

RESEARCH ON NON-INVASIVE DROWSINESS DETECTION METHOD FOR HARVESTER DRIVERS BASED ON MULTI-FEATURE FUSION

基于多特征融合的非侵入性收获机驾驶员困倦检测方法的研究

Wei LIU, Kai RONG, Yi NIU, Ruixue LI, Haoxuan HONG, Guohai ZHANG

Collage of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo, China

Tel: +86 15965534882; E-mail: guohaizhang@sdut.edu.cn

Corresponding author: Guohai Zhang

DOI: <https://doi.org/10.35633/inmateh-77-12>

Keywords: drowsiness detection, CNN, LSTM, harvester operator, deep learning, face images

ABSTRACT

Driver drowsiness severely impairs the normal operation of harvesters, leading to casualties and economic losses. Effectively detecting driver drowsiness in harvesters remains a significant challenge. This paper introduces a lightweight convolutional neural network that identifies driver drowsiness in harvester operators by analyzing the driver's eyes, mouth, and head posture. The model comprises a lightweight CNN, a Long Short-Term Memory (LSTM) network, and an attention layer, achieving high efficiency and low latency. Experimental results demonstrate that the CNN-LSTM-Attention model effectively balances accuracy and computational efficiency, enabling rapid and precise drowsiness detection. This approach significantly improves safety during combine harvester operation.

摘要

驾驶员困倦严重影响收获机正常作业，造成人员伤亡和经济损失。对收获机驾驶员困倦的有效检测依然是重大挑战。本文介绍了一种轻量化卷积神经网络，利用驾驶员眼睛、嘴巴以及头部姿态对收获机驾驶员困倦进行识别。该模型由轻量化的 CNN、长短时记忆网络以及一个注意力层组成，具有高效、低延迟的特点。实验结果表明，CNN-LSTM-Attention 模型很好的平衡了结果准确性和计算效率，能够快速准确的识别困倦，对收获机驾驶安全产生重大影响。

INTRODUCTION

The International Labor Office's publication "Safety and Health in Agriculture: Code of Practice" highlights that agricultural machinery is diverse and hazardous, emphasizing that safety features must be prioritized in equipment operation. This underscores the international consensus on prioritizing agricultural machinery safety. Against the backdrop of highly mechanized and intensive agricultural production, driver drowsiness resulting from prolonged, high-intensity harvesting operations has become a major safety hazard in agricultural machinery. Combine harvesters play a pivotal role in modern agricultural production systems (Dias et al., 2024). However, compared to automotive driving environments, their operating environments are notably more complex and physically demanding. Particularly during peak harvest seasons, operators face intense workloads and extended hours, making them highly susceptible to drowsiness that jeopardizes both operational safety and quality (Indri et al., 2018). While driver drowsiness detection technology has matured in automotive research, significant progress has also been made in the agricultural machinery sector. However, existing studies predominantly focus on drowsiness detection for operators of general agricultural equipment like tractors, while research on drowsiness detection for operators of more complex combine harvesters remains relatively scarce (Zhu et al., 2016). Consequently, there exists significant innovation potential and necessity for developing drowsiness detection models tailored to combine harvester operators (Sikander and Anwar, 2019).

Driver drowsiness detection systems can be broadly categorized into four types: computer vision-based, physiological feature-based, vehicle behavior-based, and hybrid-based, as shown in Figure 1. In recent years, with the rapid advancement of deep learning technology, research focus has increasingly shifted toward non-invasive computer vision methods that cause minimal driver distraction (Koay et al., 2022). These methods detect drowsiness by capturing visually observable signs of drowsiness. Within this category, eye and mouth states are commonly selected as effective feature indicators. However, combine harvester operators exhibit significant changes in head posture during drowsiness, such as head drooping and frequent nodding.

These changes possess a quantitative correlation with drowsiness yet are often overlooked in existing research (Sengupta, 2024). Particularly in combine harvester operations, drivers must frequently turn their heads to monitor crop distribution and adjust paths, resulting in significantly higher head movement angles and frequencies compared to other driving scenarios. This leads to increased neck muscle strain and accumulated mental drowsiness, posing higher risks of driving drowsiness during prolonged operations (Dipti., 2025). Furthermore, the complex and harsh operating environment of harvesters, coupled with unstable lighting conditions, often necessitates drivers wearing protective gear like sunglasses and masks. These factors interfere with traditional facial feature capture methods based solely on eyes and mouth, reducing their detection robustness. Therefore, this study proposes a multi-feature fusion method for detecting driver drowsiness in combine harvesters. Building upon eye and mouth state detection, this approach integrates head pose estimation technology. By fusing multiple facial features—including eye and mouth states alongside head pose—it aims to enhance the generalization capability and accuracy of drowsiness detection models. This adaptation better accommodates the complex and variable operating environment of combine harvesters, thereby effectively safeguarding harvest quality and operational safety.

This study constructs a lightweight CNN-LTSM-Attention model. By removing redundant convolutional layers and adding LSTM layers to capture temporal features, the model enhances computational efficiency while maintaining satisfactory detection accuracy. Details are as follows: (1) A lightweight CNN model was developed to detect drowsiness and alertness states in combine harvester operators; (2) This model comprises a lightweight CNN network with two convolutional layers, an LSTM layer, and an Attention layer. The lightweight CNN network excels at analyzing a balance between computational accuracy and computational efficiency; (3) To validate the proposed CNN-LSTM-Attention model, performance was evaluated based on both accuracy and efficiency. Its effectiveness is verified through comparisons with independent CNN and LSTM networks, while model accuracy is benchmarked against mainstream deep learning networks ResNet101 and VGG16.

The remainder of this paper is organized as follows: First, existing research on driver drowsiness is reviewed, and the research framework, dataset, and model architecture are presented. Building upon this foundation, optimal performance is achieved through model parameter adjustment. To validate the model's superiority, comparative evaluations are conducted against existing models. Finally, based on the experimental results, an in-depth discussion is provided, key findings and conclusions are summarized, and potential future research directions are outlined.

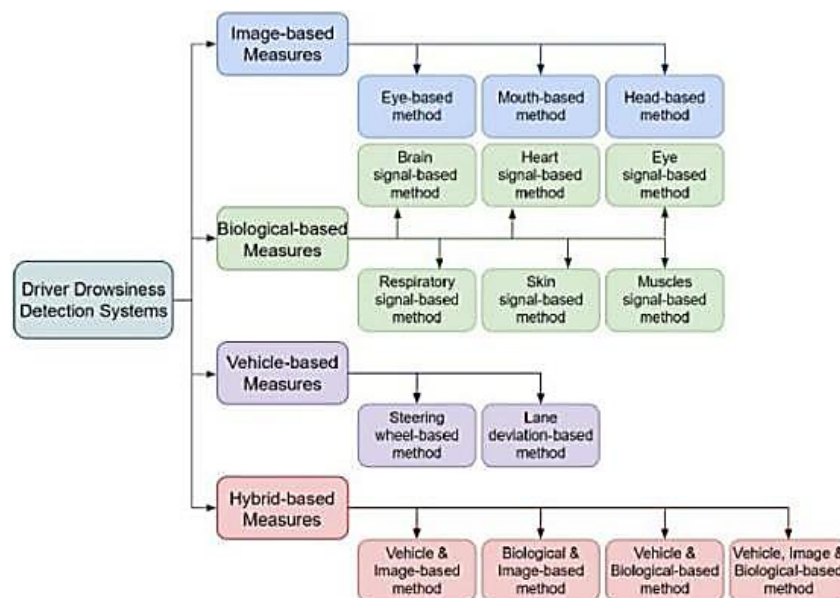


Fig. 1 – Driver drowsiness detection measures

(Albadawi, Y. et al., 2024)

LITERATURE SURVEY

Physiological signals serve as objective quantitative indicators of drowsiness, offering advantages of real-time monitoring and non-invasiveness (Ren et al., 2023). Current research on physiological signals for detecting drowsiness in combine harvester operators primarily focuses on electrocardiogram (ECG) signals and surface electromyography (sEMG) signals.

Kong *et al.* (2008) introduced physiological signals into the agricultural machinery field by comprehensively evaluating operational drowsiness using heart rate and working duration. Chen identified correlations between tractor operator drowsiness and electromyographic signals (Chen *et al.*, 2016). While these studies represent significant progress, both detection methods involve invasive devices that interfere with normal operation. Furthermore, physiological signals generate large data volumes with low computational efficiency. Consequently, this approach is unsuitable for harvesters requiring high driving precision and limited computational resources.

Computer vision technology detects driver drowsiness by capturing facial behavioral characteristics (Shang *et al.*, 2022), emerging as a core technical approach to enhancing road safety (Sun *et al.*, 2023). Compared to physiological signals, this method enables non-contact detection. Research focuses on feature extraction accuracy, model generalization capability, and real-time deployment adaptability. Zhan *et al.* integrated driver eye movement and head motion features to achieve a comprehensive assessment of driver drowsiness levels (Zhang *et al.*, 2018). The PERCLOS criterion is commonly used as a drowsiness assessment benchmark. When integrated with mouth features, the system's fault tolerance significantly improves, meeting real-time detection requirements (Albadawi *et al.*, 2022). However, this approach demands certain image clarity and driving environment stability. Accurate capture of facial key features forms the foundation of detection (Guan, 2024). The MediaPipe framework can extract 52 numerical facial features, including blink rate and eye-opening degree (Joshi *et al.*, 2025). Among traditional machine learning classifiers trained on these features, random forests achieve a test accuracy of 97%, offering optimal performance balance between drowsiness and alertness categories. The drawback lies in the substantial computational data volume and poor real-time capability. Deep learning models represent the core driver of performance breakthroughs. Deep convolutional neural networks (DCNNs) demonstrate outstanding performance across multiple datasets, achieving classification accuracies as high as 99.6%. They enable real-time classification of video frames using OpenCV (Zaman *et al.*, 2024). Transfer learning further enhances model efficiency. The Xception model combined with transfer learning achieved 99.36% accuracy in driver facial image classification (Muthuraja *et al.*, 2024). ResNet50 achieved a validation accuracy of 95% on image datasets through transfer learning, with a recall rate of 0.97 for drowsiness detection, effectively reducing false negatives (Joshi *et al.*, 2025). This demonstrates that constructing suitable drowsiness recognition models and selecting appropriate features are crucial for enhancing the robustness and accuracy of drowsiness detection systems, with transfer learning serving as one of the key methods for building effective models.

MATERIALS AND METHODS

To detect driver drowsiness in harvesters, this study proposes a deep learning approach integrating a lightweight convolutional neural network (CNN), attention mechanism, and long short-term memory (LSTM) network. The model identifies driver drowsiness by visually learning features such as eye, mouth, and head posture, aiming to achieve high operational efficiency and classification accuracy. The research involves several key stages: data acquisition, preprocessing, dataset segmentation, model design, training, and performance evaluation. Figure 2 illustrates how the model detects drowsiness by learning visual features of eye, mouth, and head posture. Furthermore, this study employs transfer learning to freeze and fine-tune two mainstream deep learning models—VGG16 and ResNet101—and conducts model performance evaluations. Using metrics such as accuracy, precision, recall, and F1-Score, it validates the superior performance of the improved models.

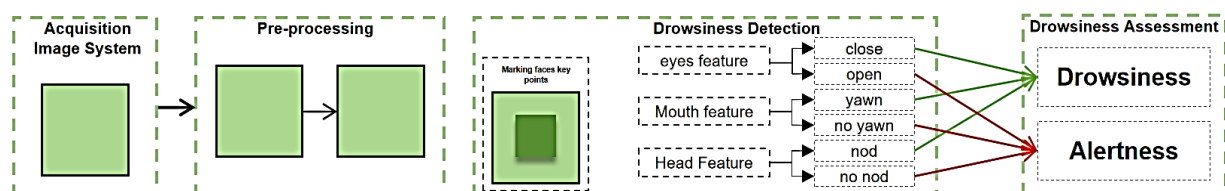


Fig. 2 – Drowsiness Detection System with Facial Multi-Information Fusion

Datasets

The lightweight CNN-LSTM model developed in this paper, along with the deep learning framework for visual image detection (VGG16, RenNet50, and DenseNet), requires training and testing using a dataset comprising full-head image features of combine harvester operators during operation. This enables more precise expression of driver drowsiness levels.

The dataset includes images representing the states of drowsy and alert combine harvester operators, with experiments conducted with volunteer consent. Between 2024 and 2025, facial and head images of 22 drivers—varying in age, gender, work duration, and drowsiness levels—were collected in Shandong Province while harvesting diverse crops. To ensure model generalization, objective factors like gender and environment were balanced. Additionally, to mirror real-world conditions, volunteers were permitted to wear masks, sunglasses, or regular glasses. This dataset was captured unobtrusively in authentic work environments, avoiding psychological pressure on volunteers. The diversity of driving conditions and demographic characteristics enhances the dataset's authenticity, improving model performance and strengthening its applicability in real-world scenarios.

Using expert scoring to evaluate the degree of drowsiness changes over time in 22 drivers, most drivers transitioned from alertness to drowsiness within 30 minutes. After removing images of drivers who did not exhibit drowsiness, the drowsiness states of 20 drivers within 30 minutes were recorded, as shown in Figure 3. To capture the dynamic transition of drowsiness, key segments were extracted from each video—specifically, the phase where drivers transitioned from alertness to drowsiness. During this process, computer vision techniques converted videos into image sequences, ensuring each driver contributed an equal number of alert and drowsy state images to the dataset to prevent model bias caused by sample imbalance. Examples of alert and drowsy states are clearly illustrated in Figure 4. To adapt to the models used in this study, image dimensions were standardized by resizing each image to 224×224 pixels. Pixel values underwent normalization to initialize network weights.

To ensure the effectiveness of model training and evaluation, the dataset was divided into training, validation, and test subsets at an 80:10:10 ratio, following the strategy proposed by Sahu et al. (2021). Ultimately, images from 16 drivers were selected for training, comprising 48,000 images. The validation and test set each contained images from two drivers, comprising 6,000 images per set. The temporally partitioned dataset facilitates model training by enabling simultaneous capture of spatial and temporal features. Furthermore, this structured segmentation effectively prevents data overlap, mitigates overfitting, ensures model adaptability to novel data, and enhances generalization capabilities.

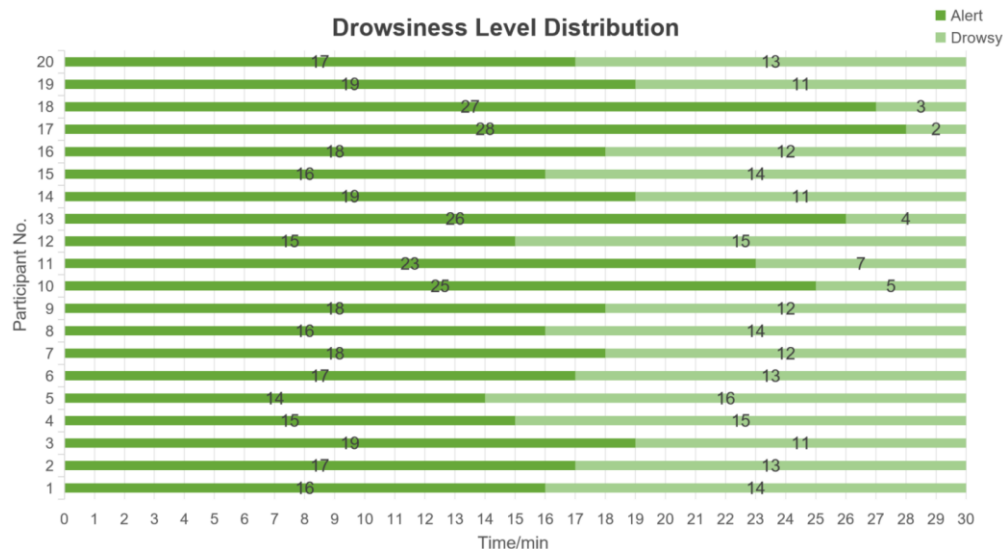


Fig. 3 – Drowsiness Level Distribution



Fig. 4 – Image Examples from the Driver Drowsiness State Dataset
(a) Normal; (b) Drowsy - Nodding; (c) Drowsy - Yawning; (d) Drowsy - Eye Closed

Model Construction

A lightweight CNN-LSTM-Attention model is proposed to detect driver drowsiness in harvesters in real time. By combining CNN's spatial feature extraction capability for static images with LSTM's ability to capture temporal relationships in image sequences, this model not only enables early drowsiness detection but also tracks the progression of drowsiness states, thereby improving detection accuracy and reducing false alarms. Specifically, a lightweight CNN architecture with two convolutional layers is first employed to extract key spatial features such as eyes, mouth, and head posture from input images. These flattened spatial features serve as input to the LSTM network. Considering the continuity of image sequences during driving and the strong correlation between adjacent frames, the LSTM models temporal characteristics of the extracted data, capturing feature changes over time—such as nodding accompanied by eye closure in consecutive frames. The model incorporates an attention mechanism to enhance focus on critical spatial and temporal features while suppressing irrelevant information. To counter overfitting, a Dropout layer follows the attention layer. The final dense layer provides two output units for classification, as illustrated in Figure 5. This architecture reduces redundant convolutional layers, improves computational efficiency, and simultaneously captures drowsiness features and their temporal variations across both spatial and temporal dimensions, thereby enhancing detection accuracy. Unlike conventional drowsiness detection for automotive driving, this model simultaneously learns multi-source information from eye, mouth, and head posture, thereby enhancing robustness.

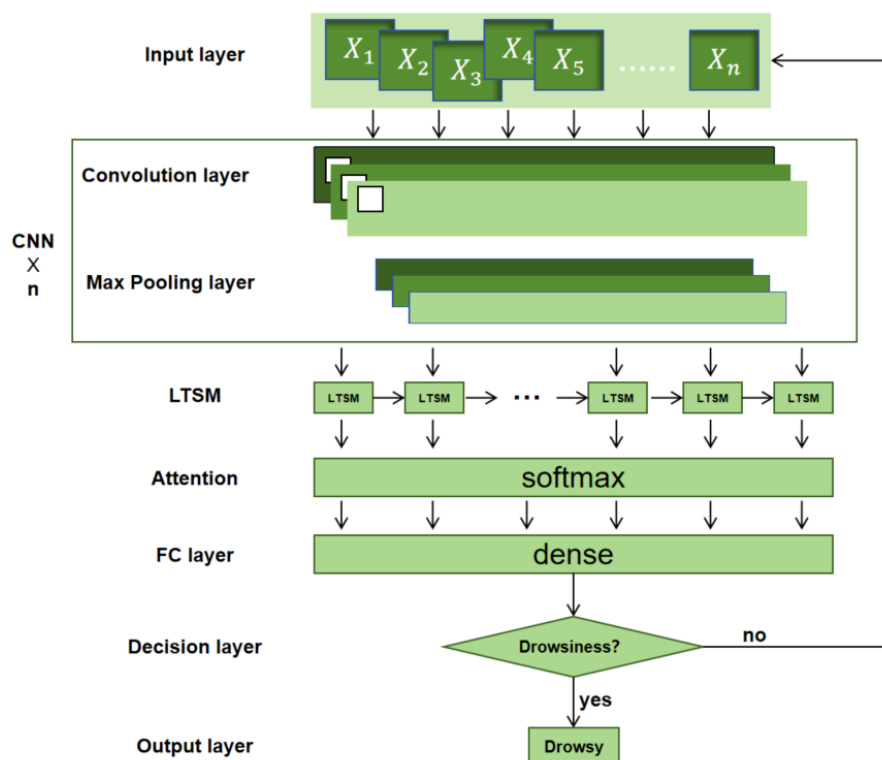


Fig. 5 – Proposed CNN-LSTM-Attention Model

Transfer Learning via popular deep learning pre-trained models

To validate the effectiveness of the proposed model, a comparative analysis was conducted in terms of accuracy and efficiency. Regarding accuracy, comparisons were made with the popular deep learning networks VGG16 and Resnet101 in the field of driver drowsiness detection to emphasize the precision of classification results. Regarding efficiency, comparisons were made with standalone CNN and LSTM networks to highlight the efficiency improvements achieved through model fusion. Training networks from scratch consumes enormous time and resources. Transfer learning serves as one approach to address this challenge, representing a collection of techniques rather than a single strategy (Atila, 2021). This process primarily relies on two steps: feature extraction and fine-tuning. The former utilizes certain pre-trained network weights to handle different problems, while the latter requires adjusting specific layers of the model to meet the unique demands of the new task.

The VGG16 model, developed by the Visual Geometry Group at the University of Oxford, stands as a classic among convolutional neural networks. Its structure is simple yet deep, delivering outstanding recognition performance (Tan et al., 2024) and finding widespread application in computer vision recognition (Mahmoud and Ahmed, 2024). Its primary characteristic lies in capturing multi-scale features from images, leading to superior learning outcomes (Yuan et al., 2020). The ResNet series of models effectively addresses the vanishing gradient and exploding gradient issues encountered during training of deep neural networks by introducing “residual blocks,” thereby learning more complex and abstract feature representations (Su et al., 2024). The residual block architecture employs skip connections, aiding both training and feature extraction (Sun et al., 2021). VGG16 and ResNet101 demonstrate formidable capabilities in image classification tasks (Di et al., 2025), frequently serving as feature extractors or foundational components for transfer learning to address domain-specific classification challenges (Zhou, 2024).

During deep learning model training, parameter tuning is crucial for optimizing model performance and training efficiency (Alameen., Alhothal., 2023). In this study, all pre-trained models were trained with a learning rate of 0.001, and the ADAM optimizer was selected to accelerate model convergence and enhance training stability (Gautam T., 2025). The output layer employs the Sigmoid function for binary classification tasks. For the proposed novel drowsiness detection model, the study initially adopted classical universal parameter settings from deep learning as the baseline (Joshi et al., 2025). After multiple iterative training rounds, targeted parameter adjustments yielded the optimal parameter combination, as shown in Table 1.

Table 1

Different hyper-parameters selected in the proposed CNN model					
Hyper-parameter	Optimizer	Learning rate	Number of epochs	Batch size	Dropout
Initial	=0.9, $\beta_2=0.999\beta_1$	0.001	15	32	0.25
Final	=0.9, $\beta_2=0.999\beta_1$	0.0008	20	32	0.3

Performance Evaluation Metrics

Model performance is evaluated using several metrics, including accuracy, precision, recall, and F1 score. Accuracy, a widely used evaluation metric in deep learning, is typically employed to assess the overall effectiveness of a classifier. Precision quantifies the ratio of correctly predicted positive samples to the total expected positive samples, reflecting the model's reliability for positive classification. To balance false positives and false negatives, recall is introduced as a judgment metric. Typically, higher model accuracy correlates with lower recall, so the *F1* score is calculated to evaluate the balanced effectiveness of the model. To compute these metrics, drowsiness is defined as positive and alertness as negative. Test results are categorized into four classes: True Positives (*TP*, drowsy samples correctly identified as drowsy), True Negatives (*TN*, normal samples correctly identified as normal), False Positives (*FP*, normal samples incorrectly predicted as drowsy), and False Negatives (*FN*, drowsy samples incorrectly predicted as normal).

Calculating these metrics enables an effective and accurate assessment of the model's computational results.

The calculation formulas are as follows:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (1)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = 2 * \frac{(TP*FP)}{(TP+FP)} \quad (4)$$

Experimental software and hardware devices

To verify the superiority of the proposed model, it is essential to select an appropriate hardware-software combination for inter-model performance comparison and practical application validation, in line with the requirements of low power consumption, anti-interference, and real-time performance for harvesters. In this study, the low-power NVIDIA Jetson Nano 4GB embedded development board, integrated with the harvester's CAN bus, was adopted as the inference terminal. This setup not only meets the low-power demand of the

harvester's on-board system but also enables signal transmission from detection results to feedback. To adapt to variations in the harvester's working environment and enhance the anti-interference capability of the model application, a binocular camera was used in conjunction with the OpenCV gamma correction algorithm to capture images of the driver's head and face. The two devices are connected via USB 3.0, with a data transmission latency of <50 ms, which satisfies the requirements of real-time monitoring. Detailed descriptions of the software and hardware are presented in the table below:

Table 2

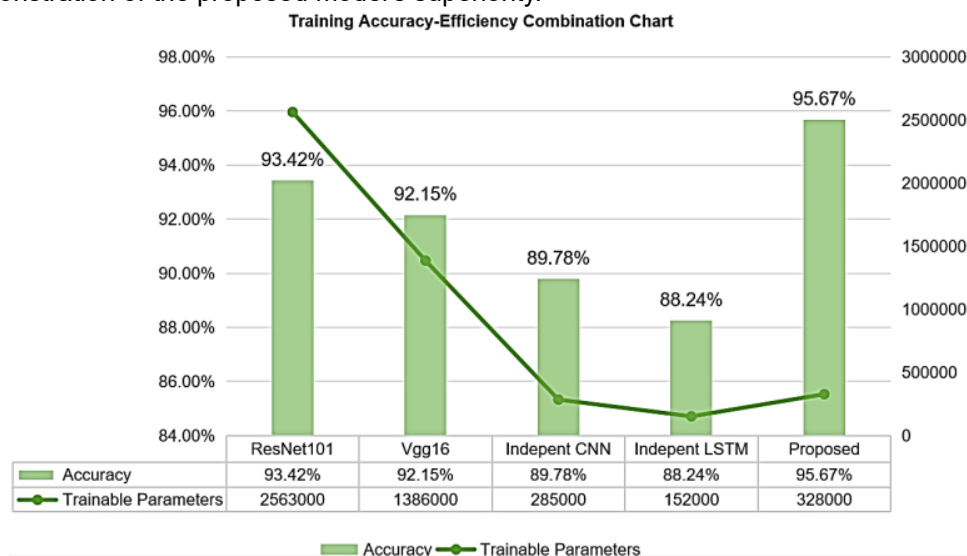
Hardware and Software Used		
Hardware Used	Data Acquisition	Binocular Camera (1624×1240 resolution)
	Model Training	Intel Core i7-13700K CPU 32 GB RAM、NVIDIA RTX 4060 Ti GPU (8 GB VRAM)
	Application Deployment	NVIDIA Jetson Nano Embedded Terminal、Harvester CAN Bus Module
Software Used	Data Processing	OpenCV 4.8
	Model Development	PyTorch 2.0.1+cu118

RESULTS AND DISCUSSIONS

Performance Evaluation Results

This study evaluates the effectiveness of the lightweight CNN-LSTM-Attention model in identifying driver drowsiness in harvesters. It compares the computational efficiency and accuracy of this model against classical deep learning models (VGG16, ResNet101) and single-structure models (standalone CNN, standalone LSTM). To ensure a fair comparison, all models were trained using identical parameter configurations. As shown in Figure 6, compared to VGG16 and ResNet101, the proposed model requires fewer trainable parameters and achieves higher computational efficiency. Compared to standalone CNN and LSTM models, the proposed model delivers more accurate computations and more reliable results.

After multiple training iterations, the results for accuracy and loss are shown in Figure 7, excluding the standalone LSTM model with lower accuracy. ResNet101 performs well with a validation accuracy of 93.42% and a validation loss of 0.152, but its 2.563 million trainable parameters indicate high computational load and low efficiency. VGG16 achieved a validation accuracy of 92.15% and a loss of 0.234, slightly underperforming ResNet101. However, its trainable parameters were reduced to 1.386 million, improving computational efficiency at the cost of accuracy. The standalone CNN and LSTM models perform poorly compared to classical deep learning models, exhibiting low accuracy for binary classification tasks. The proposed lightweight CNN-LSTM-Attention model outperforms all pre-trained architectures, achieving a validation accuracy of 95.67% and a validation loss of 0.157. Its core model contains only 328,000 trainable parameters, significantly reducing computational resource consumption. To further evaluate model performance, metrics were calculated as shown in Table 3. A confusion matrix (Figure 8) was plotted based on each model's metrics, providing a more intuitive demonstration of the proposed model's superiority.

**Fig. 6 – Training Accuracy and Efficiency Combination Chart**

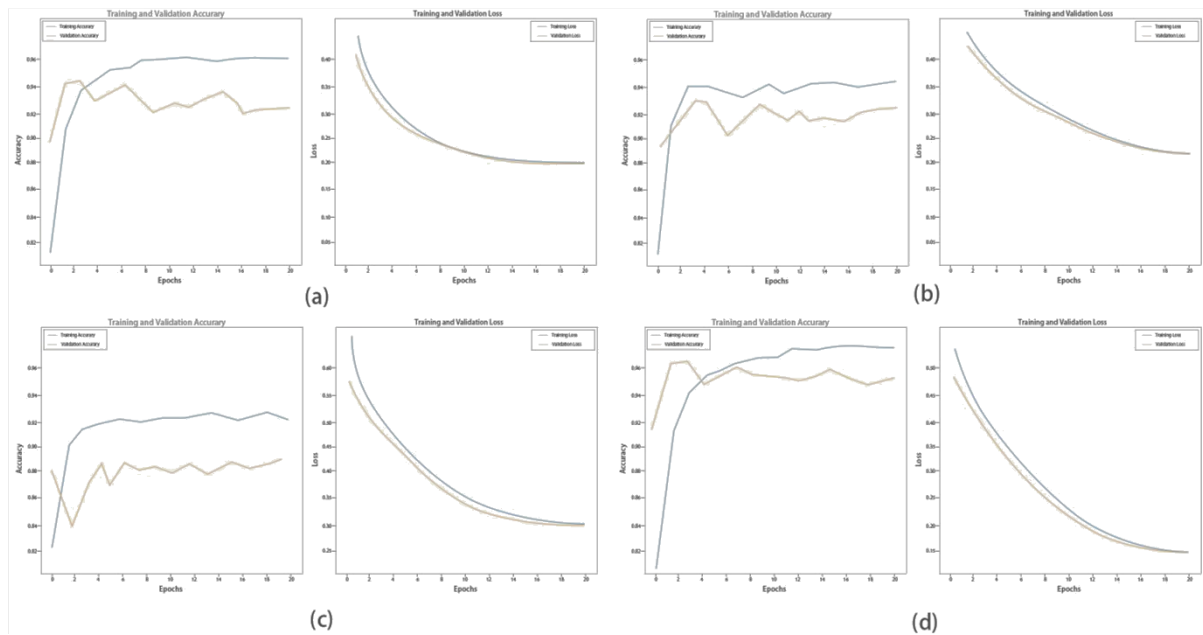


Fig.7 – Training and Validation Accuracy and Loss Plots
 (a) ResNet101; (b) VGG16; (c)Independent CNN; (d)CNN-LSTM-Attention

Table 3

Model Performance Evaluation Table

Model	Class Name	Precision	Recall	F1-score	Accuracy
ResNet101	Drowsiness	0.9315±0.004	0.9275±0.001	0.9293±0.001	0.9567±0.002
	Alertness	0.9371±0.002	0.9402±0.002	0.9384±0.003	
VGG16	Drowsiness	0.9182±0.001	0.9032±0.002	0.9102±0.001	0.9215±0.001
	Alertness	0.9251±0.002	0.9322±0.002	0.9274±0.003	
Independent-CNN	Drowsiness	0.8921±0.001	0.8812±0.002	0.8862±0.003	0.8978±0.001
	Alertness	0.9032±0.002	0.9002±0.001	0.9012±0.001	
Independent-LSTM	Drowsiness	0.8762±0.002	0.8672±0.003	0.8715±0.003	0.8824±0.002
	Alertness	0.8885±0.001	0.8901±0.002	0.8892±0.002	
CNN-LSTM-Attention	Drowsiness	0.9521±0.003	0.9491±0.002	0.9504±0.001	0.9567±0.002
	Alertness	0.9613±0.002	0.9642±0.001	0.9621±0.002	

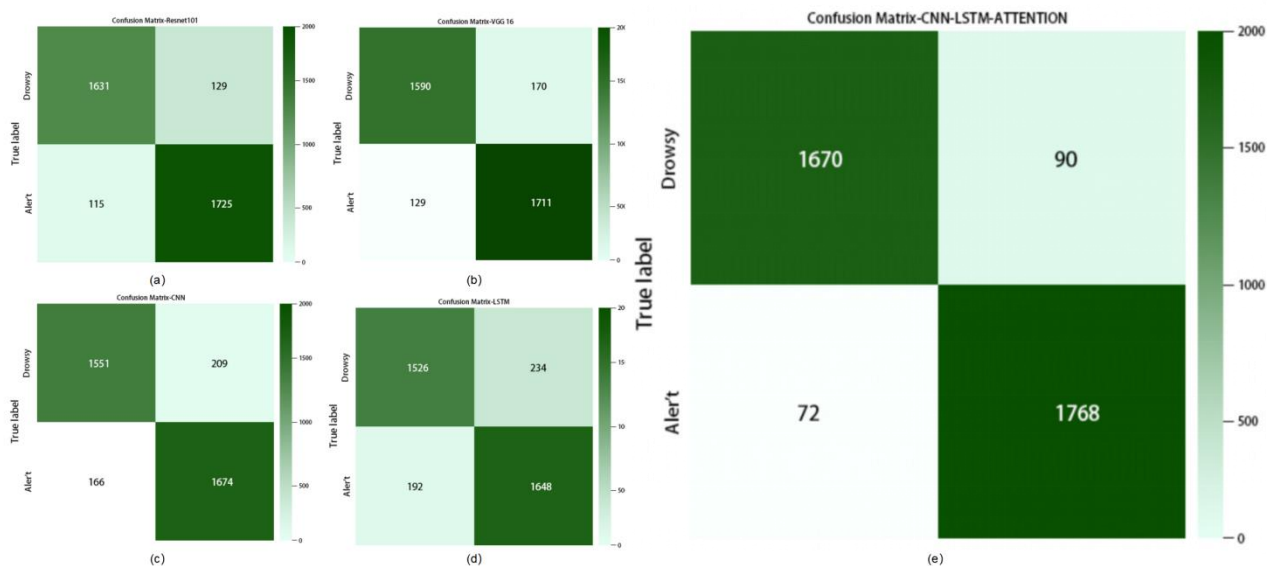


Fig. 8 - Confusion matrices
 (a)ResNet 101; (b)VGG 16; (c)Independent CNN; (d) Independent LSTM; (e)CNN-LSTM-Attention

Real-vehicle testing

To further validate the generalization capability and robustness of the proposed CNN-LSTM-Attention model in real-world operational environments, field experiments with actual machinery were conducted in June 2025 during the winter wheat harvest season at the winter wheat experimental field of Shandong University of Technology in Zibo City, Shandong Province, China. The experiment utilized a Lovol GR80 combine harvester with a camera mounted on the driver's side front. Approval was obtained from local agricultural machinery safety regulatory authorities, and all drivers signed informed consent forms. Figure 9 shows the front view (a) and side view (b) of the actual vehicle experiment process. Eight combine drivers (aged 29–52, including 3 females) were recruited to complete real wheat harvesting tasks. To validate the model's robustness across diverse conditions, test scenarios included: midday intense sunlight (80,000 lux), overcast conditions (<20,000 lux), drivers wearing masks or polarized sunglasses, and wearing hats. Each scenario was tested for 2 hours, with driver drowsiness changes recorded using a dual-expert scoring method throughout the experiment. Eight drivers conducted ten experimental sessions. Due to high operational intensity and prolonged continuous work, both expert assessments and model detection indicated fatigue in every session. To prevent accidents, a driving assistant was present during each experiment. Figure 10 displays the model's UI interface during drowsiness detection. The system ultimately integrates with the harvester's CAN bus to alert drivers via a drowsiness-alert mechanism, prompting timely rest breaks. Given drowsiness's cumulative nature, each experimental image set featured only one drowsiness baseline. Real-vehicle validation results are presented in Table 4.



Fig.9 – Vehicle test
(a)Front View; (b)Side View

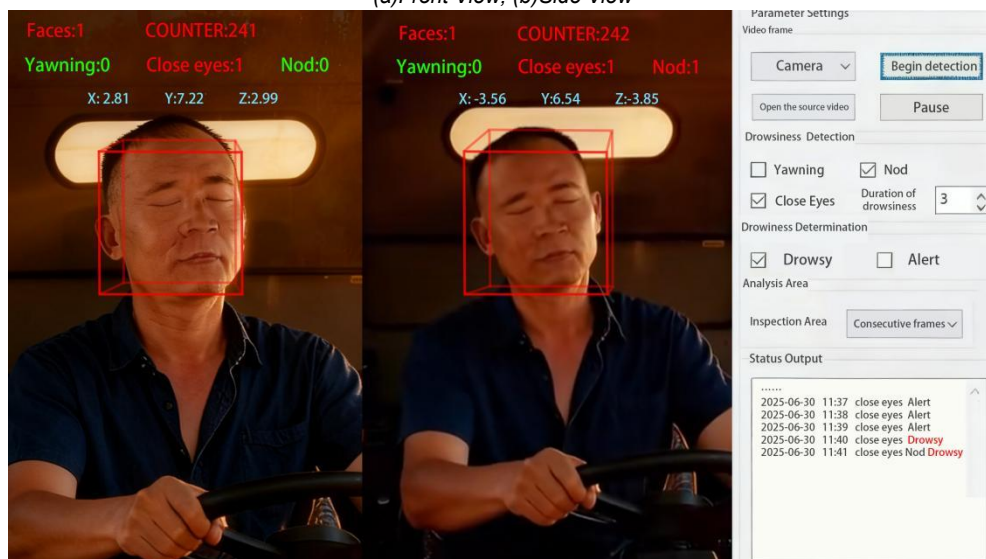


Fig.10 – UI Page Display Diagram

Table 4

Performance of Embedded Drowsiness Detection Model in Field Real-Vehicle Tests			
Test Scenario	Accuracy (%)	False Alarm Rate (%)	Miss Rate (%)
Noon Strong Light	93.81	3.92	2.31
Cloudy Day	94.62	3.10	2.28
Wearing a Mask	92.54	4.75	2.71

Test Scenario	Accuracy (%)	False Alarm Rate (%)	Miss Rate (%)
Wearing Sunglasses	94.35	3.30	2.35
Wearing a Hat	93.64	3.68	2.68
Wearing Mask + Hat	90.52	6.12	3.36
Wearing Sunglasses + Hat	90.33	6.25	3.42

CONCLUSIONS

This study addresses the challenges of complex operating environments for combine harvesters and the difficulty of detecting driver drowsiness. It proposes a lightweight CNN-LSTM-Attention model based on multi-feature fusion to achieve non-invasive driver drowsiness detection. This model innovatively integrates the static spatial feature extraction capability of lightweight convolutional neural networks (CNN), the temporal feature capture capability of long short-term memory networks (LSTM), and the attention mechanism's focused enhancement of key fatigue features. It effectively resolves the critical challenge of balancing accuracy and efficiency in traditional detection models.

Performance comparisons with four alternative models demonstrate significant advantages in detecting harvester operator fatigue. The proposed method achieves a validation accuracy of 95.67%, surpassing all other models, while requiring only 328,000 trainable parameters—far fewer than other high-accuracy models—thus achieving superior computational efficiency. Furthermore, real-vehicle validation confirms the model's robustness. Across diverse complex scenarios, the model maintains detection accuracy between 90.33% and 94.62%, with a false alarm rate below 6.25% and a false negative rate below 3.42%. By linking with the harvester's dashboard via CAN bus, it provides fatigue alert feedback, effectively mitigating operational safety risks.

This study provides a technically sound and practical solution for fatigue detection in agricultural machinery operators, though optimization opportunities remain. Future research will focus on developing models tailored for harvester terminals, aiming to achieve lighter model structures and faster inference speeds. Additionally, the classification and processing workflow for fatigue detection results will be further optimized. By designing a “tiered” early warning mechanism customized for agricultural machinery operation scenarios, an architecture will be established where different levels of drowsiness trigger distinct alert methods.

REFERENCES

- [1] Alameen, S.A., & Alhothali, A.M. (2023). A lightweight driver drowsiness detection system using 3DCNN with LSTM. *Computer Systems Science and Engineering*, Vol. 44, Saudi Arabia
- [2] Albadawi, Y., AlRedhaei, A., Takruri, M. (2023). Real-Time Machine Learning-Based Driver Drowsiness Detection Using Visual Features. *Journal of Imaging*, Vol. 9(5), Peru&Brazil
- [3] Albadawi, Y., Takruri, M., Awad, M. (2022). A Review of Recent Developments in Driver Drowsiness Detection Systems. *Sensors*, Vol. 22, Jordan
- [4] Atila, Ü., Uçar, M., Akyol, K., & Uçar, E. (2021). Plant leaf disease classification using EfficientNet deep learning model. *Ecological Informatics*, Vol. 61, Turkey
- [5] Chen, Y.J., Zhao, Y.C., Kong, D.G., & Others. (2016). Test analysis of heart rate and drowsiness of agricultural machinery drivers in field operations (田间作业中农机驾驶员心率与疲劳的测试分析). *Journal of Agricultural Mechanization Research*, Vol. 38, China
- [6] Dias, J., Silva, M., & Costa, P. (2024). Agricultural mechanization trends: Focus on combine harvester safety. *Journal of Agricultural Safety and Health*, Vol. 30, 45-62.
- [7] Di, M., Peng, Z., & Chen, S. (2025). Performance comparison and application of deep-learning-based image recognition models (基于深度学习的图像识别模型性能比较及应用). *Second International Conference on Big Data, Computational Intelligence, and Applications (BDCIA 2024) Proceedings, SPIE*, 2025, 13550, China
- [8] Indri J., Yulian J., Elly R. (2018). The Impact of Using Combine Harvester Technology on Social Economic Conditions of Swamp Rice Farmers and Harvest Workers in South Sumatera. *Jurnal Manajemen & Agribisnis*, Vol. 15 No. 3, pp. 299-308, Indonesia
- [9] Dipti, D. (2025). Driver Drowsiness Detection Using Hybrid Model CNN and LSTM. *International Journal of Scientific Research in Engineering and Management*, Vol. 9, India
- [10] Gautam, T. (2025). Novel CNN Based Model Using LSTM for Driver Drowsiness Detection. *International Journal for Research in Applied Science and Engineering Technology*, Vol. 13, India

- [11] Guan, Z. (2024). Research Progress on the Development of Drowsiness Driving Detection Based on Deep Learning (基于深度学习的疲劳驾驶检测发展研究进展). *Theoretical and Natural Science*, Vol. 52, China
- [12] Joshi, P., Adhikari, M., Shrestha, S., & Shaik, S. (2025). Real-Time Driver Drowsiness Detection Using CNN, MediaPipe, and ML Classifiers. *SoutheastCon 2025 Proceedings*, IEEE, USA
- [13] Koay, H. V., Chuah, J. H., Chow, C.-O., & Chang, Y.-L. (2022). Detecting and recognizing driver distraction through various data modality using machine learning: A review, recent advances, simplified framework and open challenges (2014–2021). *Engineering Applications of Artificial Intelligence*, Vol. 115, Malaysia
- [14] Kong, D. G., Zhang, S., Zhu, Z. Y., & Others. (2008). Analysis and evaluation of driver drowsiness in mechanized seeding operations (机械化播种作业中驾驶员疲劳分析与评价). *Transactions of the Chinese Society of Agricultural Machinery*, Vol. 39, China
- [15] Mahmoud, A., & Ahmed, S. (2024). Deep Learning for Computer Vision: Innovations in Image Recognition and Processing Techniques. *CyberSystem Journal*, Vol. 1, Egypt
- [16] Muthuraja, M., Shanthi, N., Aravindhraj, N., Sarankanth, K., Senthana Vignas, M., & Shivaani, J. (2024). An Acceptable Deep Learning Approach for Detecting Drowsiness. *Proceedings of the 9th International Conference on Communication and Electronics Systems, ICCES 2024*, India
- [17] Ren, B., Guan, W., Zhou, Q., & Wang, Z. (2023). EEG-Based Driver Drowsiness Monitoring within a Human–Ship–Environment System: Implications for Ship Braking Safety (人-船-环境系统下基于脑电信号的驾驶员疲劳监测及其对船舶制动安全的意义). *Sensors*, Vol. 23, China
- [18] Sahu, A., Pradhan, S., & Panda, R. (2021). A novel dataset and lightweight model for driver drowsiness detection in agricultural vehicles. *Computers and Electronics in Agriculture*, Vol. 187, 106345.
- [19] Sengupta, S. (2024). Enhanced Driver Drowsiness Detection System using Deep Learning and Image Processing Techniques. *International Journal of Scientific Research in Engineering and Management*, Vol. 8, India
- [20] Shang, Y., Yang, M., Cui, J., Cui, L., Huang, Z., & Li, X. (2022). Driver Emotion and Drowsiness State Detection Based on Time Series Fusion (基于时间序列融合的驾驶员情绪与疲劳状态检测). *Electronics*, Vol. 12, China
- [21] Sikander, Gulbadan, & Shahzad Anwar. (2019). Driver Drowsiness Detection Systems: A Review. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 20, Pakistan
- [22] Sun, Q. Y., & Zhang, Z. D. (2021). Crowd density estimation based on YOLOv3 enhanced model fusion (基于YOLOv3改进模型融合的人群密度估计). *Application of Computer Systems*, Vol. 4, China
- [23] Sun, Z., Miao, Y., Jeon, J. Y., Kong, Y., & Park, G. (2023). Facial feature fusion convolutional neural network for driver drowsiness detection. *Engineering Applications of Artificial Intelligence*, Vol. 126, South Korea
- [24] Tan, Y., Wu, L., Zhang, Z., Zhang, X., & Zhou, J. (2024). VGG16 hardware design and implementation for CNN in image recognition (用于CNN图像识别的VGG16硬件设计与实现). *Second International Conference on Informatics, Networking, and Computing (ICINC 2023) Proceedings*, SPIE, China
- [25] Yuan, B., Han, L., Gu, X., & Yan, H. (2020). Multi-deep features fusion for high-resolution remote sensing image scene classification (高分辨率遥感图像场景分类的多深度特征融合方法). *Neural Computing and Applications*, Vol. 33, China
- [26] Zaman, A., Chatterjee, P., & Sharma, R. (2024). Real-Time Drivers' Drowsiness Detection: A Deep Learning Approach. *2024 2nd International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings 2024) Proceedings*, India
- [27] Zhang, T., Chen, Z., & Ouyang, C. (2018). Research on driver drowsiness detection (驾驶员疲劳检测研究). *MIPPR 2017: Pattern Recognition and Computer Vision Proceedings*, SPIE, China
- [28] Zhou, N. (2024). Image recognition in depth: Comparative study of CNN and pre-trained VGG16 architecture for classification tasks (图像识别深度研究: CNN与预训练VGG16架构的分类任务对比). *Second International Conference on Physics, Photonics, and Optical Engineering (ICPPOE 2023) Proceedings*, SPIE, China
- [29] Zhu, R., Wang, J., Tang, H., Zhou, W., Pan, Z., Wang, Q., & Duo, T. (2016). Analysis and evaluation of combine harvester driver drowsiness based on heart rate variability (基于心率变异性的联合收割机驾驶员疲劳分析与评价). *Transactions of the Chinese Society of Agricultural Engineering*, Vol. 32, China