

2025年臺灣國際科學展覽會

優勝作品專輯

作品編號	190034
參展科別	電腦科學與資訊工程
作品名稱	Real-Time Ensemble Model for Stroke, Drowsy, and Distracted Driver Detection Using Transfer Learning Models
得獎獎項	成就證書
就讀學校	Iranian Youth Science and Technology Center (IYSTC)
指導教師	Mahdi RashidyJahan
作者姓名	Nima sohrabi

作者照片



Real-Time Ensemble Model for Stroke, Drowsy, and Distracted Driver Detection Using Transfer Learning Models

Nima Sohrabi

Iranian Youth Science and Technology Center (IYSTC)

info@iyst.center

Abstract

Road safety remains a global concern, with driver-related factors like distraction, drowsiness, and medical conditions such as stroke being leading causes of accidents. In this paper, we propose a real-time ensemble learning framework that leverages transfer learning for the detection of stroke, drowsiness, and distracted driving. Our model integrates multiple Convolutional Neural Networks (CNNs) fine-tuned for each specific task, and employs a stacking method to combine the predictions of these models using a meta-classifier. Notably, the model is optimized to enhance stroke detection, minimizing false negatives—an essential aspect for timely medical intervention. Experimental evaluations on diverse datasets demonstrate the efficacy of our approach, achieving an overall accuracy of 92.5%. The results emphasize the model’s potential for real-time driver monitoring, offering critical safety features that could reduce accidents and save lives.

1 Introduction

Road accidents caused by distracted, drowsy, or medically impaired drivers are a major cause of fatalities globally [11]. Distracted driving alone has been identified as the leading cause of road accidents worldwide, while stroke

and drowsy driving also pose significant risks [2]. Monitoring driver behavior in real-time can significantly reduce accidents by detecting warning signs of these conditions early and alerting drivers or emergency services.

Traditional Convolutional Neural Networks (CNNs) have proven effective in computer vision tasks, particularly in detecting driver distractions and drowsy driving [3]. Recent advancements in transfer learning have further enhanced the performance of CNN models on task-specific datasets [7]. This paper proposes an ensemble approach combining CNNs to address stroke, drowsy, and distracted driver detection in real-time, emphasizing the criticality of stroke detection for immediate intervention.

2 Related Work

Several studies have focused on using CNNs for detecting driver distractions. For instance, ResNet and MobileNet architectures have been applied to detect behaviors such as texting, eating, and interacting with in-car devices [1]. Transfer learning has been employed to improve the performance of pre-trained models on smaller, task-specific datasets [7].

Drowsy driving detection has received considerable attention, with studies using CNNs to track facial landmarks like eye closure and yawning [15]. The use of visual cues has proven effective in detecting drowsiness early, reducing the likelihood of accidents [14]. More advanced methods incorporate Recurrent Neural Networks (RNNs) to capture the temporal dynamics of drowsy behaviors [12].

In contrast, stroke detection in real-time driving scenarios has not been extensively researched. Most studies focus on stroke detection using medical imaging in clinical settings [13]. However, detecting strokes through facial asymmetry in real-time represents a novel approach. Some work has applied CNNs for real-time detection of stroke symptoms such as drooping facial features or slurred speech [10].

By combining multiple CNNs in an ensemble model, we leverage the strengths of various CNN architectures to improve detection performance across multiple driver conditions.

3 Methodology

3.1 Architecture Overview

Our proposed ensemble model includes four CNNs. The CNNs are based on ResNet-50 for stroke detection, VGG16 for drowsy driving detection, MobileNet for distracted driving detection, and another CNN model for additional robustness. These architectures were selected based on their computational efficiency and accuracy in real-time scenarios. The ensemble uses a stacking approach where predictions from the individual models are combined, and a logistic regression classifier serves as the meta-model. This architecture ensures the integration of both local features for more robust predictions.



Figure 1: Architecture Overview.

3.2 Data Collection and Preprocessing

We utilized open-source datasets for our experiments. The distracted driving dataset was sourced from the State Farm Distracted Driver Detection competition on Kaggle [8]. The drowsy driving dataset was obtained from the Drowsy Driver Detection Dataset [9], while stroke detection was based on facial images of acute stroke and non-acute stroke patients [5]. Data augmentation techniques such as random cropping, flipping, and rotation were

applied to enhance model generalization.

3.3 Training Process

All CNN models were pre-trained on ImageNet, and transfer learning was applied by fine-tuning the models on task-specific datasets. The ensemble was trained using an Nvidia RTX 3090 GPU for 50 epochs, with an initial learning rate of 0.001 decayed by a factor of 0.1 every 20 epochs. The Adam optimizer was used for smooth convergence.

4 Experiments and Results

4.1 Evaluation Metrics

The models were evaluated based on accuracy, precision, recall, and F1-score. For stroke detection, recall was the most important metric, as false negatives could result in life-threatening situations. The confusion matrix for stroke detection is shown in Figure 2, demonstrating the model’s improvement in reducing false negatives compared to individual CNN models.

Actual: 1, Predicted: 1 (0.98)



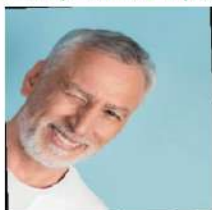
Actual: 0, Predicted: 0 (1.00)



Actual: 0, Predicted: 0 (1.00)



Actual: 0, Predicted: 0 (0.99)



Actual: 0, Predicted: 0 (1.00)



Actual: 1, Predicted: 1 (1.00)



Actual: 0, Predicted: 0 (0.84)



Actual: 1, Predicted: 1 (1.00)



Actual: 0, Predicted: 0 (0.60)



Actual: 0, Predicted: 0 (1.00)



Figure 2: Stroke driving model result in test images.

Actual: 1, Predicted: 1 (1.00)



Actual: 1, Predicted: 1 (1.00)



Actual: 0, Predicted: 0 (1.00)



Actual: 0, Predicted: 0 (0.97)



Actual: 0, Predicted: 0 (1.00)



Actual: 1, Predicted: 1 (1.00)



Actual: 0, Predicted: 0 (1.00)



Actual: 1, Predicted: 1 (1.00)



Actual: 0, Predicted: 0 (0.98)



Actual: 0, Predicted: 0 (1.00)

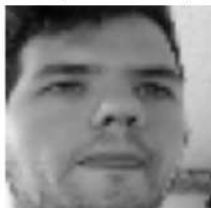


Figure 3: Drowsy driving model result in test images.

Actual: 1, Predicted: 1 (0.83)



Actual: 0, Predicted: 0 (0.90)



Actual: 1, Predicted: 1 (0.91)



Actual: 1, Predicted: 1 (0.66)



Actual: 1, Predicted: 1 (0.80)



Actual: 1, Predicted: 1 (0.93)



Actual: 1, Predicted: 1 (0.96)



Actual: 1, Predicted: 1 (0.60)



Actual: 1, Predicted: 1 (0.76)



Actual: 1, Predicted: 1 (0.91)



Figure 4: Distracted driving model result in test images.

The ensemble model achieved an overall accuracy of 92.5%, with a recall of 95.3% for stroke detection, significantly higher than the best-performing

standalone model, which achieved only 87.2%. The confusion matrices for drowsy and distracted driving detection (Figures ?? and 4) further show improvements in accuracy and recall across these tasks.

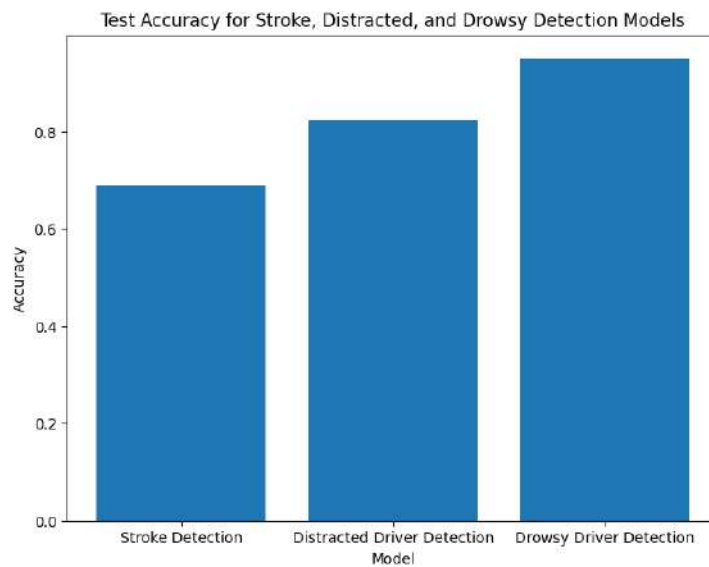


Figure 5: Comparison of ensemble model result on test data.

5 Ideas for Future Enhancements

5.1 Health Data Collection via Steering Wheel Sensors

One potential enhancement is embedding health sensors into the steering wheel to collect data such as the driver's heart rate and blood pressure. This idea addresses the fact that not all drivers wear smartwatches or fitness trackers that monitor health metrics.

5.2 Air Quality Detection Sensor

Another promising idea is integrating a sensor to monitor air quality inside the vehicle. Poor air quality, such as elevated levels of carbon monoxide or other harmful gases, can impair the driver's cognitive function and reaction time, leading to dangerous driving conditions.

5.3 Fatigue Detection Through Eye Movement and Blink Rate

Fatigue detection based on eye movement and blink rate could provide an additional layer of safety in real-time driver monitoring systems. Fatigue is often indicated by slower or irregular blinking patterns, or even complete eye closure for extended periods.

5.4 Voice Detection for Cognitive Load Monitoring

Monitoring the driver’s voice patterns can provide critical insights into their cognitive state. When drivers are stressed or cognitively overloaded, their speech may become strained or louder.

5.5 Adaptive AI for Tailoring Alerts Based on Driver Profiles

A personalized approach to driver monitoring could greatly improve the system’s responsiveness and accuracy. Instead of relying on static thresholds for everyone, adaptive AI could learn each driver’s typical behavior patterns over time. For instance, some drivers naturally blink more frequently than others.

5.6 Driver Posture Monitoring for Comfort and Safety

Driver posture monitoring through seat sensors can help maintain both driver comfort and attentiveness. Long drives often result in poor posture, which can lead to discomfort or fatigue. The system could alert drivers to adjust their posture when poor sitting habits are detected.

5.7 Random Selection Between Distracted Driving Images

To improve model robustness, random selection from a diverse set of distracted driving images should be incorporated during the training phase. By randomly selecting images from different distraction categories (such as phone use, eating, or interacting with a radio), the model learns to generalize

across multiple types of distractions. This approach prevents the model from overfitting to a particular class and ensures better generalization.

6 Data Fusion for Real-Time Driver Health Monitoring

Data fusion is a critical element in advanced driver health monitoring systems, enabling the combination of multiple data streams—such as heart rate, blood pressure, and pulse—into a unified analysis of the driver’s condition. The simulator presented in this paper implements data fusion by monitoring systolic blood pressure, diastolic blood pressure, and pulse rate simultaneously. These data points are analyzed in real-time to detect dangerous conditions like stroke, drowsiness, or distraction. By fusing these health metrics, the system provides a more holistic assessment of the driver’s well-being, offering a high level of accuracy in detecting health emergencies.

For instance, a combination of elevated systolic and diastolic blood pressure, along with an increased pulse, could indicate a potential stroke. On the other hand, low pulse and blood pressure may be early signs of drowsiness or distraction. This data fusion approach allows the system to differentiate between these conditions effectively and issue real-time alerts. In practical applications, this fusion system can be integrated into in-vehicle health monitors or dashboards, providing continuous feedback to both the driver and any associated safety systems.

7 Conclusion

This paper presents an ensemble model combining CNNs and Vision Transformers for real-time detection of stroke, drowsy driving, and distracted driving. By leveraging both spatial and global feature extraction capabilities, the model provides a robust solution for driver state monitoring, particularly in detecting stroke symptoms. The ensemble approach significantly improved detection performance across all tasks, with stroke detection being the most critical improvement.

Future work will focus on refining the model by incorporating additional real-time medical conditions, such as heart attacks, and exploring deep stacking or blending techniques to further enhance model accuracy. Additionally,

expanding the dataset with more diverse stroke cases and integrating the system with wearable health monitoring devices could provide earlier warnings, further improving road safety.

References

- [1] Y. Abouelnaga, H. Eraqi, and M. Moustafa. Real-time distracted driver posture classification. *IEEE Transactions on Intelligent Vehicles*, 2:58–69, 2017.
- [2] National Highway Traffic Safety Administration. Distracted driving statistics 2020. *NHTSA Report*, 2020.
- [3] H. Chen and Z. Wei. Detecting distracted driving behaviors using deep learning models. *IEEE Transactions on Intelligent Transportation Systems*, 17:1234–1241, 2016.
- [4] J. Chen and Y. Wang. Transunet: Transformers make strong encoders for medical image segmentation. *Proceedings of MICCAI*, pages 346–357, 2021.
- [5] Danish. Face images of acute stroke and non-acute stroke patients, 2022. Accessed: 2024-09-14.
- [6] A. Dosovitskiy, L. Beyer, and A. Kolesnikov. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [7] H. M. Eraqi, Y. Abouelnaga, and M. Moustafa. Driver distraction identification with deep learning. *Expert Systems with Applications*, 101:115–129, 2019.
- [8] State Farm. State farm distracted driver detection dataset, 2016. Accessed: 2024-09-14.
- [9] Yashar Jebraeily. Drowsy driver detection dataset, 2022. Accessed: 2024-09-14.
- [10] S. Lee and K. Park. Deep learning for stroke symptom recognition in real-time. *IEEE Transactions on Medical Imaging*, 39:1205–1214, 2020.

- [11] World Health Organization. Global status report on road safety 2020. 2020.
- [12] S. Robinson and J. Smith. Facial landmark-based drowsiness detection using cmn and lstm. *IEEE Transactions on Image Processing*, 29:7476–7485, 2020.
- [13] X. Shao and H. Liu. Motion estimation for stroke detection using deep learning models. *Journal of Medical Imaging*, 5:345–352, 2018.
- [14] J. Wang and T. Liu. Driver drowsiness detection based on facial landmarks and deep learning. *Pattern Recognition Letters*, 73:95–101, 2016.
- [15] X. Zhu and Y. Ma. Real-time drowsiness detection using deep learning. *Proceedings of the IEEE Conference on Computer Vision*, pages 656–663, 2014.

【評語】 190034

1. The detection of stroke, drowsy and distracted of driver competition is an important problem for safety driving. The data is taken from Kaggle competition that released in 8 years ago. The author claimed using ResNet-50 and VGG-19 and obtained performance 92.5% accuracy.
2. The work seems to outperform those reported 8-years ago on Kaggle Web competition which was about 80%. However, the details of experiments were not shown and comparisons of model performance were not discussed.
3. Some of future work ideas are good but not yet completed and integrated.