNet: A deep learning solution for distracted driving detection," in Proc. IMC, Las Vegas Nevada, USA, pp. 22–2

[7] M. S. Young, J. M. Mahfoud, G. H. Walker, D. P. Jenkins and N. A. Stanton. (2008). "Crash dieting: The effects of eating and drinking on driving performance," Accident Analysis and Prevention, vol. 40, no. 1, pp. 142–148.

[8] S. Koppel, J. Charlton, C. Kopinathan and D. Taranto. (2011). "Are child occupants a significant source of driving distraction?," Accident Analysis and Prevention, vol. 43, no. 3, pp. 1236–1244.

[9] A. Theofilatos, A. Ziakopoulos, E. Papadimitriou and G. Yannis. (2018). "How many crashes are caused by driver interaction with passengers? A meta-analysis approach," Journal of Safety Research, vol. 65, pp. 11–20.

[10] F. Zhang, S. Mehrotra and S. C. Roberts. (2019). "Driving distracted with friends: Effect of passengers and driver distraction on young drivers' behavior," Accident Analysis and Prevention, vol. 132, pp. 1–9.

[11] K. Shinohara, T. Nakamura, S. Tatsuta and Y. Iba. (2010). "Detailed analysis of distraction induced by in-vehicle verbal interactions on visual search performance," International Association of Traffic and Safety Sciences Research, vol. 34, no. 1, pp. 42–47.

[12] M. Krause, A. S. Conti, M. Henning, C. Seubert, C. Heinrich et al. (2015). , "App analytics: Predicting the distraction potential of in-vehicle device applications," in Proc. AHFE, Las Vegas, USA, pp. 2658–2665.

[13] J. Edquist, T. Horberry, S. Hosking and I. Johnston. (2011). "Effects of advertising billboards during simulated driving," Applied Ergonomics, vol. 42, no. 4, pp. 619–626.

[14] J. Edquist, T. Horberry, S. Hosking and I. Johnston. (2011). "Effects of advertising billboards during simulated driving," Applied Ergonomics, vol. 42, no. 4, pp. 619–626.

[15] P. Pattarapongsin, B. Neupane, J. Vorawan, H. Sutthikulsombat, and T. Horanont, "Real-time drowsiness and distraction detection using computer vision and Deep Learning," Proceedings of the 11th International Conference on Advances in Information Technology, 2020. doi:10.1145/3406601.3406638

[16] I. Jahan et al., "4D: A real-time driver drowsiness detector using Deep Learning," Electronics, vol. 12, no. 1, p. 235, 2023. doi:10.3390/electronics12010235

[17] I. Kayadibi, G. E. Güraksın, U. Ergün, and N. Özmen Süzme, "An eye state recognition system using transfer learning: AlexNet-based deep convolutional neural network," International Journal of Computational Intelligence Systems, vol. 15, no. 1, 2022. doi:10.1007/s44196-022-00108-2

[18] K. A. AlShalfan and M. Zakariah, "Detecting driver distraction using deep-learning approach," Computers, Materials & Continua, vol. 68, no. 1, pp. 689–704, 2021. doi:10.32604/cmc.2021.015989

[19] J. Wang and Z. Wu, "Driver distraction detection via multi-scale domain Adaptation Network," IET Intelligent Transport Systems, vol. 17, no. 9, pp. 1742–1751, 2023. doi:10.1049/itr2.12366

[20] "MRL Eye Dataset," MRL, http://mrl.cs.vsb.cz/eyedataset (accessed Dec. 18, 2023).

[21] Pattern Recognition and Neural Computing, http://parnec.nuaa.edu.cn/ (accessed Dec. 18, 2023).

[22] "State Farm distracted driver detection," Kaggle, https://www.kaggle.com/c/state-farm-distracted-driver-detection (accessed Dec. 18, 2023).

[23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018. doi:10.1109/cvpr.2018.00474