Majalah Bisnis & IPTEK
Vol. 19 No. 2, October 2026, 194 - 220
ISSN Online: 2502-1559. DOI: https://doi.org/10.55208/bistek
Published by P3M, STIE Pasundan, Bandung, Indonesia.

Optimization of the YOLOv8-Nano Deep Learning Model Through Knowledge Distillation for Driver Drowsiness Detection on Android Devices

Muhamad Fauzan[1], Feri Alpiyasin[2]
[1] *STMIK Mardira Indonesia, Bandung, Indonesia*
[2] *Telkom University, Indonesia*
Email: vanz66hu@gmail.com[1], feryalpiyasin@gmail.com[2]

Abstract

Driver drowsiness is a major concern in transportation safety and requires an automated detection system capable of operating efficiently on mobile devices. This study aims to optimize the deep learning model YOLOv8-Nano for driver drowsiness detection through Knowledge Distillation while maintaining its lightweight characteristics for Android deployment. The dataset consists of 3,090 images representing six classes, namely eye_closed, eye_open, yawn, no_yawn, head_dropped, and head_straight. A YOLOv8-Large model was trained as the teacher model for 100 epochs using the default random seed, while YOLOv8-Nano was evaluated through four experimental configurations: E0 as the baseline, E1 with logit distillation, E2 with logit and bounding box distillation, and E3 with logit, bounding box, and feature distillation. Each configuration was trained for 30 epochs using random seeds 42, 43, and 44, and the results were summarized using mean and standard deviation. The results show that E3 achieved the highest performance, with a Precision of 0.9260 ± 0.0066, Recall of 0.9018 ± 0.0050, F1-score of 0.9137 ± 0.0036, mAP@0.5 of 0.9323 ± 0.0015, and mAP@0.5:0.95 of 0.6368 ± 0.0030. Compared with the E0 baseline, E3 improved mAP@0.5 from 0.3845 ± 0.0360 to 0.9323 ± 0.0015. The optimized model maintains 3.2 million parameters, 8.7 GFLOPs, and a model size of 6.2 MB, with an inference time of 182–343 ms and an observed frame rate of 2.91–5.82 FPS on the target Android device. Functional testing using the Black Box method achieved a Pass result for all tested functions. These results indicate that Knowledge Distillation, particularly the combination of logit, bounding box, and feature distillation in E3, can improve the detection performance of YOLOv8-Nano while preserving its lightweight characteristics for Android deployment.

Keywords: *Deep Learning, YOLOv8-Nano, Knowledge Distillation, Driver Drowsiness Detection, Android, Object Detection*

Introduction

Driver drowsiness is an important issue in transportation safety because decreased alertness can impair a driver's attention and responsiveness, thereby increasing the risk of traffic accidents. The development of automated driver monitoring systems has therefore become an important research area, particularly through the application of computer vision and deep learning. Recent research indicates that deep learning has been widely applied to driver drowsiness detection using visual, physiological, and multimodal information, while practical implementation remains challenged by dataset diversity, evaluation consistency, and computational requirements (Fonseca & Ferreira, 2025) . The application of deep learning and convolutional neural networks to image-based detection has also demonstrated its potential to extract complex visual features and support automated recognition systems (Alpiyasin et al., 2025).

Computer vision-based driver drowsiness detection can identify visual characteristics associated with reduced driver alertness through observable conditions of the eyes, mouth, and head posture. In this research, the target conditions are represented by six detection classes, namely eye_closed and eye_open for eye conditions, yawn and no_yawn for mouth conditions, and head_dropped and head_straight for head posture. Object detection is particularly relevant because it enables the identification of these conditions together with their spatial locations through bounding boxes. YOLO-based models have been increasingly applied to real-time driver monitoring because of their ability to perform object detection with relatively efficient inference. Recent studies have demonstrated the application of YOLOv8 for driver fatigue and drowsiness detection, including systems designed to identify visual indicators of driver fatigue in real-time environments (Debsi et al., 2024; Hidayat & Ibnugraha, 2025) .

YOLOv8 provides several model variants with different computational capacities, enabling the model family to be adapted to different deployment requirements. YOLOv8 provides model variants with different computational capacities, allowing its application to be adapted to different deployment requirements (Jocher et al., 2023) . Lightweight object detection

is particularly relevant for edge and mobile environments because such systems operate under constraints related to computation and memory. Recent research on lightweight object detection has emphasized the need to balance detection performance with computational and memory efficiency, particularly for resource-constrained deployment environments (Cao et al., 2023; Khan & Rafat, 2025) . Therefore, improving the performance of YOLOv8-Nano while retaining its lightweight characteristics becomes an important consideration for driver drowsiness detection on Android devices.

Knowledge Distillation (KD) is an approach that enables a smaller student model to learn from a larger and more capable teacher model. The objective is to transfer useful knowledge from the teacher into the student so that the student can improve its performance without directly adopting the computational complexity of the teacher. Recent studies have shown that Knowledge Distillation can improve lightweight object detectors while supporting deployment under resource-constrained conditions (Cao et al., 2023; Khan & Rafat, 2025; Lan & Tian, 2024) . This characteristic makes Knowledge Distillation relevant to the optimization of YOLOv8-Nano, where detection performance needs to be improved while maintaining suitability for mobile deployment.

The knowledge transferred from the teacher model can take several forms and provide complementary guidance to the student model. Logit knowledge transfers information from the teacher's prediction outputs, allowing the student to learn the relationships contained in the teacher's prediction distribution. Bounding box knowledge transfers information related to object localization, guiding the student to produce bounding boxes that are more consistent with the teacher's predictions. Feature knowledge transfers intermediate feature representations learned by the teacher, providing additional visual information for the student's learning process. The use of these complementary knowledge components provides an opportunity to evaluate their individual and combined contributions to the optimization of YOLOv8-Nano. In this research, the three components

are applied progressively through the experimental configurations E1, E2, and E3 to analyze their respective effects on detection performance.

Previous studies have demonstrated the effectiveness of YOLOv8-based approaches for driver fatigue and drowsiness detection. Debsi et al. (2024) proposed an improved YOLOv8-based method for detecting driver distraction and fatigue, while Hidayat and Ibnugraha (2025) developed a real-time driver fatigue monitoring system using YOLOv8. Other recent studies have focused on improving YOLOv8-based fatigue detection through architectural modifications to increase detection accuracy and computational efficiency (Gu et al., 2025; Xie et al., 2025) . However, the studies discussed above primarily focus on modifying the detection architecture or developing driver monitoring systems, while the use of Knowledge Distillation to progressively transfer logit, bounding box, and feature knowledge into YOLOv8-Nano for driver drowsiness detection and Android deployment has received comparatively limited attention. The broader literature also identifies computational efficiency and real-world deployment as continuing challenges in deep learning-based driver drowsiness detection (Fonseca & Ferreira, 2025) .

Based on these considerations, this research proposes the optimization of YOLOv8-Nano through Knowledge Distillation for driver drowsiness detection on Android devices. YOLOv8-Nano is used as the student model, while a larger model is employed as the teacher. The optimization process is evaluated progressively through four experimental configurations: E0 as the YOLOv8-Nano baseline, E1 using logit distillation, E2 using logit and bounding box distillation, and E3 combining logit, bounding box, and feature distillation. Each configuration is evaluated across three independent runs using random seeds 42, 43, and 44 to assess the consistency of the observed performance. The study aims to analyze the contribution of each Knowledge Distillation component to detection performance and to determine an optimized YOLOv8-Nano configuration that provides an appropriate balance between detection performance and computational efficiency for Android deployment.

Research Method

This research employs three complementary methodological components: Research Methodology, Development Method, and Model Optimization Method. The Research Methodology uses the Design Science Research Methodology (DSRM) to organize the overall research process and the development and evaluation of the resulting artifact. The Development Method uses Agile to develop the Android application, while the Model Optimization Method applies Knowledge Distillation to improve the performance of the YOLOv8-Nano model. The integration of these components is intended to produce an optimized lightweight object detection model for driver drowsiness detection on Android devices.

The research methodology adopts the Design Science Research Methodology (DSRM) because the research produces an artifact in the form of an optimized YOLOv8-Nano model and an Android-based driver drowsiness detection application. Recent studies have applied DSRM to artifact development and evaluation in information technology research, particularly when the research objective is to produce and evaluate a solution to a defined problem (Haryanti et al., 2022, 2024) .

The DSRM implementation in this research consists of six stages: Problem Identification and Motivation, Objectives of a Solution, Design and Development, Demonstration, Evaluation, and Communication.

Problem Identification and Motivation

The research begins by identifying the need for an automated driver drowsiness detection system that can operate efficiently on mobile devices. The use of deep learning-based object detection provides an opportunity to identify visual indicators of driver drowsiness automatically. However, the computational limitations of mobile devices require a model that can provide an appropriate balance between detection performance and computational efficiency. Therefore, YOLOv8-Nano is selected as the student model and optimized using Knowledge Distillation.

Objectives of a Solution

The objective of the proposed solution is to improve the detection performance of YOLOv8-Nano while maintaining its suitability for

Android deployment. The optimization is performed by transferring knowledge from a larger teacher model to the YOLOv8-Nano student model. The contribution of different knowledge components is investigated through four experimental configurations, namely E0, E1, E2, and E3.

Design and Development

The design and development stage includes dataset preparation, teacher model training, student model training, Knowledge Distillation, and Android application development. The dataset consists of 3,090 labeled images representing six target classes: eye_closed, eye_open, yawn, no_yawn, head_dropped, and head_straight.

The teacher model is trained once for 100 epochs using the default random seed. The YOLOv8-Nano student model is subsequently trained under four experimental configurations. Each configuration is conducted three times for 30 epochsusing random seeds 42, 43, and 44. The repeated experiments are performed to assess the consistency of the model performance under different random initializations. The results for each configuration are summarized using the mean and standard deviation.

Demonstration

The demonstration stage applies the trained models to test images to verify their ability to detect the target driver conditions. The detection output consists of the predicted class, bounding box, and confidence score. The selected optimized model is subsequently integrated into the Android application to demonstrate its operation in a mobile detection environment.

Evaluation

The evaluation stage measures the performance of E0, E1, E2, and E3 using Precision, Recall, mAP@0.5, and mAP@0.5:0.95. The training loss is also analyzed to examine the learning behavior of the student model and the effect of the Knowledge Distillation components.

For the Android implementation, the selected model is evaluated based on the available deployment indicators, including inference time, frame rate, model size, and memory usage. The evaluation results are used to determine whether the optimized model provides an appropriate balance between detection performance and computational efficiency.

Communication

The final stage communicates the research findings through the comparison of E0, E1, E2, and E3. The analysis includes quantitative performance results, loss behavior, qualitative detection results, and Android implementation results. The findings are used to determine the contribution of each Knowledge Distillation component and to identify the optimized model configuration.

Development Method

The Android application is developed using the Agile development method. Agile methods are widely used in software development to support iterative development and continuous refinement of software products (Edison et al., 2022) . In this research, the development process consists of six stages: Plan, Design, Develop, Test, Deploy, and Review.

Plan

The Plan stage identifies the functional requirements of the Android application, including camera input, driver image acquisition, object detection, visualization of detection results, confidence scores, and drowsiness warning functionality.

Design

The Design stage defines the application interface and detection workflow. The design includes the camera interface, detection result display, bounding box visualization, confidence score, and the mechanism used to determine the driver's condition.

Develop

The Develop stage implements the Android application and integrates the selected YOLOv8-Nano model. The implementation includes image acquisition, preprocessing, model inference, post-processing, and visualization of the detection results.

Test

The Test stage verifies the functionality of the Android application and the execution of the detection model on the target device. Functional testing is conducted to ensure that the application operates according to the

defined requirements, while model testing verifies that detection results can be generated correctly.

Deploy

The Deploy stage places the completed application and optimized detection model on the target Android device. The application is then used to perform driver drowsiness detection through the device camera.

Review

The Review stage evaluates the application and identifies aspects requiring improvement. The review considers the detection results, application functionality, and computational performance of the optimized model on the Android device.

Model Optimization Method

The Model Optimization Method applies Knowledge Distillation (KD) to transfer knowledge from a larger teacher model to the YOLOv8-Nano student model. Knowledge Distillation has been studied as a means of improving lightweight object detectors while maintaining lower computational requirements, particularly for resource-constrained environments (Cao et al., 2023; Park et al., 2024).

Teacher–Student Model

The Knowledge Distillation framework consists of a teacher model and a student model. The teacher model provides additional knowledge to guide the learning process of YOLOv8-Nano. Both models receive the same input data, while the outputs and intermediate representations of the teacher are used as additional supervision for the student model. This teacher–student structure is commonly used in recent object detection distillation approaches to improve compact detection models.

Hard Loss

Hard Loss represents the original learning objective derived from the ground-truth annotations. This loss provides the primary supervision for YOLOv8-Nano to learn the target classes and their corresponding object locations.

Logit Loss

Logit Loss transfers prediction-level knowledge from the teacher to the student. The difference between the teacher and student prediction distributions is measured using Kullback–Leibler (KL) Divergence. Logit-based distillation has been investigated as an effective mechanism for transferring prediction information in object detection models.

Bounding Box Loss

Bounding Box Loss transfers localization knowledge from the teacher model to the student model. The difference between the teacher and student bounding box predictions is measured using the Smooth L1 formulation. This component provides additional supervision for object localization.

Feature Loss

Feature Loss transfers intermediate visual representations from the teacher model to the student model. The difference between the teacher and student feature maps is measured using Mean Squared Error (MSE). Feature-level distillation provides additional representation-level supervision for the student model. Recent object detection studies have investigated feature imitation as one of the mechanisms for knowledge transfer between teacher and student models.

Total Distillation Loss

The total training objective combines the original hard loss with the Knowledge Distillation components:

$$L_{total} = L_{hard} + \lambda_{logit} L_{logit} + \lambda_{box} L_{box} + \lambda_{feature} L_{feature}$$

where L_hard represents the ground-truth supervision loss, L_logit represents the logit distillation loss, L_box represents the bounding box distillation loss, and L_feature represents the feature distillation loss. The parameters λ_logit, λ_box, and λ_featurecontrol the contribution of each distillation component according to the experimental configuration.

Experimental Scenarios

The optimization process is evaluated through four experimental configurations. E0 represents the YOLOv8-Nano baseline without Knowledge Distillation. E1 introduces logit distillation. E2 combines logit

and bounding box distillation, while E3 combines logit, bounding box, and feature distillation.

Each student configuration is trained three times using random seeds 42, 43, and 44 under a fixed budget of 30 epochs per run with default Ultralytics hyperparameters (initial learning rate lr0 = 0.01, momentum = 0.937, weight decay = 0.0005) without task-specific hyperparameter tuning. Although standalone YOLOv8 models typically require 100 to 300 epochs to reach full convergence (Jocher et al., 2023; Terven et al., 2023; Yu et al., 2025), restricting student training to 30 epochs with uniform default hyperparameters establishes a strictly controlled ablation baseline to evaluate how Knowledge Distillation accelerates convergence under constrained training budgets. The performance of each configuration is reported using the mean and standard deviation across the three runs. The experimental configuration is presented in Table 1.

Table 1. Experimental Configuration

Exp.	Hard Loss	Logit Loss	Bounding Box Loss	Feature Loss
E0	✓	–	–	–
E1	✓	✓	–	–
E2	✓	✓	✓	–
E3	✓	✓	✓	✓

The teacher model is trained separately once for 100 epochs using the default random seed and provides the reference knowledge for the Knowledge Distillation experiments.

The comparison between E0 and E1 is used to analyze the contribution of logit knowledge, the comparison between E1 and E2 evaluates the contribution of bounding box knowledge, and the comparison between E2 and E3 evaluates the additional contribution of feature knowledge. This progressive experimental design enables the effect of each Knowledge Distillation component on YOLOv8-Nano to be analyzed systematically.

Results and Discussion

This section presents the results of the YOLOv8-Nano optimization using Knowledge Distillation for driver drowsiness detection and its implementation on an Android device. The discussion covers the characteristics of the dataset, the performance of the teacher and student models, the results of the E0–E3 experiments, the contribution of each Knowledge Distillation component, qualitative detection results, and the implementation and functional evaluation of the optimized model on Android. Each student-model experiment was conducted three times using random seeds 42, 43, and 44 with 30 epochs, while the teacher model was trained once for 100 epochs using the default random seed.

Model Training and Evaluation Results

The dataset used in this research consists of 3,090 images with six target classes, namely eye_closed, eye_open, head_dropped, head_straight, no_yawn, and yawn. The dataset was divided into training, validation, and testing sets with proportions of 70%, 15%, and 15%, respectively.

Table 2. Dataset Distribution

Dataset	Number of Images	Percentage
Train	2,163	70%
Validation	464	15%
Test	463	15%
Total	**3,090**	**100%**

The training set contains 2,163 images, while the validation and test sets contain 464 and 463 images, respectively. This division provides separate data for model learning, validation during training, and final performance evaluation.

Based on the object annotations, the dataset contains 9,269 annotated objects distributed across the six classes.

Table 3. Distribution of Annotated Objects by Class

Class	Number of Objects	Proportion
eye_closed	1,340	14.5%
eye_open	1,749	18.9%

head_dropped	871	9.4%
head_straight	2,220	24.0%
no_yawn	1,754	18.9%
yawn	1,335	14.4%
Total	**9,269**	**100%**

The distribution of the annotated objects is further visualized to provide a clearer representation of the differences in the number of objects among the six target classes. Figure 1 presents the number of annotated objects for each class, while Figure 2 presents the same distribution in terms of percentage.

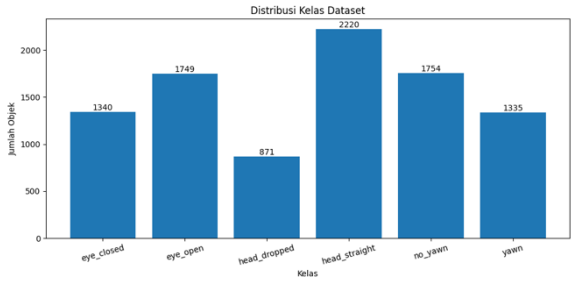


Figure 1. Distribution of annotated objects by class

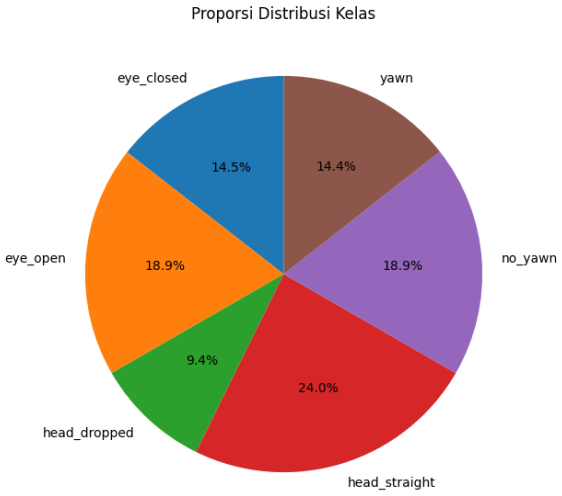


Figure 2. Proportion of dataset classes

The head_straight class has the largest number of annotated objects, with 2,220 instances or 24.0% of the total annotations. The smallest class is head_dropped, with 871 instances or 9.4%. The remaining classes have distributions between 14.4% and 18.9%. The dataset therefore contains all target classes with different frequencies, with head_droppedhaving the lowest representation and head_straight having the highest.

- Teacher Model

The teacher model was trained once for 100 epochs using the default random seed. The model was used to provide additional prediction, localization, and feature information during the Knowledge Distillation process.

Table 4. Teacher Model Performance

YOLOv8-Large				
Precision	**Recall**	**F1-score**	**mAP@0.5**	**mAP@0.5:0.95**
0.935	0.911	0.922840	0.940	0.651

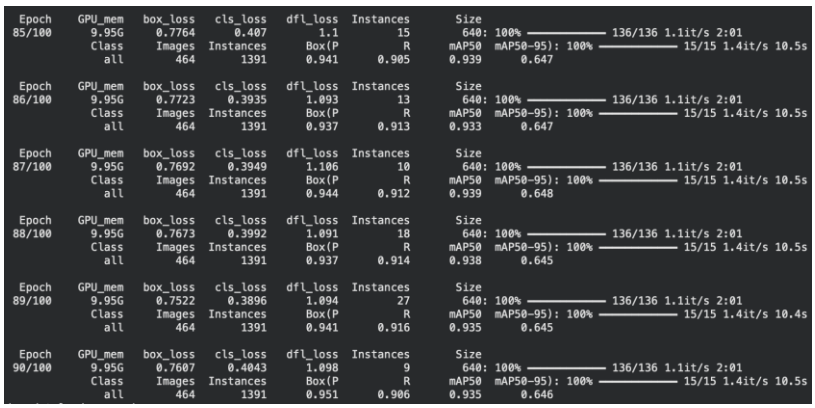


Figure 3. Training Results of the Teacher Model

The teacher model achieves a Precision of 0.935, Recall of 0.911, F1-score of 0.922840, mAP@0.5 of 0.940, and mAP@0.5:0.95 of 0.651. These results indicate that the teacher model provides a stronger detection reference than

the lightweight YOLOv8-Nano baseline and can therefore be used as a source of additional knowledge for the student model.

Knowledge Distillation Experimental Results

Four experimental configurations were used to evaluate the effect of Knowledge Distillation. E0 serves as the baseline, while E1, E2, and E3 progressively introduce logit, bounding box, and feature distillation. Each configuration was trained three times for 30 epochs using random seeds 42, 43, and 44.

- E0 – Baseline YOLOv8-Nano

E0 represents the YOLOv8-Nano model trained without Knowledge Distillation.

Table 5. E0 Results

Metric	Seed 42	Seed 43	Seed 44	Mean ± SD
Precision	0.57992	0.52155	0.42091	0.5075 ± 0.0804
Recall	0.42879	0.40280	0.43153	0.4210 ± 0.0159
F1-score	0.49303	0.45455	0.42615	0.4579 ± 0.0336
mAP@0.5	0.42202	0.38118	0.35018	0.3845 ± 0.0360
mAP@0.5:0.95	0.26127	0.24494	0.23133	0.2459 ± 0.0150

The baseline E0 configuration achieves a mean mAP@0.5 of 0.3845 (38.45%) and a mean mAP@0.5:0.95 of 0.2459, with a relatively high Precision standard deviation of 0.0804 across the three random seeds. Compared with external benchmarks—where pretrained YOLOv8-Nano achieves 37.3% mAP@0.5:0.95 on MS COCO over 300–500 epochs (Jocher et al., 2023; Terven et al., 2023) and domain-specific driver fatigue detectors trained for 100 or more epochs reach 80.5% to 90%+ mAP@0.5 (Debsi et al., 2024; Xie et al., 2025)—the lower baseline score of 0.3845 in E0 is attributable to three factors. First, the 30-epoch training window is intentionally short; recent YOLOv8 optimization studies confirm that standalone training without teacher guidance exhibits severe gradient fluctuations and slow convergence during the initial 100–130 epochs (Yu et al., 2025). Second, E0 was trained using un-tuned default hyperparameters (such as the standard

learning rate schedule) to maintain an identical baseline across all ablation runs. Third, the ultra-compact YOLOv8-Nano backbone (3.2M parameters) has limited representational capacity when simultaneously learning six fine-grained, visually overlapping facial and head-posture classes solely from hard one-hot labels (Cao et al., 2025). Consequently, E0 reflects an under-converged standalone baseline under a restricted 30-epoch budget, providing a clear reference for measuring the convergence-accelerating effect of Knowledge Distillation.

- E1 – Logit Distillation

E1 introduces logit distillation, allowing YOLOv8-Nano to receive prediction-level knowledge from the teacher model.

Table 6. E1 Results

Metric	Seed 42	Seed 43	Seed 44	Mean ± SD
Precision	0.80701	0.81392	0.78686	0.8026 ± 0.0141
Recall	0.82644	0.81835	0.83192	0.8256 ± 0.0068
F1-score	0.81661	0.81613	0.80876	0.8138 ± 0.0044
mAP@0.5	0.85849	0.85948	0.85799	0.8587 ± 0.0008
mAP@0.5:0.95	0.52715	0.52440	0.52772	0.5264 ± 0.0018

E1 substantially improves the performance of the baseline. The mean mAP@0.5 increases from 0.3845 in E0 to 0.8587 in E1, while mAP@0.5:0.95 increases from 0.2459 to 0.5264. The standard deviation also decreases considerably, indicating more consistent performance across the three random seeds. The result indicates that logit knowledge provides substantial additional supervision for YOLOv8-Nano.

- E2 – Logit and Bounding Box Distillation

E2 extends E1 by incorporating bounding box distillation to provide additional localization knowledge.

Table 7. E2 Results

Metric	Seed 42	Seed 43	Seed 44	Mean ± SD
Precision	0.80701	0.81392	0.78686	0.8026 ± 0.0141
Recall	0.82644	0.81835	0.83192	0.8256 ± 0.0068

F1-score	0.81661	0.81613	0.80876	0.8138 ± 0.0044
mAP@0.5	0.85849	0.85948	0.85799	0.8587 ± 0.0008
mAP@0.5:0.95	0.52715	0.52440	0.52772	0.5264 ± 0.0018

The addition of bounding box distillation produces another substantial improvement. The mean mAP@0.5 increases from 0.8587 in E1 to 0.9217 in E2, while mAP@0.5:0.95 increases from 0.5264 to 0.6224. The results indicate that localization knowledge from the teacher provides useful additional guidance for object detection.

- E3 – Logit, Bounding Box, and Feature Distillation

E3 combines all three Knowledge Distillation components: logit, bounding box, and feature distillation.

Table 8. E3 Results

Metric	Seed 42	Seed 43	Seed 44	Mean ± SD
Precision	0.92986	0.92969	0.91840	0.9260 ± 0.0066
Recall	0.89627	0.90597	0.90302	0.9018 ± 0.0050
F1-score	0.91276	0.91768	0.91065	0.9137 ± 0.0036
mAP@0.5	0.93102	0.93388	0.93208	0.9323 ± 0.0015
mAP@0.5:0.95	0.64025	0.63463	0.63544	0.6368 ± 0.0030

E3 achieves the highest overall performance among the student-model configurations. The model obtains a mean Precision of 0.9260, Recall of 0.9018, F1-score of 0.9137, mAP@0.5 of 0.9323, and mAP@0.5:0.95 of 0.6368. Compared with E2, the addition of feature distillation increases mAP@0.5 from 0.9217 to 0.9323 and mAP@0.5:0.95 from 0.6224 to 0.6368.

Comparison and Analysis of Knowledge Distillation

The results of E0–E3 are summarized to analyze the progressive contribution of each Knowledge Distillation component.

Table 9. Performance Comparison of E0–E3

Metric	E0	E1	E2	E3

Precision	0.5075 ± 0.0804	0.8026 ± 0.0141	0.9251 ± 0.0088	**0.9260 ± 0.0066**
Recall	0.4210 ± 0.0159	0.8256 ± 0.0068	0.8857 ± 0.0089	**0.9018 ± 0.0050**
F1-score	0.4579 ± 0.0336	0.8138 ± 0.0044	0.9049 ± 0.0021	**0.9137 ± 0.0036**
mAP@0.5	0.3845 ± 0.0360	0.8587 ± 0.0008	0.9217 ± 0.0021	**0.9323 ± 0.0015**
mAP@0.5:0.95	0.2459 ± 0.0150	0.5264 ± 0.0018	0.6224 ± 0.0021	**0.6368 ± 0.0030**

The results demonstrate a progressive improvement in detection performance as additional Knowledge Distillation components are introduced. The largest improvement occurs from E0 to E1. The mean mAP@0.5 increases by 0.4742, while mAP@0.5:0.95 increases by 0.2805. This indicates that logit distillation provides the largest contribution among the evaluated components.

The transition from E1 to E2 produces a further increase of 0.0630 in mAP@0.5 and 0.0960 in mAP@0.5:0.95. This result demonstrates the additional value of bounding box knowledge for localization.

The transition from E2 to E3 produces a smaller but consistent improvement of 0.0106 in mAP@0.5 and 0.0144 in mAP@0.5:0.95. Although the marginal improvement is smaller than the previous stages, E3 achieves the highest overall performance.

Compared with the baseline E0, E3 increases mAP@0.5 from 0.3845 to 0.9323 (+0.5478) and mAP@0.5:0.95 from 0.2459 to 0.6368 (+0.3909) under the exact same 30-epoch schedule and default hyperparameter configuration. While E0 suffers from early-epoch under-convergence and suboptimal feature discrimination at 30 epochs (Terven et al., 2023; Yu et al., 2025), the continuous soft supervision transferred from the 100-epoch YOLOv8-Large teacher in E1–E3 acts as an effective regularizer and gradient guide (Cao et al., 2025). This teacher guidance enables YOLOv8-

Nano to bypass the slow warm-up phase of standalone training and reach performance comparable to the 100-epoch teacher (0.940 mAP@0.5) within only 30 epochs.

- Stability Across Random Seeds

The repeated experiments also show a reduction in variability after Knowledge Distillation is introduced. The standard deviation of Precision decreases from 0.0804 in E0 to 0.0141 in E1, 0.0088 in E2, and 0.0066 in E3. The standard deviation of mAP@0.5 decreases from 0.0360 in E0 to 0.0008 in E1, 0.0021 in E2, and 0.0015 in E3.

These results indicate that the Knowledge Distillation configurations produce more consistent results across different random seeds than the baseline configuration.

Qualitative Detection Results

Quantitative evaluation is complemented by a qualitative analysis of the visual detection outputs. A subset of 30 sample test images containing upright-posture drivers with varying ocular (eye_open, eye_closed) and oral (yawn, no_yawn) expressions was evaluated to visually inspect global head-posture localization before and after Knowledge Distillation, while full multi-class detection across all six target classes is evaluated across the complete test set in Figure 5.



Figure 4. Qualitative Comparison of Detection Results Before and After Optimization

Figure 4 illustrates the qualitative comparison on the 30 sample images. Under the under-converged E0 baseline (top panel), predictions on several

frames fail to exceed the visualization confidence threshold or exhibit unstable localization. In contrast, the E3 configuration (bottom panel) consistently localizes the driver region with high confidence. It should be noted that in this multi-label driver monitoring dataset, global head posture (head_straight or head_dropped) co-occurs spatially with localized eye and mouth states (eye_closed, eye_open, yawn, and no_yawn) within the same frame; thus, a driver who is yawning or closing their eyes while facing forward simultaneously carries a head_straight annotation alongside the corresponding yawn or eye_closed annotation. For visual clarity in the 30-image montage (Figure 4) and the single-instance log in Table 10, the export script recorded the top-1 highest-confidence global bounding box per image (head_straight) rather than overlaying all overlapping local eye and mouth boxes simultaneously.

Table 10. Primary Global Head-Posture Detection Confidence on 30 Sample Images (E3)

Image	Detected Class	Confidence
01.jpg	head_straight	0.9573
02.jpg	head_straight	0.9946
03.jpg	head_straight	0.9551
04.jpg	head_straight	0.9624
05.jpg	head_straight	0.9412
...	...	...
30.jpg	head_straight	0.9148

Table 10 summarizes the top-1 global head-posture detection output logged for the 30 upright-posture sample images under the E3 configuration, recording the primary bounding-box class and confidence score derived from the combined Hard Loss, Logit Loss, Box Loss, and Feature Loss optimization.

Across the 30 sample frames in Table 10, the primary global head-posture bounding box (head_straight) achieved confidence scores ranging from 0.9099 (28.jpg) to 0.9946 (02.jpg). While Table 10 logs only the

dominant global head-posture box per frame, the E3 model actively detects all six target classes—including drowsiness indicators such as closed eyes (eye_closed), yawning (yawn), and nodding/dropped head (head_dropped)—as confirmed quantitatively by the class-level evaluation on the full test set presented in the Confusion Matrix (Figure 5), which records 193 true positive detections for eye_closed, 250 for eye_open, 190 for yawn, 246 for no_yawn, 93 for head_dropped, and 300 for head_straight.

Confusion Matrix and Class-Level Detection

The confusion matrix is used to examine the distribution of correct and incorrect predictions across the target classes. The matrix includes the six target classes and a background category.

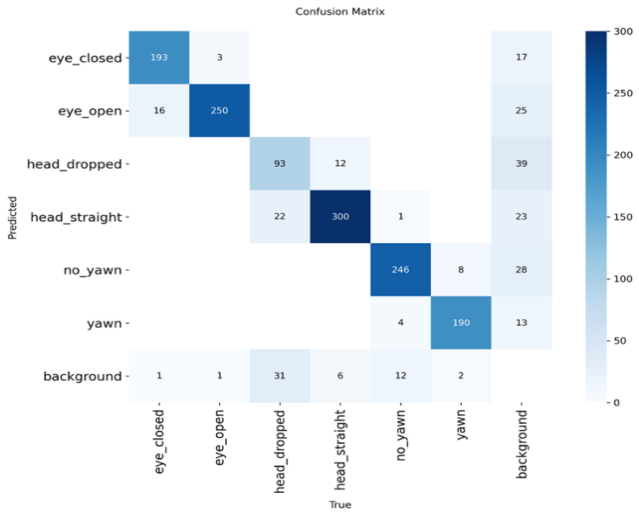


Figure 5. Confusion Matrix of the Selected Model

The diagonal values of the confusion matrix indicate correctly detected instances. The highest diagonal value is observed for head_straight with 300 instances, followed by eye_open with 250, no_yawn with 246, eye_closed with 193, yawn with 190, and head_dropped with 93.

The matrix also indicates several misclassifications among target classes. For example, 22 true head_dropped instances are predicted as head_straight, while 12 true head_straight instances are predicted as head_dropped. The presence of these errors indicates that the visual

characteristics of some eye, mouth, and head conditions can overlap and remain challenging for the detection model.

The confusion matrix therefore complements the overall evaluation metrics by showing the detailed distribution of correct predictions and errors among the target classes.

Android Implementation and Performance Evaluation

The optimized YOLOv8-Nano model is subsequently implemented on an Android device to evaluate its suitability for mobile deployment. The computational characteristics of YOLOv8-Large, YOLOv8-Nano, and YOLOv8-Nano KD are compared.

Table 11. Model Performance and Computational Efficiency

Spec	**YOLOv8**		
	Large	**Nano**	**Nano KD**
Params (M)	43.7	3.2	**3.2**
GFLOPs	165.2	8.7	**8.7**
Size (MB)	87.7	6.2	**6.2**
Inference Time (ms)	532–983	176–321	**182–343**
FPS	1–1.9	2.64–5.32	**2.91–5.82**

The results show that Knowledge Distillation does not increase the structural complexity of the student model. YOLOv8-Nano KD retains 3.2 million parameters, 8.7 GFLOPs, and a model size of 6.2 MB, identical to the original YOLOv8-Nano. This indicates that the optimization takes place during the training process without increasing the size of the deployed student architecture.

YOLOv8-Nano KD achieves an inference time of 182–343 ms and an observed frame rate of 2.91–5.82 FPS on the target Android device. The baseline YOLOv8-Nano records an inference time of 176–321 ms and 2.64–5.32 FPS. Meanwhile, the larger YOLOv8-Large requires substantially greater computational resources, with 43.7 million parameters, 165.2 GFLOPs, an 87.7 MB model size, and an inference time of 532–983 ms.

Although the inference time of YOLOv8-Nano KD is slightly higher than the baseline, its detection performance is substantially better while maintaining the same model size and computational complexity. This demonstrates that Knowledge Distillation can improve the detection capability of the lightweight model without directly increasing its parameter count or GFLOPs.

Functional Testing

Black Box testing was conducted to verify the functionality of the Android application. The testing focuses on the main functions involved in the driver drowsiness detection process.

Table 12. Black Box Testing Results

Feature	Test Scenario	Status
Camera	Open the camera	Pass
Frame Input	Provide driver frame	Pass
Object Detection	Perform detection on input	Pass
Bounding Box	Detect target object	Pass
Class Label	Detect target condition	Pass
Confidence Score	Generate detection result	Pass
Drowsiness Warning	Detect drowsiness condition	Pass

The Black Box testing results show that all tested application functions operate according to the expected behavior. Every test scenario obtained a Pass status, indicating that the main Android application functions can be executed successfully from input acquisition through presentation of the detection results and drowsiness warning.

Overall Discussion

The experimental results demonstrate that Knowledge Distillation substantially improves the detection performance of YOLOv8-Nano under

a constrained 30-epoch training schedule. Whereas the standalone baseline E0 obtains a mean mAP@0.5 of only 0.3845 due to early-epoch under-convergence, default un-tuned hyperparameters, and the limited representational capacity of the 3.2M-parameter nano architecture on fine-grained facial states (Terven et al., 2023; Cao et al., 2025; Yu et al., 2025), the complete E3 configuration reaches 0.9323 within the same 30 epochs. This confirms that multi-level teacher supervision effectively compensates for short training schedules and compact backbone capacity.

The largest improvement occurs after the introduction of logit distillation in E1. The mean mAP@0.5 increases from 0.3845 to 0.8587, indicating that prediction-level knowledge provides substantial additional supervision for the student model. The addition of bounding box distillation in E2 further increases the mean mAP@0.5 to 0.9217, indicating the contribution of localization information. Feature distillation in E3 produces the highest final performance, with a mean mAP@0.5 of 0.9323 and mAP@0.5:0.95 of 0.6368.

In addition to improved performance, the Knowledge Distillation configurations demonstrate lower variation across the three random seeds compared with the baseline. This is particularly evident in the standard deviation of Precision and mAP@0.5, indicating more consistent training outcomes.

The comparison with YOLOv8-Large also demonstrates the efficiency advantage of YOLOv8-Nano. YOLOv8-Nano KD maintains 3.2 million parameters, 8.7 GFLOPs, and a 6.2 MB model size while achieving substantially higher detection performance than the baseline YOLOv8-Nano. Therefore, the optimization improves the capability of the lightweight model without increasing its structural complexity.

The Android implementation confirms that the optimized YOLOv8-Nano can be deployed on a mobile device while preserving its lightweight characteristics. Although its inference time is slightly higher than the baseline in the observed measurement, the model provides substantially better detection performance. The successful results of Black Box testing

further indicate that all main application functions operate according to the defined requirements.

Overall, the results indicate that the combination of logit, bounding box, and feature distillation in E3 provides the best detection performance among the evaluated YOLOv8-Nano configurations, while maintaining the lightweight model characteristics required for Android deployment.

Conclusion

This research demonstrates that Knowledge Distillation can effectively optimize the YOLOv8-Nano model for driver drowsiness detection while maintaining its lightweight characteristics for Android deployment. The YOLOv8-Nano baseline (E0) achieved a mean mAP@0.5 of 0.3845 ± 0.0360, while the progressive application of Knowledge Distillation improved the detection performance through E1, E2, and E3.

The addition of logit distillation in E1 produced the largest performance improvement compared with the baseline, increasing the mean mAP@0.5 to 0.8587 ± 0.0008. The addition of bounding box distillation in E2 further increased the mean mAP@0.5 to 0.9217 ± 0.0021, demonstrating the contribution of localization knowledge. The complete configuration in E3, which combines logit, bounding box, and feature distillation, achieved the highest overall performance, with a Precision of 0.9260 ± 0.0066, Recall of 0.9018 ± 0.0050, F1-score of 0.9137 ± 0.0036, mAP@0.5 of 0.9323 ± 0.0015, and mAP@0.5:0.95 of 0.6368 ± 0.0030.

The results also show that the use of Knowledge Distillation produces more consistent performance across the three random seeds 42, 43, and 44. In addition, the optimized YOLOv8-Nano KD maintains 3.2 million parameters, 8.7 GFLOPs, and a 6.2 MB model size, indicating that the improvement in detection performance does not require an increase in the structural complexity of the student model.

The implementation on Android shows that YOLOv8-Nano KD can be deployed while maintaining its lightweight characteristics. The model achieves an inference time of 182–343 ms with an observed frame rate of

2.91–5.82 FPS on the target device. Functional testing using the Black Box method also shows that all tested application functions obtain a Pass result.

Overall, E3 is the best-performing configuration among the evaluated YOLOv8-Nano models and provides the most favorable balance between detection performance and model efficiency for Android-based driver drowsiness detection.

References

Alpiyasin, F., Hermanto, C., Khoirunnisa, N., Rahmadanti, C., & Mutmainah, N. (2025). Implementation of ResNet 101 Architecture in Convolutional Neural Network (CNN) Algorithm to Detect Diseases in Chili Plant Leaves Based on Image Processing. Majalah Bisnis & IPTEK, 18(2), 164–175. https://doi.org/10.55208/bistek.v18i2.407

Cao, S., Li, M., Hays, J., Ramanan, D., Wang, Y.-X., & Gui, L. (2023). Learning Lightweight Object Detectors via Multi-Teacher Progressive Distillation. Proceedings of the 40th International Conference on Machine Learning, Proceedings of Machine Learning Research, 202, 3577–3598.

Cao, X., Hu, Y., & Zhang, H. (2025). LKD-YOLOv8: A Lightweight Knowledge Distillation-Based Method for Infrared Object Detection. Sensors, 25(13), 4054. https://doi.org/10.3390/s25134054

Debsi, A., Ling, G., Al-Mahbashi, M., Al-Soswa, M., & Abdullah, A. (2024). Driver distraction and fatigue detection in images using ME-YOLOv8 algorithm. IET Intelligent Transport Systems, 18(10), 1910–1930. https://doi.org/10.1049/itr2.12560

Edison, H., Wang, X., & Conboy, K. (2022). Comparing Methods for Large-Scale Agile Software Development: A Systematic Literature Review. IEEE Transactions on Software Engineering, 48(8), 2709–2731. https://doi.org/10.1109/TSE.2021.3069039

Fonseca, T., & Ferreira, S. (2025). Drowsiness Detection in Drivers: A Systematic Review of Deep Learning-Based Models. Applied Sciences, 15(16), 9018. https://doi.org/10.3390/app15169018

Gu, Q., Yin, S., Wang, D., Li, X., & Yin, H. (2025). Mining truck driver fatigue driving detection based on improved YOLOv8. China Safety Science Journal, 35(1), 60–66. https://doi.org/10.16265/j.cnki.issn1003-3033.2025.01.0147

Haryanti, T., Rakhmawati, N. A., & Subriadi, A. P. (2024). Measuring the digital transformation maturity level independently with the design

science research methodology. Systems Engineering, 27(1), 159–176. https://doi.org/10.1002/sys.21714

Haryanti, T., Rakhmawati, N. A., Subriadi, A. P., & Tjahyanto, A. (2022). The Design Science Research Methodology (DSRM) for Self-Assessing Digital Transformation Maturity Index in Indonesia. Proceedings of the 2022 IEEE 7th International Conference on Information Technology and Digital Applications (ICITDA). https://doi.org/10.1109/ICITDA55840.2022.9971171

Hidayat, R. N., & Ibnugraha, P. D. (2025). The Dashboard of Remote Monitoring System for Driver Fatigue. Journal Europeen Des Systemes Automatises, 58(6), 1089–1100. https://doi.org/10.18280/jesa.580601

Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLOv8. https://github.com/ultralytics/ultralytics

Khan, N. A., & Rafat, A. M. S. (2025). Optimizing Deep Learning Models for Resource-Constrained Environments With Cluster-Quantized Knowledge Distillation. Engineering Reports, 7(5), e70187. https://doi.org/10.1002/eng2.70187

Lan, Q., & Tian, Q. (2024). Gradient-Guided Knowledge Distillation for Object Detectors. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 424–433.

Park, S., Kang, D., & Paik, J. (2024). Cosine similarity-guided knowledge distillation for robust object detectors. Scientific Reports, 14, 18888. https://doi.org/10.1038/s41598-024-69813-6

Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction, 5(4), 1680–1716. https://doi.org/10.3390/make5040083

Xie, Z., Hj Salam, M. S., & Taliba, J. (2025). Fatigue Driving Recognition Using Improved YOLOv8. International Journal of Innovative Computing, 15(2), 191–199. https://doi.org/10.11113/ijic.v15n2.549

Yu, H., Luo, Q., Peng, W., Zheng, L., Ju, J., & Zhuo, H. (2025). PKD-YOLOv8: A Collaborative Pruning and Knowledge Distillation Framework for Lightweight Rapeseed Pest Detection. Sensors, 25(16), 5004. https://doi.org/10.3390/s25165004

Conflicts Of Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the

manuscript, or in the decision to publish the results.

Use Of Artificial Intelligence (AI)
The authors transparently disclose the use of AI Tool exclusively for the purpose of linguistic editing and grammatical refinement of the original draft. The authors maintain full responsibility for the manuscript's substantive content, data integrity, andfinal scientific conclusions.sion to publish the results.


© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC Attribution-NonCommercial-ShareAlike 4.0) license (https://creativecommons.org/licenses/by-nc-sa/4.0/).