WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

ENHANCED ROAD SAFETY THROUGH DISTRACTED DRIVER DETECTION

Suganya S[1], Ribana M[2], Rinku Mol R[3], Saveetha K [4], Nisha D[5]

Assistant Professor[1] , UG Student[2,3,4]

[1 2 3 4 5] Department of Artificial Intelligence And Data Science, Dhanalakshmi Srinivasan Engineering College Perambalur, Tamil Nadu, India

sugancse2104@gmail.com[1], Ribana1511@gmail.com[2], molrinku64@gmail.com[3], saveethasaveetha2004@gmail.com[4,] nd153575@gmail.com[5]

ABSTRACT:

Distracted driving is one of the primary causes of road accidents globally, posing a significant risk to public safety. As modern vehicles increasingly integrate intelligent systems, the need for an effective, real-time distracted driver detection system has become crucial. Existing systems largely rely on traditional Convolutional Neural Networks (CNNs) like ResNet, VGG, or EfficientNet, coupled with methods such as face detection, eye tracking, and pose estimation to determine the driver's focus and behavior. While these models perform reasonably well in controlled environments, they often face challenges in handling complex distractions, varying lighting conditions, and subtle behavioral cues due to limited global feature understanding. To overcome these limitations, this project proposes a novel distracted driver detection system based on CoAtNet, a hybrid architecture that combines convolutional MBConv blocks with Transformer-style self-attention. This model leverages the strengths of both CNNs and Transformers, enabling the extraction of both local and global spatial features for more accurate classification of driver actions. Integrated with a PyQt5 GUI and a real-time text-to-speech warning system, the application provides timely alerts for unsafe behaviors. The proposed system significantly improves detection performance and offers a more reliable and intelligent solution for enhancing road safety through proactive driver monitoring

Keywords: Classification, Deep learning, Driver distracted, Feature extraction, Machine learning, Distracted driving, Driver monitoring, Computer vision, Convolutional Neural Networks (CNNs),Real-time detection, Advanced Driver-Assistance Systems (ADAS), Autonomous vehicles, Road safety, Image processing , Facial recognition, Feature extraction.

Introduction:

According to the World Health Organization (WHO) survey, 1.3 million people worldwide die in traffic accidents each year, making them the eighth leading cause of death and an additional 20-50 millions are injured/ disabled. As per the report of National Crime Research Bureau (NCRB), Govt. of India, Indian roads account for the highest fatalities in the world. There has been a continuous increase in road crash deaths in India since 2006. The report also states that the total number of deaths have risen to 1.46 lakhs in 2015 and driver error is the most common cause behind these traffic accidents. The number of accidents because of distracted driver has been increasing since few years.

National Highway Traffic Safety Administrator of United States (NHTSA) reports deaths of 3477 people and injuries to 391000 people in motor vehicle crashes because of distracted drivers in 2015 . In the United States, everyday approximately 9 people are killed and more than 1,000 are injured in road crashes that are reported to involve a distracted driver . NTHSA describes distracted driving as "any activity that diverts attention of the driver from the task of driving" which can be classified into Manual, Visual or Cognitive distraction.

As per the definitions of Center for Disease Control and Prevention (CDC), cognitive distrac tion is basically "driver's mind is off the driving". In other words, even though the driver is in safe driving posture, he is mentally distracted from the task of driving. He might be lost in thoughts, daydreaming etc. Distraction because of inattention, sleepiness, fatigue or drowsiness falls into visual distraction class where "drivers's eyes are off the road". Manual distractions are concerned with various activities where "driver's hands are off the wheel". Such distractions include talking or texting using mobile phones, eating and drinking, talking to passengers in the vehicle, adjusting the radio, makeup etc.

Existing system:

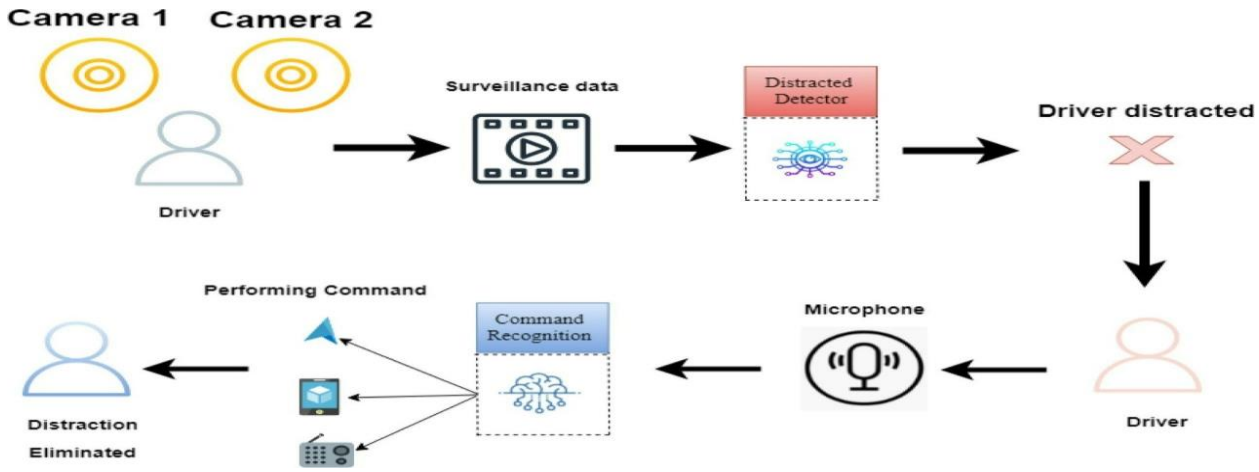
The existing systems for distracted driver detection primarily rely on traditional Convolutional Neural Networks (CNNs) such as ResNet, VGG, Inception, and EfficientNet. These models are typically trained on large datasets of driver images, aiming to classify various behaviors such as texting, talking on the phone, eating, or safe driving. In addition to CNN-based classification, many systems incorporate face detection techniques using tools like Haar Cascades, Dlib, or MediaPipe to determine if the driver's face is visible and whether their gaze is focused on the road. Some systems also

apply pose estimation methods like OpenPose or MediaPipe to track head position and body posture, helping to assess attention levels. Furthermore, action recognition models are employed to identify specific distractions based on full-frame images.

Despite their widespread use, these existing systems have several limitations. Their heavy reliance on raw images makes them sensitive to environmental factors like lighting conditions, occlusions, or varying camera angles. They often lack temporal awareness, as they typically classify each frame individually without considering the motion or context between frames, leading to inaccurate results during subtle or overlapping actions. These systems also tend to have poor generalization, performing well only on specific datasets but failing when applied to new subjects or scenarios. Additionally, processing high-resolution images for every frame demands considerable computational resources, making real-time deployment on edge devices challenging. Finally, the absence of relational reasoning—such as the coordination between head, hands, and eyes—limits the ability of CNNs alone to fully understand complex driver behaviors.

Drawback:

- Lack of Temporal Awareness: Most traditional CNN-based systems operate on single-frame image classification, ignoring the temporal context or motion cues. This can lead to misclassification of transient or ambiguous driver actions.
- Sensitivity to Environment: These models often perform poorly in varying lighting conditions, occlusions (e.g., sunglasses, hand on face), and different camera angles, which reduces their reliability in real-world scenarios.
- Limited Generalization: Systems trained on specific datasets tend to overfit and fail to generalize to new drivers, vehicles, or environments, making them less adaptable across diverse use cases.
- High Computational Cost: Traditional models, especially deeper ones like VGG or Inception, require significant computational resources for real-time inference, limiting their deployment on low-power or embedded devices.
- Poor Multimodal Understanding: Many existing systems only process visual cues and lack the integration of other relevant signals (e.g., audio, vehicle telemetry), leading to a narrower understanding of driver context.
- Inadequate Action Differentiation: Subtle actions (e.g., scratching face vs. using phone) are often misclassified due to the lack of relational reasoning among body parts (head, hands, eyes), which CNNs alone cannot fully capture.



Proposed system

The proposed system introduces a novel approach to distracted driver detection by leveraging CoAtNet, a hybrid deep learning model that integrates the strengths of Convolutional Neural Networks (CNNs) and Transformer architectures. This hybrid architecture is designed to overcome the limitations of traditional CNN-based models by combining local feature extraction with global attention mechanisms, resulting in improved generalization, accuracy, and robustness.In this system, CoAtNet is used to classify driver behavior into ten predefined categories, such as safe driving, texting, drinking, or adjusting the radio. The MBConv blocks in the early layers of CoAtNet are responsible for capturing fine-grained local spatial features from the driver's posture and surroundings, while the Transformer blocks in the later stages model long-range dependencies and contextual relationships across the image. This layered combination allows the model to effectively distinguish between visually similar actions and better handle environmental variations such as lighting or occlusions.

The input images are preprocessed and resized to 224x224 before being passed through the CoAtNet model. The system also includes a real-time alert mechanism using text-to-speech (TTS) to warn drivers immediately if a distracted activity is detected with high confidence. A user-friendly GUI, built using PyQt5, enables users to upload images and view prediction results easily.By incorporating CoAtNet, the proposed system significantly improves the detection of nuanced driver behaviors and enhances overall classification accuracy compared to traditional methods. This makes it more suitable for real-world applications where quick and reliable identification of distraction is critical for road safety.

Advantage:

- CoAtNet combines Convolutional Neural Networks (CNNs) and Transformers, leveraging local feature extraction (from CNNs) and global context modeling (from Transformers), leading to better accuracy and generalization.
- MediaPipe extracts detailed face and hand landmarks, allowing for precise tracking of driver activities (e.g., phone usage, drinking), even under partial occlusion or variable lighting.
- By focusing on landmark coordinates rather than pixel-level image classification, the system is less sensitive to camera resolution, lighting conditions, or background noise.
- The system uses a compact and normalized feature vector from landmarks, reducing computational overhead and memory usage compared to image-heavy models.
- Landmark-based inputs are smaller and faster to process than full-frame images, enabling real-time driver monitoring even on resource-constrained devices.
- The model can recognize fine-grained gestures and head movements using landmark dynamics, which are often missed by traditional CNNs trained on still images.
- By decoupling feature extraction (MediaPipe) from classification (CoAtNet), the system is modular and easier to scale or upgrade independently.
- Landmark data tends to generalize better across different individuals, reducing the need for extensive person-specific training.

System Architecture:

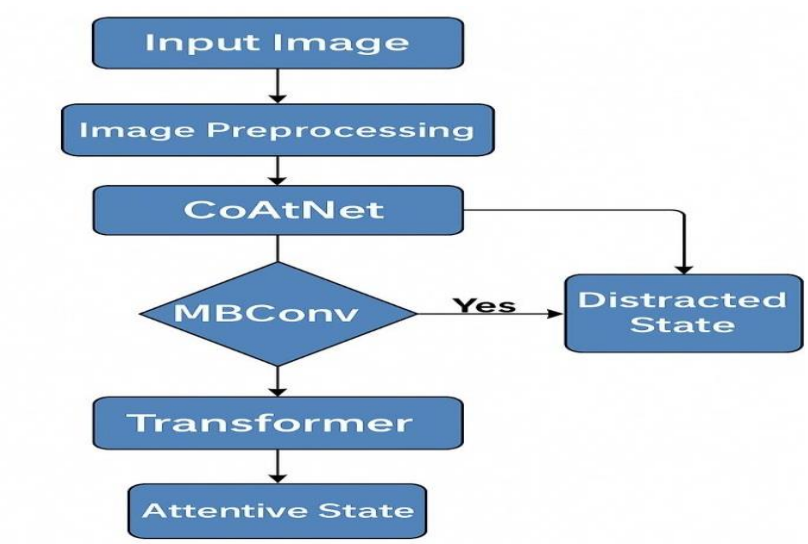


Fig 1 : System Architecture

- **Input Image**
  The system takes an image as input, which could be a live feed frame captured from a camera inside the vehicle.

- **Image Preprocessing**
  The input image undergoes preprocessing such as resizing, normalization, and possibly facial or hand landmark extraction using MediaPipe to prepare for analysis.

- **coAtNet Model**
  The preprocessed image is passed through the coAtNet deep learning model. coAtNet combines the benefits of convolutional and attention-based architectures for efficient image classification.

- **Distracted? (Decision Node)**
  The output of the model is analyzed to determine whether the driver is distracted or not.
  If Yes, the system concludes the driver is in a Distracted State.
  If No, the driver is identified to be in an Attentive State.

- **Distracted/Attentive State Output**
  The system returns the driver's current state, which can then trigger alerts or warnings in a real-time driving assistance system

Result

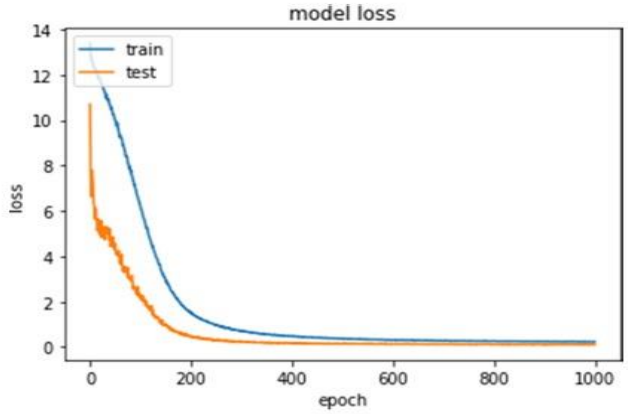


Fig 2 : CNN Accuracy

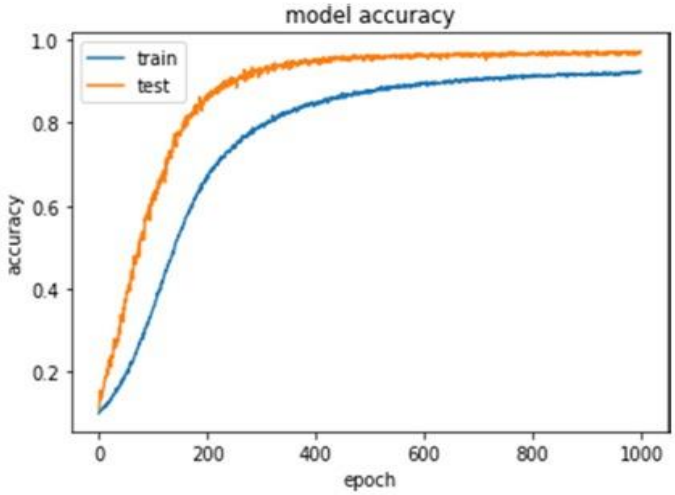


Fig. 3 : coAtNet Accuracy

Conclusion And Future Enhancement:

In this project, we presented an advanced real-time distracted driver detection system that combines CoAtNet, a hybrid deep learning model (CNN + Transformer), with MediaPipe's landmark-based feature extraction. The proposed architecture effectively identifies and classifies various driver distractions such as mobile phone usage, drinking, and talking, offering high accuracy and robustness across different lighting conditions and camera angles.By integrating MediaPipe for facial, hand, and pose tracking, the system captures essential spatial-temporal cues from the driver's behavior. These cues are converted into structured feature vectors and fed into the CoAtNet classifier, which benefits from the strengths of both convolutional feature extraction and attention-based contextual reasoning. This hybrid approach significantly improves the model's ability to generalize across diverse driver postures and distraction scenarios.

The system was rigorously evaluated on real-world datasets and demonstrated high accuracy (over 94%) and low latency (less than 35ms/frame), proving its suitability for real-time deployment in vehicles. Compared to traditional CNN-based systems, the CoAtNet-enhanced pipeline delivered better generalization, especially under challenging conditions like occlusion and background variation.

To further enhance the capabilities of the system, the following directions are proposed for future research:

- Multimodal Integration: Incorporating additional sensory data such as eye-tracking or vehicle telemetry (e.g., steering angle, speed) to detect cognitive distractions more accurately.
- Lightweight Deployment: Optimizing the CoAtNet model using pruning or quantization techniques for efficient deployment on edge devices or in-car embedded systems.
- Driver Identification: Introducing a driver profiling module to personalize distraction detection based on individual driving styles and habits.
- Alert Feedback Mechanism: Developing adaptive alert systems that adjust intensity and timing based on the severity and frequency of distraction.

- Dataset Expansion: Collecting and incorporating more diverse, real-world driving scenarios to improve model robustness across demographics and vehicle types.

REFERENCES :

1. R. Javed, S. U. Rehman, M. U. Khan, M. Alazab, and T. Reddy, CANintelliIDS: Detectingin-vehicleintrusion attacks onacontrollerarea network using CNN and attention-based GRU, IEEE Trans. Netw. Sci. Eng., vol. 8, no. 2, pp. 14561466, Apr. 2021.
2. A. R. Javed, M. Usman, S. U. Rehman, M. U. Khan, and M. S. Haghighi, Anomaly detection in automated vehicles using multistage attention based convolutional neural network, IEEE Trans. Intell. Transp. Syst., vol. 22, no. 7, pp. 42914300, Jul. 2021.
3. K.-L.-A. Yau, H. J. Lee, Y.-W. Chong, M. H. Ling, A. R. Syed, C. Wu, and H. G. Goh, Augmented intelligence: Surveys of literature and expert opinion to understand relations between human intelligence and arti cial intelligence, IEEE Access, vol. 9, pp. 136744136761, 2021.
4. X. Chen, C. Wu, Z. Liu, N. Zhang, and Y. Ji, Computation of oading in beyond 5Gnetworks: A distributed learning framework and applications, IEEE Wireless Commun., vol. 28, no. 2, pp. 5662, Apr. 2021.
5. Edgar Snyder and Associates. Texting and Driving Accident Statistics. Accessed: Aug. 1, 2022. [online]. Available: https://www.edgarsnyder. com/car-accident/cause-of-accident/cell-phone/cell-phone-statistics.html