# A comparative analysis of lightweight deep learning models for driver distraction detection

[1],[2],[3]Gökhan Azizoğlu, [1],*Ahmet Nusret Toprak
[1]Erciyes University, Faculty of Engineering, Department of Computer Engineering
[2]Erciyes University, Graduate School of Natural and Applied Sciences, Computer Engineering Graduate Program
[3]Sivas Cumhuriyet University, Faculty of Engineering, Department of Software Engineering
*antoprak@erciyes.edu.tr

**Abstract:** Driver distraction is one of the leading causes of road accidents worldwide, making its accurate and efficient detection a critical component of intelligent transportation systems. While deep learning approaches have achieved strong performance in this domain, many state-of-the-art models are computationally expensive and unsuitable for deployment on embedded, real-time platforms. To address this challenge, this study presents a comparative analysis of lightweight deep learning models for driver distraction detection, focusing on the trade-offs between accuracy and computational efficiency. We evaluate ten popular lightweight architectures, which are MobileNetV2, MobileNetV3, MobileNetV4, ShuffleNetV2, SqueezeNet, EfficientNet-B0, ResNet18, MixNet, EfficientViT-B0, and GhostNetV2, using the AUC and the StateFarm distracted driver datasets. The models are assessed across key metrics, including accuracy, precision, recall, F1-score, inference speed (FPS), and parameter efficiency, with additional deployment experiments conducted on the NVIDIA Jetson Xavier platform. The results show that MixNet and ResNet18 achieve the highest accuracy on the more challenging AUC dataset, with 95.01% and 94.69%, respectively, while ShuffleNetV2 and MixNet reach near-perfect accuracy (99.96%) on the StateFarm dataset. In terms of efficiency, SqueezeNet delivers the fastest inference speed (116.61 FPS) with the smallest parameter count (0.74M), while MobileNetV4 provides a balanced trade-off between speed and accuracy. These findings highlight that the optimal model depends on deployment priorities: accuracy-focused applications may prefer MixNet, whereas resource-constrained, real-time systems benefit most from SqueezeNet or MobileNet variants. The study provides practical insights for integrating lightweight deep learning models into next-generation driver monitoring systems.

**Keywords: intelligent transportation, road safety, driver distraction, lightweight model, convolutional neural networks**

## 1. Introduction

Intelligent Transportation Systems (ITS) play a pivotal role in enhancing the efficiency, safety, and environmental sustainability of modern mobility by utilizing advanced technologies, including artificial intelligence, real-time data analytics, and automation. (Azizoglu et al., 2024). Among the various issues addressed by ITS, road safety is one of the most significant challenges. Despite ongoing advancements in vehicle technology and the implementation of active safety mechanisms, driver distraction remains a significant contributor to traffic accidents worldwide (Azizoglu and Toprak, 2024). Recent statistics show that approximately 1.19 million people die each year due to road traffic accidents, with a considerable portion of these incidents caused by visual, manual, or cognitive distractions while driving (World Health Organization, 2025). Nowadays, as connected and semi-autonomous vehicles become increasingly

widespread, the accurate and efficient detection of distracted driving behaviors has become a critical necessity for reducing traffic accidents and saving lives.

In the last decade, deep learning has emerged as a powerful tool for analyzing complex driving behaviors through visual and multimodal data. Convolutional neural networks (CNNs) and visual transformers (ViTs) have shown remarkable success in detecting driver distraction behaviors. However, most of the high-performing models require significant computational resources, which limits their applicability in real-time, in-vehicle environments. Embedded platforms, such as mobile GPUs and edge devices, require models that strike a good balance between accuracy, efficiency, and robustness.

This challenge has motivated researchers to develop lightweight deep learning models capable of delivering real-time inference while maintaining high predictive performance. Lv et al. (2025) proposed Si-CA MobileNet, a lightweight convolutional neural network for distracted driving detection. By integrating SimAM and Coordinate Attention mechanisms, their model enhances feature representation while maintaining efficiency, making it well-suited for deployment on resource-constrained devices. They validated the approach using the StateFarm and 100-Driver datasets. Li et al. (2024) introduced CoViT, a lightweight and efficient model that combines convolutional neural networks with vision transformers for distracted driver detection. Their experiments involved a self-recorded real-world dataset, in addition to the StateFarm and 100-Driver public datasets. Tang et al. (2024) developed CaTNet, which is a compact model that integrates a pruned ConvNeXt backbone with a Transformer-based CaT module, enabling the capture of both local and global features for distraction recognition. They evaluated their method on the AUC and StateFarm datasets, demonstrating competitive performance.

Although various lightweight architectures have been developed in the literature for driver distraction detection, the comparative effectiveness of these models in detecting driver distraction remains underexplored. While some studies have reported promising results using lightweight models, a comprehensive comparative analysis is necessary to understand the trade-offs between accuracy and inference speed, which are crucial for real-world applications.

In this paper, we present a comparative analysis of state-of-the-art lightweight deep learning models for detecting driver distraction. We evaluate MobileNetV2 (Sandler et al., 2018), MobileNetV3 (Howard et al., 2019), MobileNetV4 (Qin et al., 2024), ShuffleNetV2 (Ma et al., 2018), SqueezeNet (Iandola et al., 2016), EfficientNet-B0 (Tan and Le, 2019a), ResNet18 (He et al., 2016), MixNet (Tan and Le, 2019b), EfficientViT-B0 (Liu et al., 2022), and GhostNetV2 (Tang, 2022) architectures using the American University in Cairo (AUC) (Eraqi et al., 2019) and StateFarm (Montoya et al., 2016) datasets. Our evaluation considers not only the performance of these models in detecting driver distraction but also their inference speed, memory requirements, parameter efficiency, and potential for real-world deployment. By doing so, we aim to offer valuable insights into which models are best suited for on-device driver distraction detection and how future studies can further optimize the balance between efficiency and accuracy. Our contributions are threefold:

1. We compare state-of-the-art lightweight deep learning architectures for driver distraction detection.
2. We conduct a comprehensive analysis of model performance, evaluating both prediction accuracy and system-level factors, such as inference time and memory requirements.
3. We offer recommendations for integrating lightweight models into real-world driver distraction detection systems.

Through this study, we aim to bridge the gap between theoretical advancements in lightweight deep learning and their practical applications in safety-critical driver monitoring systems.

## 2. Materials and Methods

### 2.1. Lightweight deep learning models

#### 2.1.1. MobileNetV2

MobileNetV2 (Sandler et al., 2018) introduced the concept of inverted residuals with linear bottlenecks, improving upon its predecessor by significantly reducing computation while preserving accuracy. The network relies heavily on depthwise separable convolutions to minimize parameter count and computational cost. Its streamlined design makes it particularly suitable for mobile and embedded applications requiring efficient real-time inference.

#### 2.1.2. MobileNetV3

MobileNetV3 (Howard et al., 2019) combines architecture search with network-level optimizations, such as squeeze-and-excitation modules and the use of the H-Swish activation function. It provides a better trade-off between accuracy and inference time compared to MobileNetV2, especially on edge devices. Designed explicitly for resource-constrained environments, MobileNetV3 has been widely adopted for real-time vision tasks.

#### 2.1.3. MobileNetV4

MobileNetV4 (Qin et al., 2024) extends the MobileNet family by unifying hardware-aware network design with modern training techniques. It employs efficient attention modules and dynamic depth scaling to achieve state-of-the-art performance. MobileNetV4 demonstrates improved efficiency across a wide range of devices, making it a promising option for driver monitoring systems in real-world vehicles.

#### 2.1.4. ShuffleNetV2

ShuffleNetV2 (Ma et al., 2018) was designed with a strong focus on practical efficiency, guided by a set of direct metrics, including memory access cost and degree of parallelism. It employs channel shuffling and lightweight group convolutions to reduce computational overhead while maintaining feature diversity. This architecture achieves high efficiency on mobile devices, making it well-suited for real-time detection of driver distraction.

### 2.1.5 SqueezeNet

SqueezeNet (Iandola et al., 2016) was first introduced as achieving AlexNet-level accuracy with 50 times fewer parameters by utilizing fire modules, which consist of squeeze layers with small filters followed by expand layers with larger filters. Its compact design allows deployment on devices with limited memory and storage. While it is less accurate than newer models, SqueezeNet remains a notable example of a lightweight deep neural network.

### 2.1.6. EfficientNet-B0

EfficientNet-B0 (Tan & Le, 2019a) introduced a compound scaling method that uniformly scales network depth, width, and resolution based on a set of fixed scaling coefficients. This principled approach achieves better accuracy and efficiency than arbitrary scaling of traditional CNNs. EfficientNet-B0 is the baseline version and is widely used as a strong, lightweight backbone in real-time vision tasks.

### 2.1.7. ResNet18

ResNet18 (He et al., 2016) is a compact variant of the ResNet family, employing residual connections to ease the training of deeper networks by mitigating vanishing gradient issues. With only 18 layers, it strikes a balance between depth and efficiency, making it one of the lightest ResNet versions suitable for real-time deployment. Despite its simplicity, ResNet18 remains competitive in many vision applications.

### 2.1.8. MixNet

MixNet (Tan & Le, 2019b) extends EfficientNet by utilizing mixed depthwise convolution kernels (e.g., 3×3, 5×5, 7×7) within the same layer, enabling the model to efficiently capture multi-scale features. This design enhances accuracy while maintaining a compact model. MixNet demonstrates strong performance in mobile vision tasks, making it a candidate for distraction detection under constrained hardware.

### 2.1.9. EfficientViT-B0

EfficientViT-B0 (Liu et al., 2022) is a lightweight ViT designed specifically for mobile applications. It incorporates efficient attention mechanisms along with convolutional stems to effectively capture both local and global dependencies while keeping computational costs low. This hybrid design enables EfficientViT-B0 to utilize the powerful representation capabilities of transformers without a high inference time.

### 2.1.10. GhostNetV2

GhostNetV2 (Tang et al., 2022) builds on the original GhostNet by incorporating ghost modules that utilize inexpensive operations, thus enhancing feature diversity. It generates more feature maps through cost-effective linear operations, which significantly lowers the computational cost when compared to traditional convolutions. With its focus on efficiency, GhostNetV2 achieves high performance on edge devices, making it suitable for real-time driver distraction detection.

## 2.2. Driver Distraction Datasets

### 2.2.1. American University in Cairo (AUC) dataset

The American University in Cairo (AUC) Distracted Driver Dataset (Eraqi et al., 2019) is one of the most widely used benchmarks for evaluating driver distraction detection methods. It contains more than 14,000 labeled images captured in real-world driving scenarios, representing 10 classes of driver behavior, including safe driving, texting (using both right and left hands), talking on the phone (using both right and left hands), adjusting the radio, drinking, reaching hair or putting on makeup, and reaching behind. Fig. 1 shows sample images for each distraction class. The dataset was collected under varying illumination and pose conditions to reflect real-world challenges such as occlusions and diverse driver appearances. Its coverage of both normal and distracted behaviors makes it especially valuable for training and benchmarking lightweight deep learning models in real-time driver monitoring systems.




**Figure 1. Sample images for each distraction class from the AUC dataset**

### 2.2.2. StateFarm dataset

The StateFarm Distracted Driver Dataset (Montoya et al., 2016) is one of the earliest and most widely used datasets for detecting driver distraction. It was released as part of a Kaggle competition. The StateFarm dataset contains over 22,000 labeled images of drivers engaged in 10 distinct behaviors, including safe driving, texting (using right and left hands), talking on the phone (using right and left hands), operating the radio, drinking, reaching behind, hair or makeup, and talking to a passenger. These images were captured inside real vehicles, showcasing natural variations in lighting and the appearance of the drivers. This makes the dataset both challenging and representative of actual driving conditions. Due to its scale and diversity, the StateFarm dataset has become a standard reference for evaluating and comparing deep learning models in the field of driver distraction detection. Fig. 2 displays sample images from the StateFarm dataset.




**Figure 2. Sample images for each distraction class from the StateFarm dataset**

## 3. Findings

In this study, experiments were conducted to examine the performance of popular lightweight models widely used in the literature for driver distraction detection tasks. The performance of these models was evaluated on the AUC and StateFarm datasets, which are frequently employed in this field. Within the scope of this study, a comprehensive comparison was performed using fundamental performance metrics, including accuracy, precision, recall, and F1-score.

In the experiments, both datasets were split into 75% for training and 25% for testing. The AUC dataset was already partitioned in this ratio. However, the StateFarm dataset does not specify a split ratio, so the same 75%-25% division was applied. Table 1 presents the hyperparameters used in training the compared models.

**Table 1 Hyperparameters used in the training of the compared models**

| Hyperparameter | Value |
|---|---|
| **Learning rate** | 0.0001 |
| **Weight decay** | 0.00005 |
| **Batch size** | 16 |
| **Epochs** | 30 |
| **Optimizer** | AdamW |

The obtained findings are summarized in Tables 2 and 3, respectively. The experimental results reveal significant performance variations across the two datasets, with the AUC dataset proving to be more challenging compared to StateFarm. On the AUC dataset, MixNet achieved the highest performance with an accuracy of 95.01%, followed closely by ResNet18 (94.69%) and EfficientNet-B0 (94.67%). In contrast, EfficientViT-B0 showed the lowest performance, with an accuracy of 87.99%. The performance gap between the best and worst performing models on the AUC dataset was approximately 7%, indicating considerable variation in model effectiveness on this more complex dataset.

**Table 2 Comparison of lightweight driver distraction detection models across the AUC dataset.**

| Model | Accuracy | Recall | Precision | F1-Score |
|---|---|---|---|---|
| **SqueezeNet** | 0.9356 | 0.9366 | 0.9356 | 0.9357 |
| **ShuffleNetV2** | 0.9446 | 0.9453 | 0.9446 | 0.9447 |
| **MobileNetV3** | 0.9390 | 0.9394 | 0.9390 | 0.9391 |
| **EfficientViT-B0** | 0.8799 | 0.9093 | 0.8799 | 0.8869 |
| **MobileNetV2** | 0.9411 | 0.9423 | 0.9411 | 0.9412 |
| **MobileNetV4** | 0.9370 | 0.9374 | 0.9370 | 0.9370 |
| **MixNet** | **0.9501** | **0.9504** | **0.9501** | **0.9502** |
| **EfficientNet-B0** | 0.9467 | 0.9471 | 0.9467 | 0.9467 |
| **GhostNetV2** | 0.9342 | 0.9348 | 0.9342 | 0.9342 |
| **ResNet18** | 0.9469 | 0.9475 | 0.9469 | 0.9469 |

On the StateFarm dataset, all models demonstrated high performance, with ShuffleNetV2 and MixNet both achieving near-perfect results of 99.96% accuracy. The performance range was much narrower, with even the lowest-performing model (SqueezeNet at 98.09%) still achieving very good results. While both datasets yielded high accuracy rates, the StateFarm dataset consistently produced superior performance metrics across all models, with most architectures achieving over 99% accuracy. This suggests that StateFarm may contain more distinguishable patterns for driver distraction detection, whereas the AUC dataset presents more challenging scenarios that better differentiate model capabilities.

**Table 3 Comparison of lightweight driver distraction detection models across the StateFarm dataset.**

| Model | Accuracy | Recall | Precision | F1-Score |
|---|---|---|---|---|
| **SqueezeNet** | 0.9809 | 0.9818 | 0.9809 | 0.9810 |
| **ShuffleNetV2** | **0.9996** | **0.9996** | **0.9996** | **0.9996** |
| **MobileNetV3** | 0.9984 | 0.9984 | 0.9984 | 0.9984 |
| **EfficientViT-B0** | 0.9988 | 0.9988 | 0.9988 | 0.9988 |
| **MobileNetV2** | 0.9989 | 0.9989 | 0.9989 | 0.9989 |
| **MobileNetV4** | 0.9945 | 0.9946 | 0.9945 | 0.9945 |
| **MixNet** | **0.9996** | **0.9996** | **0.9996** | **0.9996** |
| **EfficientNet-B0** | 0.9988 | 0.9988 | 0.9988 | 0.9988 |
| **GhostNetV2** | 0.9977 | 0.9977 | 0.9977 | 0.9977 |
| **ResNet18** | 0.9950 | 0.9952 | 0.9950 | 0.9950 |

In Table 4, the evaluation conducted on the NVIDIA Jetson Xavier using 200 image inferences, sampled from the AUC dataset, highlights the computational efficiency of various lightweight deep learning models for driver distraction detection. The results reveal that SqueezeNet achieves the highest GPU throughput at 116.61 FPS, along with the lowest inference latency of 8.58 ms. Furthermore, it maintains a minimal parameter count of 0.74M, underscoring its suitability for deployment on resource-constrained embedded platforms. In contrast, Models from the MobileNet family, such as MobileNetV4 (which achieves 79.05 FPS), strike a good balance between speed and complexity. However, their higher number of Floating Point Operations (FLOPs) results in noticeably longer inference times on CPUs. Models such as EfficientNet-B0 and GhostNetV2 offer moderate performance on GPUs, with frame rates of 40.12 FPS and 31.13 FPS, respectively. Yet, they require significantly more computational resources, resulting in considerable latency on CPUs. As a secondary option, MobileNetV3 has the lowest number of FLOPs (61.46 million) among all the models compared. This characteristic makes it demanding in terms of computational efficiency and potential energy savings, despite its lower inference speed compared to SqueezeNet, at 61.34 FPS (16.30 ms).

**Table 4 Comparison of computational efficiency of lightweight driver distraction detection models.**

| Model | FPS | | Inference Time (ms) | | FLOPs | Params |
|---|---|---|---|---|---|---|
| | GPU | CPU | GPU | CPU | (M) | (M) |
| **SqueezeNet** | **116.61** | **10.07** | **8.58** | **99.31** | 733.26 | **0.74** |
| **ShuffleNetV2** | 55.50 | 1.32 | 18.02 | 760.14 | 151.70 | 1.26 |
| **MobileNetV3** | 61.34 | 1.34 | 16.30 | 743.69 | **61.46** | 1.53 |
| **EfficientViT-B0** | 64.04 | 1.03 | 15.61 | 974.39 | 104.265 | 2.14 |
| **MobileNetV2** | 68.59 | 0.49 | 14.58 | 2054.07 | 326.22 | 2.24 |
| **MobileNetV4** | 79.05 | 0.64 | 12.65 | 1564.99 | 184.77 | 2.47 |
| **MixNet** | 36.28 | 0.51 | 27.56 | 1948.06 | 237.73 | 2.58 |
| **EfficientNet-B0** | 40.12 | 0.36 | 24.93 | 2753.68 | 413.88 | 4.02 |
| **GhostNetV2** | 31.13 | 0.21 | 32.12 | 4779.96 | 183.80 | 4.87 |
| **ResNet18** | 62.08 | 5.92 | 16.11 | 168.89 | 1823.53 | 11.18 |

Fig. 3 presents a comparison of lightweight deep learning models in terms of accuracy and FPS. Each bubble represents a model, with the size of the bubble proportional to the number of parameters (in millions). Models in the upper-right region are the most desirable, as they combine higher accuracy with high FPS.

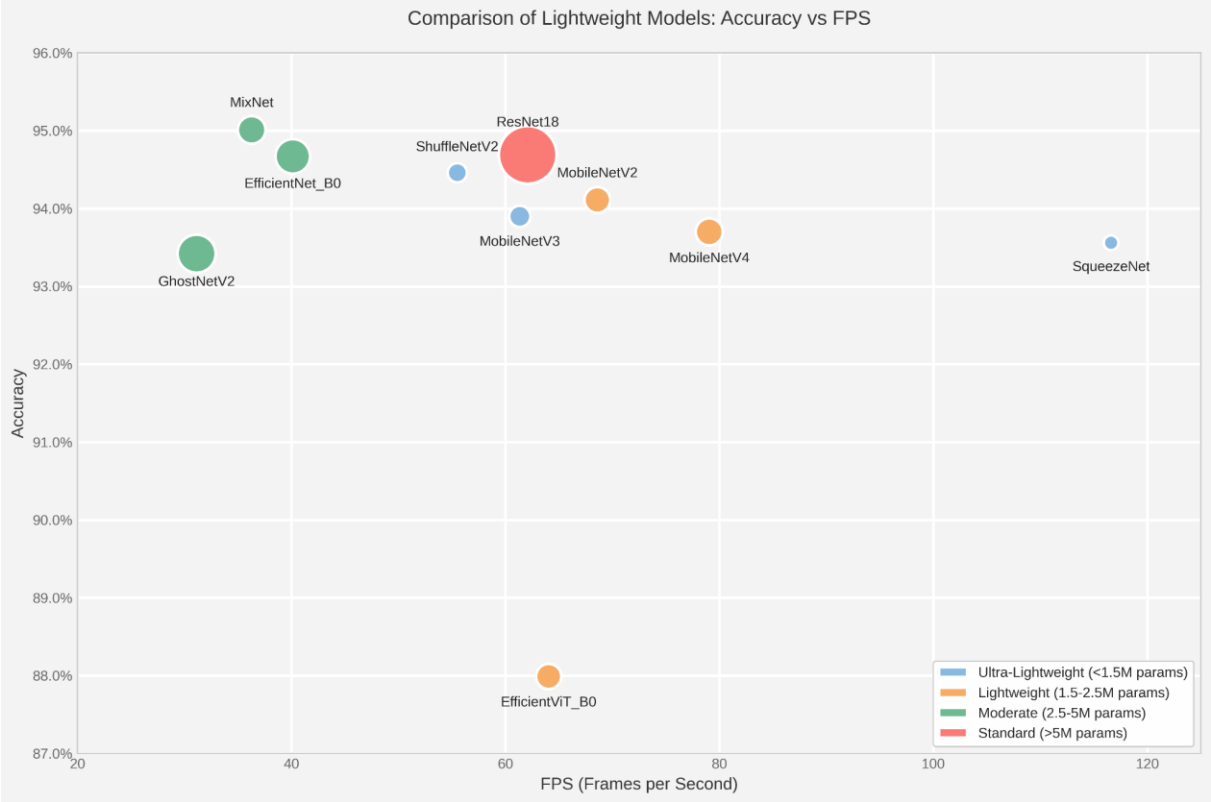


**Figure 3. Comparison of lightweight deep learning models in terms of accuracy and FPS.**

## 4. Results

The main objective of this study was to evaluate and compare the effectiveness of state-of-the-art lightweight deep learning models for driver distraction detection using both the AUC and StateFarm datasets. The results provide insights into the trade-offs between accuracy, inference speed, and computational efficiency, which are critical for deployment in real-time, in-vehicle environments.

On the AUC dataset, which presents more challenging conditions with diverse illumination, poses, and driver variations, the models displayed a wide range of performance. MixNet achieved the best performance with an accuracy of 95.01%, closely followed by ResNet18 (94.69%) and EfficientNet-B0 (94.67%). In contrast, EfficientViT-B0 yielded the lowest accuracy at 87.99%. The approximately 7% gap between the highest- and lowest performing models demonstrates that architectural design choices have a significant impact on performance.

On the StateFarm dataset, nearly all models achieved high accuracy, with ShuffleNetV2 and MixNet both reaching 99.96%. Even the lowest performing model, SqueezeNet, attained 98.09%, confirming that StateFarm offers more easily distinguishable distraction patterns. The consistently strong results across all architectures suggest that while StateFarm is valuable for benchmarking, it may not fully expose the weaknesses of lightweight models when applied to more challenging real-world conditions.

The computational efficiency analysis conducted on an NVIDIA Jetson Xavier device further highlights the trade-offs among architectures. SqueezeNet emerged as the most efficient model, achieving 116.61 FPS with only 0.74 million parameters, highlighting its suitability for embedded platforms where resource constraints are important. MobileNetV4 achieved a strong balance of speed (79.05 FPS) and accuracy (93.7%), demonstrating its potential for practical in-vehicle deployment. Conversely, models such as EfficientNet-B0 and GhostNetV2 achieved solid accuracy but incurred longer inference times (40.12 FPS and 31.13 FPS, respectively) and higher FLOP requirements, which limited their usability in real-world scenarios.

Overall, the results indicate that while MixNet and ResNet18 offer the highest accuracy on both datasets, SqueezeNet and MobileNetV4 stand out in terms of efficiency and deployment feasibility. The choice of model thus depends on application priorities: systems prioritizing maximum detection accuracy may prefer MixNet or ResNet18, while resource-constrained, real-time environments may benefit more from SqueezeNet or MobileNet variants.

**Acknowledgement**

This work was supported by the Erciyes University Scientific Research Projects Coordination Office under Grant No. FDK-2023-13350 and by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) under the 1002-A Short Term Support Module funding program with Project No. 124E762.

**References**

Azizoglu, G., and Toprak, A. N. (2024). MELD3: Integrating multi-task ensemble learning for driver distraction detection. IEEE Access, 12, pp. 186022-186034.

Azizoglu, G., Toprak, A. N., and Akkaya, S. (2024). An automatic fuel dispensing system using YOLOv5-based vehicle identification: A real-life application. In: *2024 32nd Signal Processing and Communications Applications Conference* (SIU), IEEE, pp. 1-4.

Eraqi, H. M., Abouelnaga, Y., Saad, M. H., and Moustafa, M. N. (2019). Driver distraction identification with an ensemble of convolutional neural networks. Journal of Advanced Transportation, 2019 (1), pp. 4125865.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition,* pp. 770-778.

Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., and Adam, H. (2019). Searching for MobileNetV3. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1314-1324.

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360.

Li, Z., Zhao, X., Wu, F., Chen, D., and Wang, C. (2024). A lightweight and efficient distracted driver detection model fusing convolutional neural network and vision

transformer. IEEE Transactions on Intelligent Transportation Systems, 12, pp. 19962-19978.

Liu, X., Peng, H., Zheng, N., Yang, Y., Hu, H., and Yuan, Y. (2023). EfficientViT: Memory efficient vision transformer with cascaded group attention. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14420-14430.

Lv, M., Liu, Y., Zha, Z., Zheng, X., Wang, H., Wen, Y., and Guo, Z. (2025). Si-CA MobileNet: A lightweight and efficient convolutional neural network for distracted driver detection. Neurocomputing, 654, pp. 131281.

Ma, N., Zhang, X., Zheng, H. T., and Sun, J. (2018). ShuffleNet V2: Practical guidelines for efficient CNN architecture design. In: *Proceedings of the European Conference on Computer Vision* (ECCV), pp. 116-131.

Montoya, A., Holman, D., Smith, T., and Kan, W. (2016). State Farm Distracted Driver Detection. (https://kaggle.com/competitions/state-farm-distracted-driver-detection) Accessed 30 Aug. 2025.

Qin, D., Leichner, C., Delakis, M., Fornoni, M., Luo, S., Yang, F., and Howard, A. (2024). MobileNetV4: Universal models for the mobile ecosystem. In: *European Conference on Computer Vision*, pp. 78-96.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510-4520.

Tan, M. and Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In: *Proceedings of the 36th International Conference on Machine Learning,* 97, pp.6105-6114.

Tan, M., and Le, Q. V. (2019). MixConv: Mixed depthwise convolutional kernels. *arXiv preprint arXiv:1907.09595*.

Tang, X., Chen, Y., Ma, Y., Yang, W., Zhou, H., and Huang, J. (2024). A lightweight model combining convolutional neural network and Transformer for driver distraction recognition. Engineering Applications of Artificial Intelligence, 132, pp. 107910.

Tang, Y., Han, K., Guo, J., Xu, C., Xu, C., and Wang, Y. (2022). GhostNetV2: Enhance cheap operation with long-range attention. Advances in Neural Information Processing Systems, 35, pp. 9969-9982.

World Health Organization, (2025). Road traffic injuries 2025. (https://www.who.int/health-topics/road-safety) Accessed 03 July 2025.