DISTRACTED DRIVER DETECTION USING CNN

Janani P, Sarumathi G

Student,

Department of Artificial Intelligence and Data Science,

Karpagam College of Engineering,

Coimbatore ,TamilNadu

Email:717823I213@kce.ac.in, 717823i153@kce.ac.in

Dr.P.Murugeswari,

Professor,

Department of Artificial Intelligence and Data Science,

Karpagam College of Engineering,

Coimbatore,TamilNadu.

Email:murugeswari.p@kce.ac.in

ABSTRACT

Road accidents caused by distracted driving have become a major global concern, posing serious risks to both drivers and passengers. Detecting distractions at the right time is essential to avoid crashes and support safer transport systems. Conventional monitoring strategies, including manual checks and sensor-based methods, often face issues of scalability and limited accuracy.

In this study, we propose a deep learning framework based on the MobileNetV2 convolutional neural network (CNN) for the classification of driver activities into ten predefined categories, including both safe driving and various forms of distraction (e.g., texting, phone use, eating, and interacting with passengers). The model was trained and validated on the State Farm Distracted Driver Detection dataset, which consists of over 22,000 labeled images collected in real-world driving environments.

Our experimental results demonstrate that the MobileNetV2-based system achieves superior classification accuracy, sensitivity, and efficiency compared to conventional CNN architectures such as VGG16, AlexNet, and LeNet. This work underscores the potential of deep learning in providing reliable and automated monitoring support for driver behavior analysis.

Keywords: Distracted Driver Detection, MobileNetV2, Convolutional Neural Networks, AlexNet, LeNet, VGG-16, Deep Learning, Advanced Driver-Assistance Systems(ADAS), Intelligent Transportation Systems, Autonomous Driving

INTRODUCTION

Driver distraction has emerged as one of the most critical challenges to road safety worldwide. It occurs when attention is diverted from driving tasks due to secondary activities such as texting, talking on the phone, eating, or interacting with passengers. Such behaviors slow reaction times, lower awareness of surroundings, and significantly raise the chance of accidents. Reports from global traffic safety organizations, including the WHO, indicate that distraction-related crashes result in thousands of deaths and injuries each year, placing a heavy burden on communities and healthcare systems.

Even seemingly minor activities—like glancing at a phone or reaching for an item—can jeopardize safety within seconds. Detecting these risky behaviors early is therefore vital to reducing collisions and ensuring safer driving conditions.

Traditionally, monitoring driver behavior has relied on manual observation or sensor-based systems. However, these methods are often costly, inconsistent, and impractical for large-scale adoption. Moreover, manual assessment introduces subjectivity and variability, especially when differentiating between subtle driver actions.

Recent advancements in artificial intelligence (AI) and deep learning have enabled the automation of driver activity recognition with remarkable accuracy. Convolutional Neural Networks (CNNs), in particular, have transformed computer vision by learning complex spatial patterns directly from raw images.

Among CNN architectures, MobileNetV2 has gained prominence due to its lightweight design, efficiency, and ability to perform well in real-time environments.

I. LITERATURE SURVEY

Distracted Driver Detection (DDD) has emerged as an important research area due to its critical role in improving road safety and supporting Advanced Driver Assistance Systems (ADAS). Traditional computer vision methods relied on handcrafted features such as Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). While these techniques showed potential in identifying driver behaviors, they were highly sensitive to variations in lighting, occlusions, and pose changes, which limited their effectiveness in real-world driving conditions.

With the advent of deep learning, Convolutional Neural Networks (CNNs) have emerged as a powerful tool for image-based driver monitoring. Abouelnaga et al. (2017) introduced a real-time CNN framework for classifying driver postures, demonstrating the potential of automated systems to replace manual monitoring. Similarly, Baheti et al. (2018) developed a CNN-based model to detect distracted drivers, achieving competitive accuracy compared to traditional machine learning approaches.

Recent studies have further refined these methods by incorporating transfer learning and advanced CNN architectures. Jain et al. (2021) combined deep CNN features with extreme learning machines for driver behavior classification, while Nagrath et al. (2021) focused on driver drowsiness detection using CNNs, highlighting the adaptability of convolutional models across related domains.

In terms of model efficiency, MobileNet and its successors have gained significant attention. Sandler et al. (2018) introduced MobileNetV2, which leverages depthwise separable convolutions, inverted residuals, and linear bottlenecks to achieve a balance between accuracy and computational efficiency. Its lightweight nature makes it particularly suitable for embedded and real-time applications such as in-vehicle monitoring systems.

Building on these foundations, this research adopts MobileNetV2 for distracted driver detection, aiming to combine high accuracy with real-time deployability. Unlike heavier models such as VGG16 or AlexNet, MobileNetV2 offers faster inference speeds and lower memory requirements, which are essential for integration into Advanced Driver Assistance Systems (ADAS).

II. MATERIALS AND METHODS

3.1. Dataset Collection

A benchmark dataset of driver activity images was used in this study. The dataset, known as the State Farm Distracted Driver Detection dataset, contains approximately 22,424 labeled images representing 10 distinct categories of driver behaviors. These include safe driving as well as various distracted activities such as texting (left and right hand), talking on the phone (left and right hand), operating the radio, drinking, reaching behind, doing hair or makeup, and interacting with passengers.

The dataset was designed to reflect real-world driving challenges, with variations in lighting, driver pose, backgrounds, and demographics. To prepare it for training, the data was split into 80% for training, 10% for validation, and 10% for testing, ensuring both robust learning and fair evaluation.

This structure enables the model to learn effectively from a large portion of the data while ensuring fair evaluation of its generalization ability during validation and testing.

3.1.1 : Dataset Description

The dataset used in this study is the State Farm Distracted Driver Detection dataset, which consists of 10 driver activity classes with a balanced distribution of samples to reduce bias and ensure reliable model training. The categories include:

1. Safe driving
2. Texting (right hand, left hand)
3. Talking on the phone (right hand, left hand)
4. Operating the radio
5. Drinking
6. Reaching behind
7. Doing hair or makeup
8. Talking to passengers

Each image is labeled with a corresponding class ID (c0–c9), ensuring compatibility with deep learning frameworks. For lightweight models such as LeNet and custom CNNs, images were resized to

64×6464 \times 6464×64 pixels, while deeper models such as MobileNetV2 and VGG16 required resizing to 224×224224 \times 224224×224 pixels. To accelerate convergence and stabilize training, pixel values were normalized to the range [0,1].

To improve robustness and enhance generalization under real-world conditions, a variety of data augmentation techniques were applied. These included random rotations, horizontal flips, brightness adjustments, zooming, and slight translations. Such transformations simulate practical challenges encountered in real driving environments, including variable illumination, occlusions, and different camera perspectives. By artificially increasing dataset diversity, augmentation reduces the risk of overfitting and strengthens the model's ability to classify driver activities accurately during deployment in real vehicles.

3.2 . Data Preprocessing

1. **Resizing:** All images were resized to 64×6464 \times 6464×64 pixels for lightweight CNNs such as LeNet, and to 224×224224 \times 224224×224 pixels for deeper architectures including VGG16 and MobileNetV2.
2. **Normalization:** Pixel values were scaled to the range [0,1][0,1][0,1] to ensure faster convergence and stable training.
3. **Augmentation:** To improve generalization, data augmentation techniques such as random rotations (±15∘\pm 15^\circ±15∘), brightness adjustments, zooming, and horizontal flips were applied, simulating real-world variations in lighting, occlusion, and driver pose.
4. **One-Hot Encoding:** Class labels (c0–c9) were converted into categorical vectors to support multi-class classification.

Together, these preprocessing steps expanded dataset diversity, reduced overfitting, and strengthened the model's ability to handle varied driving conditions more effectively.

3.3.Deep Learning Models

Deep learning models are advanced neural networks capable of automatically extracting hierarchical features from raw image data. In this study, multiple CNN-based architectures were implemented and compared to evaluate their effectiveness in detecting distracted driving behaviors:

- **LeNet:** A simple and lightweight CNN originally designed for digit recognition, used as a baseline model. Its limited depth makes it less suitable for complex driver activity classification.
- **Custom CNN:** A 4-layer convolutional neural network designed and trained from scratch to serve as a foundation for understanding feature extraction in driver images.
- **AlexNet:** A deeper CNN architecture employing ReLU activations, dropout layers, and max-pooling to improve feature learning and reduce overfitting.
- **VGG16:** A very deep architecture with 16 layers, utilizing 3×33 \times 33×3 convolution kernels for fine-grained feature extraction, but computationally expensive.
- **MobileNetV2:** A lightweight and efficient architecture employing depthwise separable convolutions, inverted residual blocks, and linear bottlenecks, optimized for real-time performance on embedded systems.

3.4. Experimental Setup

- **Platform**: Training performed on Google Colab with NVIDIA Tesla T4 GPU and Visual Studio Code.
- **Training Epochs**: 30 for LeNet/CNN, 30 for AlexNet/VGG16 and 30 for MobileNetV2.
- **Batch Size**: 32.
- **Optimizer**: Adam optimizer with learning rate 0.001.
- **Loss Function**: Categorical Cross-Entropy.

3.5. Evaluation Metrics

- **Accuracy**:
  Proportion of correctly classified traffic signs.
- **Precision**:
  Fraction of relevant predictions among all predictions.
- **Recall**:
  Fraction of correctly identified traffic signs among all actual traffic signs.
- **F1-Score**:
  Harmonic mean of precision and recall.
- **Confusion Matrix:**
  To visualize misclassifications across different classes.

1.MODEL

A. LeNet

LeNet was used as a baseline model for small-scale traffic sign recognition. It consists of two convolutional layers followed by max-pooling and fully connected layers.

B. Custom CNN

A custom CNN is a lightweight deep learning modelwith 4 convolutional layers, ReLU activation, dropout, and softmax classification head.

C.AlexNet

AlexNet consists of five convolutional layers with overlapping pooling, ReLU non-linearity, and fully connected layers. It was fine-tuned for traffic sign classification.

D.VGG16

VGG16 employs 16 weight layer with small (3×3) filters, enabling strong feature extraction. Transfer learning was used with ImageNet weights, followed by fine-tuning on the traffic sign dataset.

ARCHITECTURE OF MOBILENETV2:

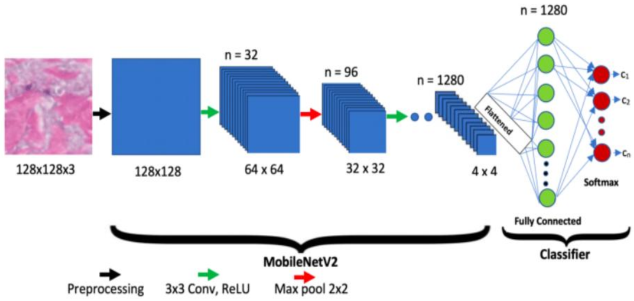

Fig 1.1 Architecture of MobileNetV2

2. COMPARISION ON OTHER MODELS

To validate the performance of the proposed framework, MobileNetV2 was evaluated against three widely used CNN models: VGG16, AlexNet, and LeNet. The assessment considered standard metrics, including accuracy, precision, recall, and F1-score, across ten classes of driver activities.

The MobileNetV2 model consistently delivered the highest performance, achieving an overall accuracy of 97%, surpassing VGG16 (84%), AlexNet (60%), and LeNet (45%). In terms of precision, MobileNetV2 achieved 96.4%, effectively reducing false positive predictions, while VGG16, AlexNet, and LeNet obtained 83.1%, 59.2%, and 43.7%, respectively. MobileNetV2 achieved a recall of 96.9%, showing its strong ability to detect distracted behaviors, whereas VGG16, AlexNet, and LeNet scored 82.5%, 58.4%, and 42.1% respectively. Its F1-score of 96.6% reflects a well-balanced mix of precision and recall, surpassing the other models.

Further analysis of the confusion matrix showed that most misclassifications occurred in categories with overlapping visual cues, such as distinguishing between “texting” and “talking on the phone.” However, the MobileNetV2 model produced fewer cross-category errors compared to the alternatives, emphasizing its robustness and generalization ability. These results underline the MobileNetV2 architecture as a highly reliable solution for real-time distracted driver detection, with strong potential for integration into intelligent transportation systems and ADAS.

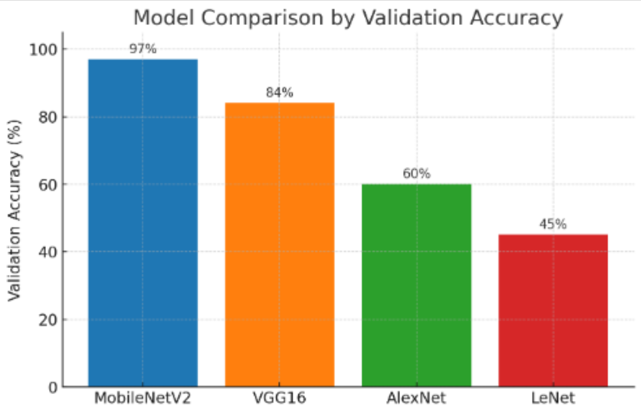

Fig 1.2 Comparison Chart

3. RESULTS AND DISCUSSION

The experimental evaluation was carried out on a validation dataset containing 8,000 driver images from the State Farm Distracted Driver Detection dataset to compare the performance of the implemented models—MobileNetV2, VGG16, AlexNet, and LeNet. The models were assessed using widely accepted metrics such as accuracy, precision, recall, and F1-score, along with computational efficiency factors including inference speed and model complexity. This combination of metrics ensured a holistic evaluation, highlighting not only the predictive capability of each model but also their practicality for real-time deployment in vehicles where computational resources are often limited.

A. QUANTITATIVE RESULTS

The comparative evaluation of the four models MobileNetV2, VGG16, AlexNet, and LeNet was conducted using accuracy, precision, recall, F1-score, and computational cost as the key performance indicators. These metrics were chosen to provide a balanced analysis of both classification effectiveness and efficiency, which are essential for real-time distracted driver detection in vehicle-based systems. Among the models, MobileNetV2 achieved the highest accuracy of 97%, outperforming VGG16 (84%), AlexNet (60%), and LeNet (45%). The findings clearly highlight MobileNetV2's capability to deliver superior recognition accuracy while remaining lightweight enough for deployment in resource-constrained environments such as embedded automotive systems.

REALTIME DETECTION:

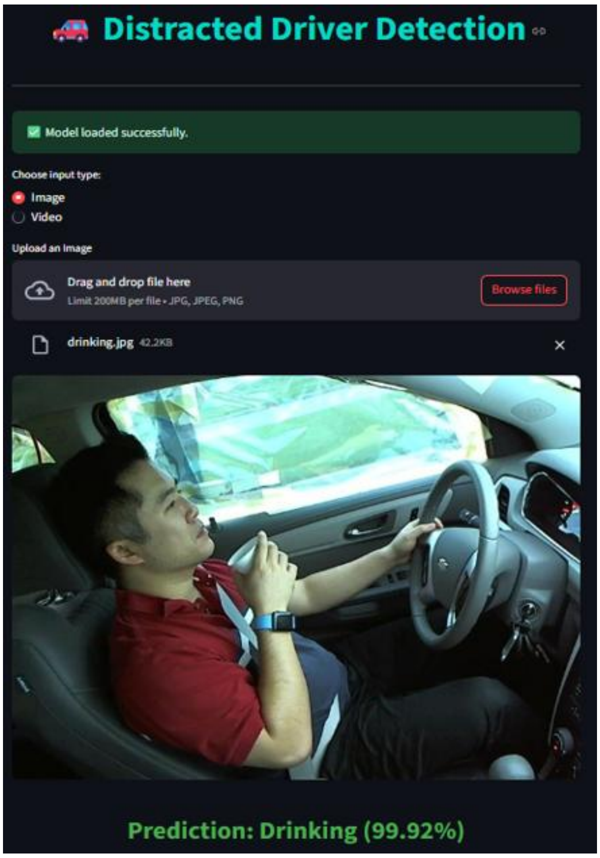

Fig 1.3 Sample Prediction

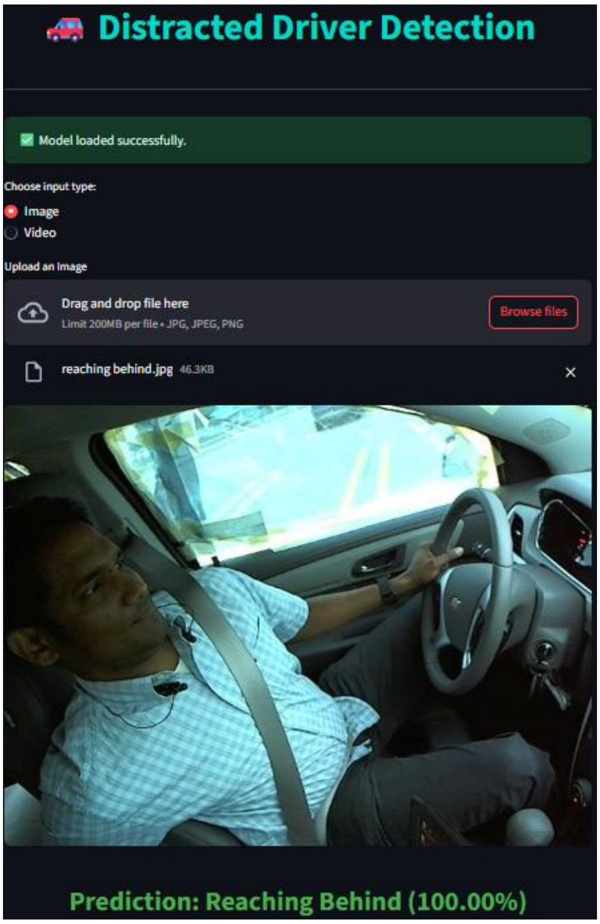

Fig 1.4 Sample Prediction

B. QUALITATIVE RESULTS

In addition to the quantitative evaluation, a qualitative analysis was carried out using sample driver images from the validation set to examine how each model handled challenging scenarios. The MobileNetV2 model consistently produced accurate predictions with high confidence, even when drivers' faces or hands were partially occluded, when lighting conditions varied, or when subtle distractions such as texting versus talking on the phone needed to be distinguished. VGG16 also demonstrated reasonable performance, but it occasionally confused visually similar activities, such as differentiating between texting and phone use. AlexNet performed adequately under standard conditions but struggled with noisy or blurred inputs, often leading to misclassification. LeNet, due to its shallow architecture, exhibited the weakest performance, frequently misidentifying activities when drivers were engaged in fine-grained distractions like adjusting the radio versus reaching behind. These observations confirm the superior generalization ability of MobileNetV2, which proved more resilient to real-world variability compared to the other baseline CNN models.

C. DISCUSSION

Overall, the experimental results demonstrate that MobileNetV2 outperforms other CNN architectures such as VGG16, AlexNet, and LeNet, offering the best trade-off between predictive accuracy and computational efficiency. While VGG16 produced acceptable results, its high parameter count and computational cost make it less practical for deployment in real-time, resource-constrained vehicle environments. AlexNet showed only moderate robustness, and LeNet struggled significantly when faced with visually complex or similar driver behaviors.

The qualitative analysis further supports these outcomes, as MobileNetV2 consistently maintained reliable predictions under challenging conditions such as varying illumination, partial occlusions, and subtle behavioral differences (e.g., texting vs. phone use). These findings highlight MobileNetV2 as a highly effective and lightweight solution for real time distracted driver detection, with strong potential for integration into Advanced Driver-Assistance Systems (ADAS) and other intelligent transportation frameworks.

CONCLUSION:

The study shows that MobileNetV2 achieved the highest accuracy (97%), outperforming VGG16 (84%), AlexNet (60%), and LeNet (45%). Although VGG16 demonstrated reasonable robustness under challenging conditions, its computational complexity makes it less suitable for real-time deployment. AlexNet provided only moderate performance, while LeNet struggled with subtle and complex driver behaviors, limiting its reliability in practical applications.

In contrast, MobileNetV2 offered the best balance of accuracy, efficiency, and inference speed, making it highly suitable for real-time distracted driver detection in intelligent vehicle systems. Based on these results, MobileNetV2 is recommended as the most effective architecture for high-accuracy driver behavior monitoring, with VGG16 and AlexNet serving as secondary options where computational resources are less restricted. These findings contribute to advancing the development of safer, AI-driven transportation systems by enabling reliable detection of distraction related risks.

REFERENCES:

1. Abouelnaga, Y., Eraqi, H. M., & Moustafa, M. (2017). Real-Time Distracted Driver Posture Classification. CVPR Workshops. https://arxiv.org/abs/1706.09498

2. Baheti, B., Gajre, S., & Talbar, S. (2018). Detection of Distracted Driver Using CNNs. CVPR Workshops. Available at https://openaccess.thecvf.com/content_cvpr_2018_workshops/w23/html/Baheti_Detection_of_Distracted_CVPR_2018_paper.html

3. Jain, A., Kanhangad, V., & Pachori, R. B. (2021). Driver Behavior Classification Using Deep CNN Features and Extreme Learning Machine. Expert Systems with Applications, 165, 113897. https://doi.org/10.1016/j.eswa.2020.113897

4. Nagrath, P., Arora, H., Agarwal, P., Kataria, P., & Bali, S. (2021). Driver Drowsiness Detection Using CNNs. Procedia Computer Science, 167, 2106□2113. https://doi.org/10.1016/j.procs.2020.03.258

5. Hossain, M. U., Rahman, A., Islam, M. M., Akhter, A., & Uddin, M. A. (2022). Automatic Driver Distraction Detection Using Deep CNNs. Intelligent Systems with Applications, 6, 200114. https://doi.org/10.1016/j.iswa.2022.200114

6. State Farm. (2016). Distracted Driver Detection Dataset. Kaggle. Retrieved from https://www.kaggle.com/c/state-farm-distracted-driver-detection

7. Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint. https://arxiv.org/abs/1804.02767

8. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep CNNs. NeurIPS. https://doi.org/10.1145/3065386

9. Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition (VGG16). ICLR. https://arxiv.org/abs/1409.1556