A Hybrid Learning Approach for Real-Time Distracted Driver Behavior Detection Using Transfer Learning

Sammita Abhay[1]; Vathsalya V.[2]; Veena B.[3]

[1]Computer Science and Engineering
Global Academy of Technology
Bangalore, India

[2]Computer Science and Engineering
Global Academy of Technology
Bangalore, India

[3]Computer Science and Engineering
Global Academy of Technology
Bangalore, India

Publication Date: 2025/12/11

Abstract: Distracted driving remains a primary contributor to many global road accidents, motivating advanced detection technologies for vehicle safety systems.

This study introduces a hybrid approach using handcrafted features (HOG, Local Binary Patterns) and deep transfer learning (VGG16), combined with SVM, Random Forest, and XGBoost classifiers. Experiments with the State Farm Distracted Driver Detection dataset achieved accurate recognition of ten behaviors, such as texting and reaching behind. And our Streamlit-based application enables real-time, user-friendly prediction.

The end resulting system is scalable, interpretable, and efficient, showing strong potential for AI powered better transportation solutions, making it suitable for practical safety applications.

Keywords: *Distracted Driving, Deep Learning, Hybrid Architecture, Transfer Learning, VGG16, Convolutional Neural Network (CNN), Histogram Of Oriented Gradients (HOG), Local Binary Pattern (LBP), Machine Learning, Ensemble Classification, Xgboost, Random Forest, Support Vector Machine (SVM), Feature Fusion, Driver Monitoring System, Road Safety, Human–Computer Interaction, Intelligent Transportation Systems (ITS), Computer Vision, Real-Time Image Recognition, Model Deployment, Streamlit Application, Tensorflow, and Autonomous Vehicles.*

How to Cite: Sammita Abhay; Vathsalya V.; Veena B. (2025) A Hybrid Learning Approach for Real-Time Distracted Driver Behavior Detection Using Transfer Learning. *International Journal of Innovative Science and Research Technology*, 10(11), 2911-2914. https://doi.org/10.38124/ijisrt/25nov1132

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25nov1132

I. INTRODUCTION

Road accidents are caused by distracted driving that has increased significantly in recent years, creating a need for intelligent AI ML systems that can automatically identify unsafe driver behavior. Reports from the World Health Organization indicate that distraction contributes to a large number of road fatalities worldwide [1].

Earlier detection techniques were relied on manual observation or basic in-vehicle sensors, which were mostly inaccurate in real-world environments. With improvements in computer vision and deep learning, it has become possible to analyze driver images more reliably with the help of AI.

In this project, we develop a hybrid system that combines handcrafted features with transfer-learning-based deep features to improve both accuracy and efficiency. The final model is deployed as a Streamlit web application, allowing users to test driver-behavior detection in real time.

II. RELATED WORK AND FOUNDATIONAL LIMITATIONS

Earlier research on distracted-driver detection mainly used handcrafted features such as HOG, LBP, and Gabor filters. These methods were simple and fast but struggled to identify complex driver poses or very small behavioral differences. Later studies shifted towards deep learning models, especially CNN-based architectures and transfer-learning models like VGG16 and ResNet which were more accurate but did not support Indian roads and different lightings.

These approaches improved accuracy but required more computing power. Some recent works explored hybrid techniques that mix deep and handcrafted features to balance performance and efficiency. This project follows a similar direction by combining VGG16 features with HOG and LBP to create a more robust classification model.

III. METHODOLOGY AND HYBRID ARCHITECTURE

A. Dataset and Preprocessing

The model was trained using the State Farm Distracted Driver Dataset, which contains over 22,000 labelled images across ten categories (c0–c9). Each category represents a specific driver behaviour such as texting, talking to a passenger, drinking, or adjusting controls.

All images were resized to 160 × 160 pixels to match the input requirement of the VGG16 model. Before training, pixel values were normalized to improve stability. To increase dataset variety and reduce overfitting, several augmentation techniques were applied, including random rotation, horizontal flipping, brightness changes, and slight translations. These transformations help the model generalize better to real-world conditions such as different lighting and camera angles.

B. Feature Extraction

➢ *Handcrafted Features:*

- *Histogram of Oriented Gradients (HOG):*

HOG was used to capture the outline and posture information of the driver. It works by analyzing how pixel intensities change in localized regions, making it useful for identifying edges such as the shape of arms or head position.

- *Local Binary Patterns (LBP):*

LBP extracts texture information by comparing each pixel to its neighbors. This helps distinguish and identify fine details, such as hand gestures or even small head movements that may not be prominent in color-based features.

➢ *Deep Features:*

The VGG16 model pre-trained on ImageNet was used to extract deep visual patterns. The convolutional layers were kept frozen to retain learned representations, while the fully connected layers were fine-tuned to adapt to driver behaviour classification. This reduces training time while improving accuracy through transfer learning.

C. Classifiers and Hybrid Fusion

Three traditional machine learning classifiers were trained using the combined feature set:

- Support Vector Machine (SVM)
- Random Forest (RF)
- Extreme Gradient Boosting (XGBoost)

To build the hybrid model, the handcrafted features (HOG + LBP) were concatenated with the VGG16 deep features to form a single feature vector. Each classifier was trained separately on this fused feature representation. Finally, model outputs were averaged to form an ensemble, improving stability and overall prediction accuracy.

D. Streamlit Deployment

A simple Streamlit web application was developed to demonstrate real-time predictions. The app allows users to upload an image of a driver, after which the trained model processes the input and displays the corresponding behavior class. Pretrained model weights are stored externally and loaded during runtime, enabling deployment on lightweight platforms such as streamlit Cloud or Hugging Face Spaces without requiring GPU resources.

IV. RESULTS, OPTIMIZATION, AND CONTRIBUTION

The model was tested using validation data from the State Farm dataset. Each classifier was evaluated separately, followed by a combined hybrid ensemble. The metrics used for comparison included accuracy, F1-score, and average inference time per image.

A. Performance Results

Table 1 Performance Results

Model	Accuracy (%)	F1-Score	Inference Time (ms/img)
SVM (HOG + LBP)	84.2	0.83	18
Random Forest	86.5	0.85	22
XGBoost	88.7	0.87	24
VGG16 Transfer Learning	92.9	0.91	31
Hybrid Ensemble (Proposed)	**94.6**	**0.93**	**33**

The hybrid approach outperformed both standalone handcrafted-feature models and deep-learning models. This improvement mainly comes from combining structural features (HOG/LBP) with the semantic deep features extracted from VGG16.

B. Optimization Techniques

To improve efficiency and accuracy, the following optimizations were applied:

- Principal Component Analysis (PCA) was used to reduce the dimensionality of handcrafted features before fusion.
- Adam optimizer with a tuned learning rate allowed faster convergence during fine-tuning of VGG16.
- Early stopping was used to avoid overfitting and stabilize validation accuracy.
- Feature scaling ensured that handcrafted and deep features contributed proportionally during training.

These adjustments allowed the hybrid model to maintain high accuracy while keeping inference time suitable for real-time applications.

C. Key Contributions

The main contributions of this project are:

- A hybrid feature extraction pipeline combining HOG, LBP, and transfer-learning-based deep features.
- A fused ensemble model that improves overall classification accuracy with minimal additional computation.
- A practical and lightweight Streamlit web application enabling real-time distracted-driver prediction.
- A deployment-friendly architecture that can be extended to mobile or embedded platforms.

V. SECURITY AND SAFETY CONSIDERATIONS

The system processes only the images provided by the user and does not store any inputs or outputs, ensuring data privacy. All predictions are generated locally or on trusted hosting platforms.

To promote safe usage:

- This application avoids collecting continuous camera feeds unless the user gives clear proper permission.
- Model files loaded at runtime are kept in protected storage to prevent unauthorized modification.
- Ethical considerations such as fairness, transparency, and responsible AI guidelines are followed as recommended by international standards.

VI. COMPARATIVE ANALYSIS AND FUTURE TARGETS

When compared to similar studies conducted on the same dataset, the proposed hybrid model demonstrates competitive performance while requiring fewer computational resources.Prior CNN-only approaches typically report around 92% accuracy. Hybrid CNN-handcrafted systems from earlier studies achieve approximately 94% accuracy.

The proposed method achieves 94.6%, aligning with the best-performing models while maintaining interpretability and efficient inference.

➢ *Future Work*

Potential improvements include:

- Including lighter architectures such as MobileNetV3 for faster mobile deployment.
- Applying quantization to reduce model size without losing accuracy.
- Extending the system to process real-time video feeds instead of single images.
- Adding Grad-CAM or similar visualization techniques to help users understand why a prediction was made.
- Deploying the model on edge devices like Raspberry Pi, making it suitable for in-vehicle integration.

VII. RESEARCH METHODOLOGY AND FUTURE VENUES

A. Research Methodology

The project followed a structured workflow to ensure consistent model development and evaluation. The main stages are:

- *Data Collection and Cleaning:*

The State Farm Distracted Driver dataset was gathered and inspected for corrupted or inconsistent images. Any unusable files were removed before training.

- *Preprocessing and Augmentation:*

Images were resized, normalized, and augmented using rotations, flips, brightness shifts, and translations to increase diversity and reduce overfitting.

- *Feature Extraction:*

Handcrafted features (HOG and LBP) were extracted first, followed by deep features using the VGG16 transfer-learning model. These two feature sets were later merged.

- *Model Training and Evaluation:*

Classifiers such as SVM, Random Forest, and XGBoost were trained individually using the fused features. Their results were compared based on accuracy, F1-score, and inference time.

- *Optimization:*

Hyperparameters were tuned, and dimensionality reduction using PCA was applied to improve speed and reduce redundancy in the feature vector.

- *Deployment and Testing:*

The best-performing hybrid model was integrated into a Streamlit application for real-time predictions. User testing was performed to confirm reliability on unseen images.

B. Future Venues of Development

Several improvements can extend the project into more advanced applications:

- *Real-Time Video Analysis:*

Expanding the system from static images to continuous video input with tracking algorithms.

- *Edge Deployment:*

Converting the model for lightweight devices such as Raspberry Pi or automotive ECUs using quantization and pruning.

- *IoT Integration:*

Sending real-time distraction alerts to fleet-management dashboards.

- *Explainability:*

Adding Grad-CAM or similar visualization tools to highlight image regions influencing predictions, improving transparency.

- *Model Generalization:*

Incorporating additional datasets or synthetic image generation to enhance robustness under different lighting and camera positions.

VIII. CONCLUSION

This project presents a hybrid learning approach for detecting distracted driver behaviour using a combination of handcrafted features and transfer-learning-based deep features. The fusion of these representations improves accuracy while keeping the model computationally efficient. The results demonstrate that the hybrid ensemble performs better than individual models, making it suitable for practical road-safety applications. The implementation of a Streamlit web interface further shows that the system can be deployed easily for real-time use. Future work will focus on optimizing the model for embedded devices and extending it to live video analysis.

REFERENCES

[1]. World Health Organization, *Global Status Report on Road Safety*, 2023. Available: https://www.who.int/publications
[2]. N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 886–893. doi: 10.1109/CVPR.2005.177
[3]. K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv preprint*, 2014. doi: 10.48550/arXiv.1409.1556
[4]. R. Verma, S. Singh and P. Kumar, “Hybrid CNN–HOG Approach for Driver Distraction Detection,” *IEEE Access*, vol. 8, pp. 112435–112447, 2020. doi: 10.1109/ACCESS.2020.3002874
[5]. European Commission, *Ethics Guidelines for Trustworthy AI*, 2021. Available:https://digital-strategy.ec.europa.eu
[6]. A. Chaudhary and V. Balasubramanian, “Driver Distraction Detection Using Transfer Learning,” in *Artificial Intelligence in Transportation*, Springer, 2021, pp. 239–255. doi: 10.1007/978-3-030-64583-0_12
[7]. H. Zhang, Y. Wang and L. Yu, “Efficient Driver Posture Classification with CNN–HOG Fusion,” *IEEE Intelligent Transportation Systems (ITS)*, 2022. doi: 10.1109/ITSC55140.2022.9922083