CTU Journal of Innovation and Sustainable Development
ISSN 2588-1418 | e-ISSN 2815-6412

DOI:10.22144/ctujoisd.2026.010

A robust ensemble framework for helmet usage classification in real-world scenarios

Tan-Duy Lam, and Tuong Le*

Faculty of Information Technology, HUTECH University, Viet Nam

**Corresponding author (lc.tuong@hutech.edu.vn)*

Article info.

Received 16 Feb 2025
Revised 11 Mar 2025
Accepted 04 Aug 2025

Keywords

Deep learning, ensemble learning, helmet usage classification, image classification

ABSTRACT

The application of machine learning models in the analysis of helmet-related images has yielded remarkable results in identifying and classifying helmet-wearing behaviours. Previous research has employed several pretrained models to predict proper or improper helmet use, achieving high accuracy on the Helmet Wearing Image Dataset (2024), a newly introduced dataset designed to enhance classification capabilities. This study aims to improve prediction performance on helmet datasets by leveraging state-of-the-art deep learning models and ensemble techniques. Using ResNet-50, MobileNetV2, and EfficientNet-B0 models, the proposed EnsemHelmet Framework uses a soft voting ensemble to optimise the classification results, achieving an outstanding accuracy of 99.24% on the experimental dataset. The results demonstrate the potential of ensemble learning to achieve high performance. This study not only improves the accuracy of the helmet-wearing recognition system but also highlights the effectiveness of ensemble techniques in optimizing performance on real-world datasets.

1. INTRODUCTION

Deep learning, a subset of machine learning, has attracted considerable attention in recent years due to its high predictive performance across many domains, especially computer vision and natural language processing (NLP). This approach relies on multi-layer artificial neural networks to model complex data representations, allowing automatic feature extraction and prediction. Advances in hardware, such as Graphics Processing Units (GPUs), and the availability of large datasets have dramatically improved the scalability and performance of deep learning models. Notable recent developments include the rise of transformer models such as BERT and GPT, which have revolutionised natural language understanding, and the widespread use of convolutional neural networks (CNNs) in image-related tasks, such as in CCTV (Closed-circuit television), and intelligent traffic management.

Deep Learning has made significant advances in practical applications in a wide range of fields. In healthcare, it enhances disease diagnosis by detecting abnormalities in medical images such as X-rays (Vo et al., 2022; Truong et al., 2024). In agriculture, it supports crop health monitoring (Nguyen et al., 2024; Zhang et al., 2024), disease detection, and resource utilization optimization through satellite and drone imagery analysis. In waste management, deep learning-powered vision systems automate smart waste sorting (Vo et al., 2019; Masand et al., 2021), improving recycling efficiency. Additionally, in CCTV surveillance, deep learning enables real-time anomaly detection (Tran et al., 2024), facial recognition (Irfan et al., 2024), and threat assessment, strengthening security and public safety. These advances, combined with hybrid architectures that integrate reinforcement learning and transfer learning, will be the foundation for the development of cutting-edge applications.

Deep learning is transforming smart transportation, enabling more advanced traffic management, enhanced safety measures, and greater overall efficiency. Pretrained models such as ResNet-50 (He et al. 2016), MobileNetV2 (Sandler et al., 2018), and EfficientNet-B0 (Tan & Le, 2019) play an important role in vehicle detection, traffic monitoring, and accident analysis. ResNet-50 excels in vehicle classification (Singh et al., 2024) and pedestrian detection, MobileNetV2 and EfficientNet-B0 optimize accuracy and efficiency for high-precision image analysis (Shourie, 2024; Ren & Cong, 2022). Additionally, CNNs and their enhanced models enhance object detection and anomaly recognition (Adewopo et al., 2023). Together, these deep learning models are driving advances in autonomous driving and smart city infrastructure, making transportation safer and more efficient. As deep learning continues to evolve, integrating these models with IoT and 5G networks will further enhance system responsiveness. This progress paves the way for fully automated traffic management and smart urban mobility solutions.

The Helmet Wearing Image Dataset (Patil et al., 2024) comprises 28,736 images featuring various helmet types (Full-Face, Half-Face, Modular, and Off-Road) in both correct and incorrect wearing configurations. Images were captured under various lighting conditions using iPhone 13 and Mi 10 T mobile phones, and pre-processed to a standard resolution of 768 × 576. This well-balanced dataset is specifically designed for machine learning tasks such as image classification and object detection, with a focus on improving motorcycle helmet safety. Notably, Patil et al. (2024) utilised a pretrained deep learning model that achieved an impressive accuracy of 98.61% for helmet usage classification on this dataset. Specifically, MobileNetV2 attained 97.02% accuracy, ResNet-50 reached 98.61%, and VGG19 achieved 91.03%. However, there is still potential to further improve accuracy by using more advanced deep learning models. Additionally, using ensemble methods can further improve prediction capabilities, enhance performance, and enable the development of more robust helmet recognition technologies.

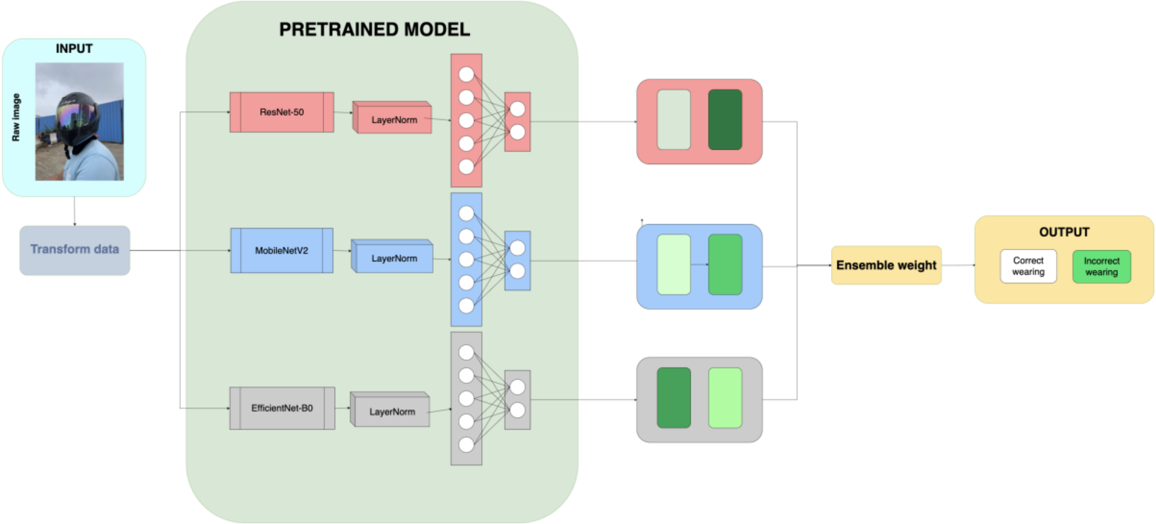


Figure 1. Overview of the EnsemHelmet Framework Architecture

This study presents the EnsemHelmet framework, an advanced approach to helmet usage classification that leverages deep learning models and ensemble techniques. Unlike previous research (Patil et al., 2024), which relied solely on ResNet-50, MobileNetV2, and VGG19, our framework integrates ResNet-50, MobileNetV2, and EfficientNet-B0, employing soft voting ensembles to enhance classification accuracy and robustness. The proposed framework achieves 99.24% accuracy on the experimental dataset, surpassing previous benchmarks. These findings reinforce the effectiveness of ensemble learning in improving classification accuracy and optimizing deep learning models for real-world applications. Furthermore, this study highlights the potential of combining multiple models to enhance prediction performance in traffic safety and other traffic image recognition tasks.

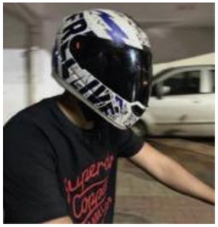
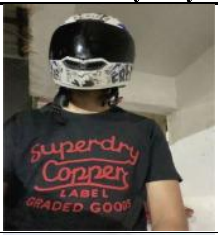
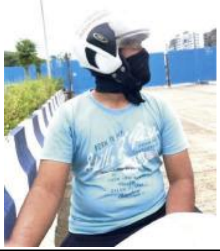
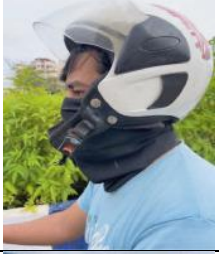
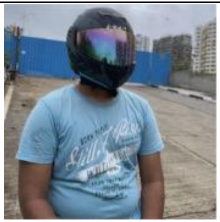
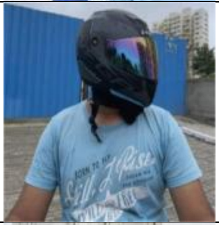
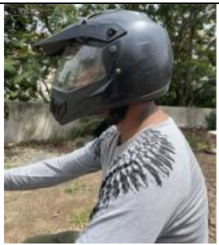
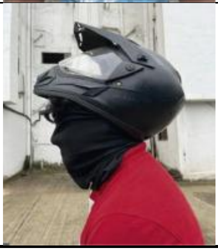
2. MATERIALS AND METHOD

The EnsemHelmet framework leverages advanced deep learning models, including ResNet-50, MobileNetV2, and EfficientNet-B0, through an ensemble mechanism to improve performance of helmet usage classification. Figure 1 provides an overview of the EnsemHelmet framework architecture.

2.1. Helmet-wearing image dataset

The dataset (Patil et al., 2024) is a comprehensive collection of 28,736 high-resolution images designed to support the development of advanced image recognition techniques to improve motorcycle helmet safety. The dataset includes diverse images of various helmet types including full-face, half-face, modular, and off-road, worn both correctly and incorrectly. These images were taken under a range of lighting conditions, including day and night, and feature diverse environmental contexts and helmet orientations. With its pre-processed structure, the dataset is ideal for machine learning tasks such as image classification, object detection, and helmet pose estimation. By focusing on promoting proper helmet use, the dataset is a valuable resource for safety research and technological advancements in motorcycle helmet recognition.

Table 1. Sample images from the helmet-wearing image dataset

Type of Helmet	Correctly way	Incorrectly way
Full-Face Helmet		
Half-Face Helmet		
Modular Helmet		
Off-Road Helmet		

This study focuses on detecting proper helmet use to enhance safety standards through automated recognition. Therefore, the dataset is divided into two classes: correct and incorrect helmet usage. In addition, the dataset was initially partitioned into a training set (70%) and a testing set (30%). However, to facilitate model tuning, 20% of the data was allocated for validation. As a result, the final dataset split consists of 50% for training, 20% for validation, and 30% for testing. The distribution of correctly and incorrectly labelled images across the subsets of the Helmet Wearing Image Dataset is shown in Table 2.

Table 2. Distribution of correctly and incorrectly labelled images across subsets of the helmet-wearing image dataset

Subset	Correctly way	Incorrectly way
Training set	6776	7025
Validation set	2904	3011
Test set	4500	4500

2.2. ResNet-50

ResNet-50. This model, introduced by He et al. (2015) at Microsoft Research, is a 50-layer deep convolutional neural network designed to mitigate the vanishing gradient problem through residual connections, enabling more efficient training of very deep networks. This allows for efficient training of very deep networks, making it extremely effective for tasks such as image classification and object detection. The key features of the ResNet-50 architecture are as follows:

• ***Residual blocks*****:** Each residual block allows the network to learn identity mappings, facilitating the flow of gradients, as shown in Equation (1):

$$y = F(x, \{W_i\}) + x \qquad (1)$$

where x is the input to the block; $F(x, \{Wi\})$ is the residual function applied to the input x; and $\{W_i\}$ represents the learnable weights of the block.

• ***Bottleneck design:*** The bottleneck structure involves a sequence of 1×1, 3×3, and 1×1 convolutions to optimise the computational efficiency in Equation (2) as follows:

$$y = Conv(1 \times 1, 3 \times 3, 1 \times 1)(x) \qquad (2)$$

2.3. MobileNetV2

This model, introduced by Sandler et al. (2018), is a lightweight convolutional neural network designed for efficient performance on mobile and embedded devices. It builds upon MobileNetV1 by incorporating Inverted Residual Blocks and Linear Bottleneck layers, which enhance computational efficiency while preserving representational power. The key features of the MobileNetV2 architecture are as follows.

• **Depthwise Separable Convolution:** This model employs depthwise separable convolutions to reduce computational complexity. The number of floating-point operations (FLOPs) is significantly reduced compared to standard convolutions:

$$FLOPs \approx \frac{1}{D_k^2} \times FLOPs\ of\ standard\ convolution \qquad (3)$$

In Equation (3), D_k denotes the kernel size.

• **Linear Bottleneck Transformation.** To mitigate information loss caused by ReLU activation in low-dimensional spaces, MobileNetV2 utilises a linear bottleneck, transforming feature representations (Equation 4) as follows:

$$X_{out} = W\ .ReLU6(X_{exp}) \qquad (4)$$

where W represents the transformation matrix, and X_{exp} is the expanded feature space.

• **Inverted Residual Block.** The inverted residual structure facilitates efficient gradient flow by leveraging shortcut connections when input and output dimensions are identical by Equation (5):

$$X_{out} = X_{input} + f(X_{input}) \qquad (5)$$

where $f(.)$ represents the transformation function applied within the residual block.

2.4. EfficientNet-B0

This model, introduced by Tan and Le (2019), is a convolutional neural network designed for optimal accuracy and efficiency by leveraging compound scaling. Unlike traditional models that scale depth, width, and resolution independently, EfficientNet-B0 employs a principled approach that scales all three dimensions simultaneously, leading to superior performance with fewer parameters. The key features of this model architecture are as follows.

• **Compound Scaling.** This model uses a compound coefficient ϕ to scale depth (d), width (w), and resolution (r) in a balanced manner:

$$d = \alpha^{\phi} d_0, w = \beta^{\phi} w_0, r = \gamma^{\phi} r_0 \qquad (6)$$

In Equation 6, α, β, γ are constants determined through grid search, and ϕ controls the overall scaling factor.

- **Mobile Inverted Bottleneck Convolutions (MBConv).** EfficientNet-B0 integrates MBConv blocks, inspired by MobileNetV2, to enhance efficiency. The transformation function (Equation 7) is given by:

$$X_{out} = X_{input} + f(X_{input}) \qquad (7)$$

where $f(.)$ includes depthwise separable convolutions and squeeze-and-excitation (SE) modules.

- **Squeeze-and-Excitation (SE) Modules. SE modules improve channel-wise feature recalibration by adaptively weighting channels by** Equation (8) as follows**:**

$$s = \sigma(W_2\delta(W_1 z)) \qquad (8)$$

where z is the global average pooled feature, W_1, W_2 are fully connected layers, δ is a ReLU activation, and σ is a sigmoid function.

2.5. Ensemble learning

To improve classification performance, an ensemble learning strategy was implemented that combined predictions from the ResNet-50, MobileNetV2, and EfficientNet-B0 models. This approach exploits the complementary strengths of three models to enhance overall accuracy, as shown in Equation (9).

$$P_e = \frac{1}{N}\sum_{i=1}^{N} P_{model_i} \qquad (9)$$

where P_e is the final ensemble prediction probability, $P_{\text{model\_}i}$ is the prediction probability from each model (e.g., ResNet-50, MobileNetV2, and EfficientNet-B0) and N is the number of models in the ensemble (in this case, $N = 3$). This approach reduces the impact of individual model biases and errors, strengthening the system's overall reliability. Additionally, by averaging predictions across cross-

Table 3. Training configuration for ResNet-50, MobileNetV2 and EfficientNet_B0 models.

No	Hyper-parameter	Value	Description
1	Batch size	32	Number of samples per training batch, balancing performance and memory.
2	Epoch	20	The number of iterations of the entire training data helps the model learning better.
3	Learning rate	0.0001	The rate at which weights are updated during training affects convergence.
4	Weight decay	0.0001	Adjustment coefficient to reduce overfitting
5	Learning Rate Scheduler	ExponentialLR (gamma=0.9)	Gradually decreases the learning rate over epochs to improve model convergence.

3. RESULTS AND DISCUSSION

3.1. Experimental setup

In this study, we conduct experiments on the Helmet Wearing Image Dataset to evaluate and compare the effectiveness of the proposed framework (EnsemHelmet) against three individual state-of-the-art pretrained CNNs models including ResNet-50, EfficientNet_B0, and MobileNetV2, as well as a VGG19 model implemented in Patil et al. (2024). The hyper-parameters of the ResNet-50, MobileNetV2, and EfficientNet_B0 models in EnsemHelmet are presented in Table 3. Additionally, the EnsemHelmet model is also experimented with default hyper-parameters, referred to as EnsemHelmet (Default). The training process in PyTorch includes batch size, learning rate, weight decay, Cross-Entropy loss, Adam optimizer, and ExponentialLR scheduler for gradual learning rate reduction per epoch as shown in Table 3. The performance of the experimental models is evaluated on the test set using accuracy and confusion matrix, which provides a detailed analysis of the classification ability between two classes.

Before being fed into the model for training and prediction, all images were preprocessed by resizing them to 224 × 224 pixels and normalizing them using the ImageNet mean and standard deviation (mean = [0.485, 0.456, 0.406] and std = [0.229, 0.224, 0.225]). To enhance generalization, we apply data augmentation techniques such as horizontal flipping, random image rotation (up to 15°), brightness and contrast adjustment, and random cropping to the training set. Meanwhile, the images in the validation and test sets only undergo resizing and normalization to maintain consistency during evaluation.

3.2. Results

Fig. 2 shows that the loss decreases over epochs, with the training (blue) and validation (orange) losses following a similar downward trend, indicating effective learning and minimal overfitting. Minor fluctuations in validation loss are expected due to dataset variations. Fig. 3 depicts the accuracy improving steadily, with training and validation curves converging near 100%, suggesting strong generalization. Occasional validation accuracy surpassing training accuracy may result from regularization techniques, reinforcing the model's robustness. Based on the stable convergence of loss and accuracy, we selected 20 epochs for training the EnsemHelmet framework.

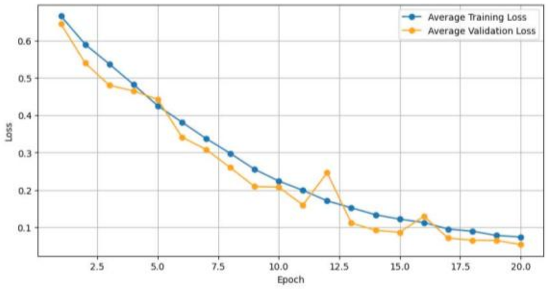


Figure 2. The training and validation loss of the EnsemHelmet framework

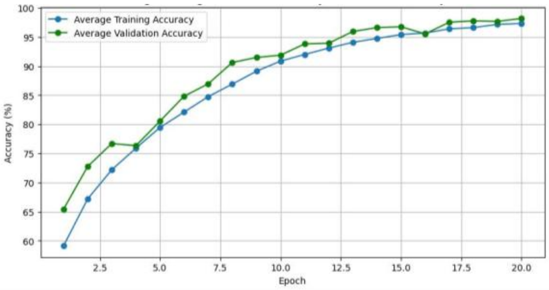
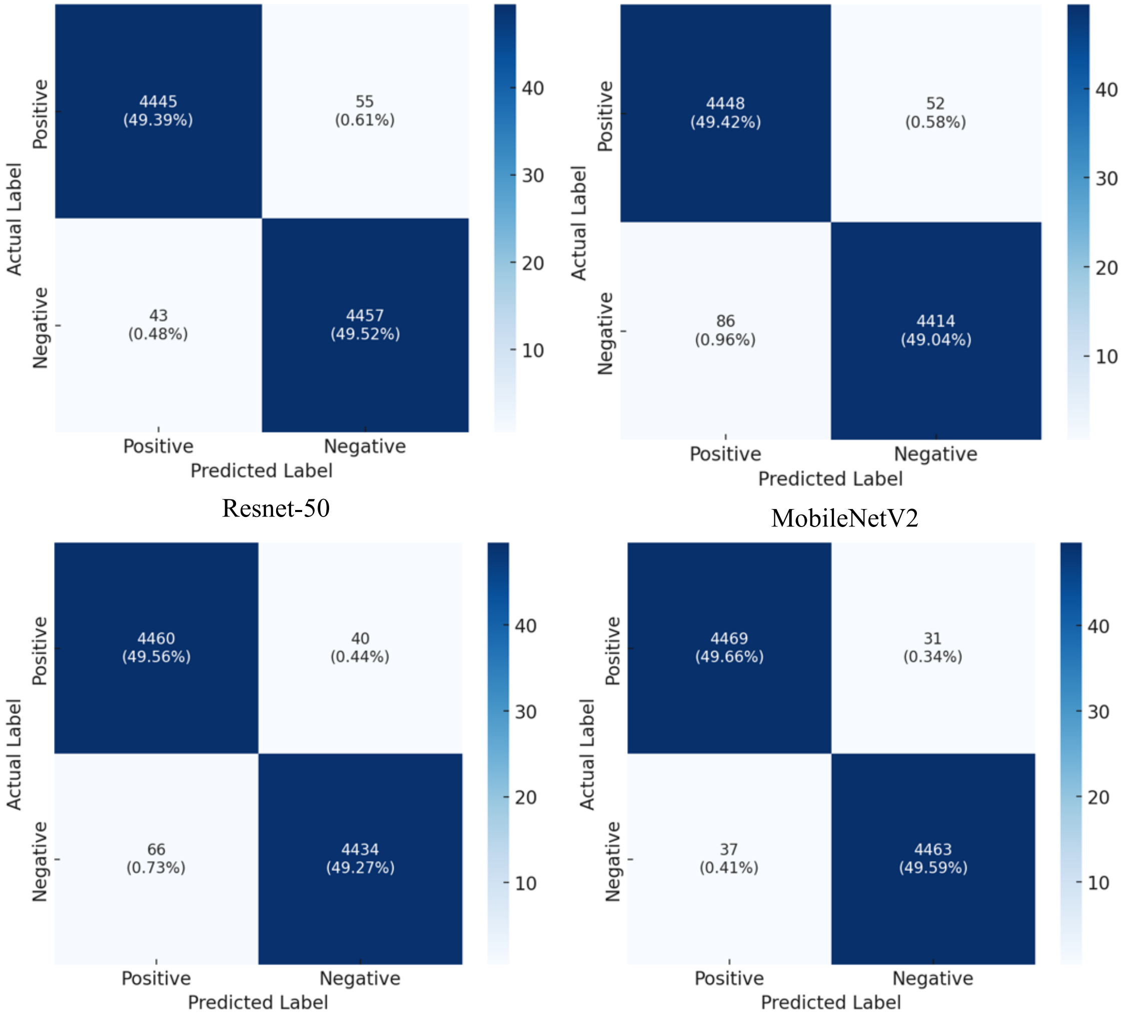


Figure 3. The training and validation accuracy of the EnsemHelmet framework

The confusion matrix results (Figure 4) demonstrate the effectiveness of various deep learning models in **helmet usage classification**. Among the individual models, EfficientNet-B0 shows superior performance compared to ResNet-50 and MobileNetV2, achieving fewer misclassifications (False Positive Rate: 40 (0.44%), False Negative Rate: 66 (0.73%)). However, the highest classification accuracy was achieved by the EnsemHelmet framework, which integrates multiple models, reducing both false positives (31 – 0.34%) and false negatives (37 – 0.41%). This combined strategy effectively minimises misclassification errors by leveraging the complementary strengths of each architecture.

A notable observation is that MobileNetV2 recorded the highest number of false negatives (86 – 0.96%), indicating a greater tendency to misclassify helmet-wearing individuals as non-helmeted. This can cause significant problems in practical safety applications, such as automatic traffic monitoring. In contrast, the EnsemHelmet framework significantly reduced both false positives and false negatives, resulting in a more reliable and robust classifier. These findings highlight the advantages of ensemble learning for improving helmet use detection, making it a promising approach for real-world applications, including law enforcement and road safety monitoring.

Table 4 summarizes the experimental results, comparing the individual models, VGG model, and the EnsemHelmet framework. Among the individual models, ResNet-50 achieved the highest accuracy at 98.91%, followed closely by EfficientNet-B0 (98.82%) and MobileNetV2 (98.46%). While VGG19 performed well, it achieved lower accuracy of 91.03%, suggesting limitations in feature extraction for this task. The EnsemHelmet (Default) framework also demonstrated strong performance, achieving 97.80% accuracy, highlighting the benefits of ensemble learning. Notably, the proposed EnsemHelmet framework outperformed all individual models, achieving 99.24% accuracy, demonstrating the effectiveness of ensemble learning in combining multiple architectures to enhance classification performance significantly.



Resnet-50

MobileNetV2

EfficientNet-B0

EnsemHelmet Framework

Figure 4. Confusion Matrices for ResNet-50, MobileNetV2, EfficientNet-B0, and the EnsemHelmet Framework

Table 4. Experimental results of individual models and the ensemhelmet framework

No	Model	Accuracy (%)
1	VGG19 (Patil et al., 2024)	91.03
2	MobileNetV2	98.46
3	Resnet 50	98.91
4	EfficientNet-B0	98.82
5	EnsemHelmet (Default)	97.80
6	EnsemHelmet (Proposed)	99.24

4. CONCLUSIONS

This study introduces the EnsemHelmet framework, an ensemble-based approach for helmet usage classification, integrating ResNet-50, MobileNetV2, and EfficientNet-B0 to enhance predictive accuracy. Experimental results indicate that ensemble learning significantly improves classification performance, achieving 99.24% accuracy, surpassing individual deep learning models. The confusion matrix analysis highlights its effectiveness in minimizing false positives and false negatives, ensuring a robust detection system. Notably, while MobileNetV2 exhibited a higher false-negative rate, the EnsemHelmet framework effectively mitigated this issue, reinforcing its practical reliability. These findings highlight the potential of fusion techniques in safety surveillance systems, contributing to enhanced road safety and citizen law enforcement.

Future research could explore incorporating additional deep learning architectures and real-time implementations to further optimise detection performance. Additionally, applying ensemble learning techniques to broader safety compliance tasks, such as detecting protective gear in industrial settings, could extend the impact of this work beyond helmet use classification.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

REFERENCES

Adewopo, V. A., Elsayed, N., ElSayed, Z., Ozer, M., Abdelgawad, A., & Bayoumi, M. (2023). A review on action recognition for accident detection in smart city transportation systems. *Journal of Electrical Systems and Information Technology*, *10*(1), 57. https://doi.org/10.1186/s43067-023-00124-y

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). IEEE. https://doi.org/10.1109/CVPR.2016.90

Irfan, E., Jacob, C., & Resmi, R. (2024, May). Facial Recognition and CCTV Integration for Enhanced Security Using Deep Learning Techniques. In *2024 IEEE Recent Advances in Intelligent Computational Systems (RAICS)* (pp. 1-5). IEEE. https://doi.org/10.1109/RAICS61201.2024.10689986

Masand, A., Chauhan, S., Jangid, M., Kumar, R., & Roy, S. (2021). Scrapnet: An efficient approach to trash classification. *IEEE Access*, *9*, 130947–130958. https://doi.org/10.1109/ACCESS.2021.3111230

Nguyen, V. H., Bui, H. H. N., & Le, T. P. (2024, November). Assessing grain size variation across rice panicles using YOLOv8 and DeepLabv3 models. In Thai-Nghe, N., Do, T.N., & Benferhat, S. (Eds.), *Intelligent Systems and Data Science. ISDS 2024* (pp. 15–29). Springer. https://doi.org/10.1007/978-981-97-9616-8_2

Patil, K., Jadhav, R., Suryawanshi, Y., Chumchu, P., Khare, G., & Shinde, T. (2024). HelmetML: A dataset of helmet images for machine learning applications. *Data in Brief*, *56*, 110790. https://doi.org/10.1016/j.dib.2024.110790

Ren, Y., Cong, L. (2023). Traffic Image Classification Algorithm Based on Deep-Learning. In Atiquzzaman, M., Yen, N., & Xu, Z. (Eds.), *Proceedings of the 4th International Conference on Big Data Analytics for Cyber-Physical System in Smart City - Volume 1. BDCPS 2022* (pp. 437–445). Springer. https://doi.org/10.1007/978-981-99-0880-6_48

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4510–4520). IEEE. https://doi.org/10.1109/CVPR.2018.00474

Shourie, P. (2024). Optimizing traffic sign detection with MobileNetV2: A lightweight deep learning approach. In *Proceedings of 2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)* (pp. 272–276). IEEE. https://doi.org/10.1109/ICTACS62700.2024.10840818

Singh, K., Patil, N., Mohite, S. G., Jadhav, S., Mohite, S., Gayakwad, M., & Joshi, R. (2024). Vehicle identification using RESNET-50: CNN approach. In *Proceedings of 2024 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICBDS61829.2024.10837029

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, (pp. 6105–6114). PMLR.

Tran, N., Nguyen, H., Ly, D., & Nguyen, H. D. (2024, November). Violence detection using skeleton data with graph convolutional networks. In Thai-Nghe, N., Do, T.N., & Benferhat, S. (Eds.), *Intelligent Systems and Data Science. ISDS 2024* (pp. 86–97). Springer. https://doi.org/10.1007/978-981-97-9616-8_7

Truong, T. D., Huynh, P. H., Nguyen, V. H., & Do, T. N. (2024). Enhancing the efficiency of lung disease classification based on multi-modal fusion model. In Thai-Nghe, N., Do, TN., & Benferhat, S. (Eds.), *Intelligent Systems and Data Science. ISDS 2024* (pp. 86–97). Springer. https://doi.org/10.1007/978-981-97-9616-8_5

Vo, A. H., Son, L. H., Vo, M. T., & Le, T. (2019). A novel framework for trash classification using deep transfer learning. *IEEE Access*, *7*, 178631–178639. https://doi.org/10.1109/ACCESS.2019.2959033

Vo, M. T., Vo, A. H., & Le, T. (2022). A robust framework for shoulder implant X-ray image classification. *Data Technologies and Applications*, *56*(3), 447–460. https://doi.org/10.1108/DTA-08-2021-0210

Zhang, W., Dang, L. M., Nguyen, L. Q., Alam, N., Bui, N. D., Park, H. Y., & Moon, H. (2024). Adapting the Segment Anything Model for plant recognition and automated phenotypic parameter measurement. *Horticulturae*, *10*(4), 398. https://doi.org/10.3390/horticulturae10040398