Available *online* at : http://chipset.fti.unand.ac.id/

Journal on Computer Hardware, Signal Processing, Embedded System and Networking

| ISSN (Online) 2722-4422 |

Computer Vision

Computer Vision-Based PPE Compliance Detection in a Pharmaceutical Laboratory

Abil Rahman[1], *Yoan Purbolingga*[*2]

[1,2] *Department of Computer Engineering, Faculty of Information Technology, Andalas University, Padang, West Sumatra, 25163, Indonesia*

ARTICLE INFORMATION

Received: August 20, 2026
Revised: October 3, 2026
Available online: October 31, 2026

KEYWORDS

PPE, *Computer Vision*, YOLOv8s, Object Detection, PWA.

CORRESPONDENCE

E-mail: yoanpurbolingga@it.unand.ac.id

A B S T R A C T

The use of Personal Protective Equipment (PPE) is important for maintaining practitioner safety in pharmaceutical laboratories; however, manual inspection may lead to oversight. This study aims to develop a system capable of automatically detecting the use of masks, laboratory coats, and gloves and providing detection results. The system was developed using a computer vision approach with the YOLOv8s model implemented on the NVIDIA Jetson Orin NX and trained using five classes: mask, labcoat, glove, no_glove, and no_labcoat. The detection results were used to determine PPE completeness, provide notifications through an LCD and speaker, and record violations through the Firebase Realtime Database for subsequent display on a Progressive Web Application (PWA). Evaluation results show that the model achieved 89% accuracy and 96.5% mAP@0.5, while testing on 40 samples resulted in a 95% detection success rate. The system also operated continuously for 8 hours without lag or crashes, with an FPS of 14–16 and power consumption of 5.7–9.2 W. These results indicate that the developed system can perform real-time PPE detection and monitoring with stable performance and can support PPE compliance supervision in a pharmaceutical laboratory.

INTRODUCTION

Pharmaceutical laboratories are educational and research environments that require consistent implementation of safety procedures. During solid dosage form processing, practitioners may be exposed to active ingredients and pharmaceutical dust; therefore, the use of Personal Protective Equipment (PPE), particularly laboratory coats, masks, and gloves, is an important part of risk control. Biosafety guidelines and pharmaceutical production practices emphasize compliance with safety procedures as an essential element in protecting personnel and maintaining process quality [1].

The main problem addressed in this study is that PPE use is still monitored through visual inspection by laboratory personnel. Manual inspection may be missed once practitioners enter the laboratory area, resulting in discontinuous monitoring. Studies on student compliance with PPE use have also shown that knowledge and attitudes do not always translate into actual compliance, indicating the need for additional monitoring mechanisms [2]. This condition motivates the application of a computer vision-based system capable of performing objective and repeated inspection.

Previous studies have developed PPE detection systems based on deep learning. Pratama, Kartikasari, and Setiawan applied YOLOv5 to detect PPE compliance in hazardous areas [3], while other research has applied deep learning to real-time visual detection [4]. In the pharmaceutical field, computer vision has also been used for product quality inspection and process automation [5][7]. Nevertheless, applying PPE detection specifically in educational pharmaceutical laboratories still requires adaptation to user characteristics, lighting conditions, camera distance, and integrated monitoring requirements.

This study focuses on the final system version realized in Chapter IV of the thesis. The initial multifunctional design was modified during implementation; therefore, this article focuses primarily on the YOLOv8s-based PPE detection subsystem implemented on the NVIDIA Jetson Orin NX and its integration with the LCD, speaker, Firebase Realtime Database, and PWA. The implementation changes included replacing the processing unit, adding a speaker and numeric keypad, adjusting the mechanical layout, adding the no_glove and no_labcoat classes to the dataset, and applying temporal validation to reduce false decisions when a person merely passes in front of the camera.

The objective of this study is to develop a system capable of detecting the completeness of masks, laboratory coats, and gloves in real time, determining compliance status, providing immediate feedback through the LCD and audio, and digitally recording violations. The evaluation covers model performance, real-time detection functionality, operational stability, the effects of distance and lighting, computational load, notification accuracy, and data integration with the PWA.

https://doi.org/10.25077/chipset.7.02.140-148.2026
Attribution-NonCommercial-ShareAlike 4.0 International. Some rights reserved

METHOD

System Architecture

The final implementation uses the NVIDIA Jetson Orin NX as the processing center. The main camera captures video of users at the laboratory entrance, and the images are processed using the YOLOv8s model. The detection results are passed to validation logic to determine the PPE status. The status is then displayed on a 7-inch LCD and communicated through a speaker. When a violation occurs, the data are sent to the Firebase Realtime Database and displayed on the PWA for remote monitoring.

The hardware changes from the initial design were an important part of the implementation. The initial system planned to use an NVIDIA Jetson Nano, whereas the final version uses the Jetson Orin NX because the computational workload includes YOLOv8s inference, video processing, audio playback, LCD display, and database communication. In addition, an Advance Digital CLS-110 speaker was added for audio notification, and a mini numeric keypad was used to switch between the PPE detection mode and the drug-dimension measurement module. In this article, the evaluation focuses on the PPE detection path.

Table 1. Final Components and Functions

Component	Function
NVIDIA Jetson Orin NX	YOLOv8s inference processing, validation logic, display, audio, and Firebase communication
Full HD 1080p Webcam	Real-time acquisition of user images/video
7-inch LCD	Displays video, bounding boxes, PPE status, FPS, and violation count
Advance CLS-110	Audio notification based on PPE status
7-Port USB 3.0 Hub	Connects the camera, display, speaker, and USB devices simultaneously
Mini numeric keypad	System mode switching input
Firebase Realtime Database	Violation data storage and synchronization
PWA	Real-time monitoring of violation counts and history

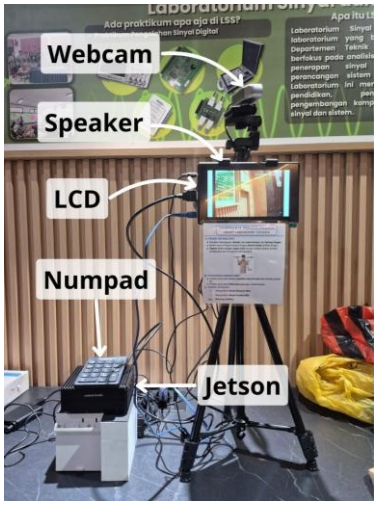

Figure 1. Hardware implementation of the system in the laboratory environment

Dataset and Annotation

The final dataset consisted of 684 images before augmentation and 1,478 images after augmentation. A total of 198 primary images were collected using a camera while considering variations in lighting, distance, viewing angle, and user characteristics. Another 486 images were obtained from secondary sources, including Google, Pinterest, and public Roboflow datasets. All images were manually annotated using Roboflow. The final dataset was divided at a ratio of 80:10:10 for training, validation, and testing.

The dataset was developed through four iterations. The initial iteration contained only three classes: mask, glove, and labcoat. The no_glove class was then added to distinguish bare hands, while no_labcoat was added because ordinary clothing, particularly white shirts, has visual characteristics similar to laboratory coats. These changes were driven by real-world implementation requirements and testing results during development [3].

Table 2. Final Dataset Composition

Class	Images	Role
Mask	390	PPE mask
Labcoat	393	PPE laboratory coat
Glove	552	PPE glove
No_labcoat	372	Condition without laboratory coat
No_Glove	354	Condition without glove
Total	2.061*	Class label total*

The values in the table represent the total number of annotations/classes used during training and are not identical to the number of unique images, because a single dataset image may contain more than one labeled object.

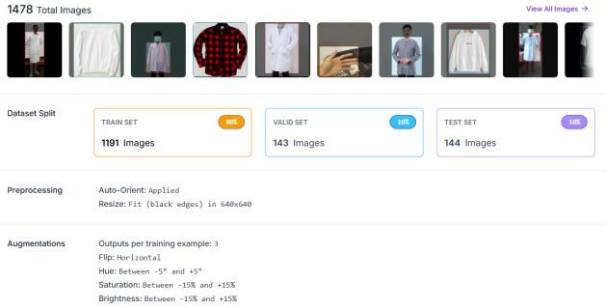

Figure 2. Final dataset in Roboflow after preprocessing and augmentation

YOLOv8s Model Training

The detection model uses YOLOv8s from Ultralytics. Training was performed offline using Python and Visual Studio Code with a transfer learning approach. Initial YOLOv8s weights pretrained on the COCO dataset were used as the starting point, followed by fine-tuning on the five PPE classes. The input size was set to 640 × 640 pixels, and training was performed for 100 epochs with a batch size of 16, 4 workers, the AdamW optimizer, and an initial learning rate of 0.001.

During Roboflow preprocessing, the images underwent Auto-Orient and Fit with black edges to preserve the image aspect ratio. Augmentation included horizontal flip, hue ±5%, saturation ±15%, and brightness ±15%. During training, augmentation was continued with rotation up to ±8°, translation 5%, scaling 30%, and mosaic 30%. This combination was selected to increase variation in position, size, and visual conditions without changing the basic characteristics of PPE objects.

Table 3. YOLOv8s Training Configuration

Parameter	Configuration
Model	YOLOv8s
Input size	640 × 640 px
Epoch	100
Batch size	16
Workers	4
Optimizer	AdamW
Initial learning rate	0,001

https://doi.org/10.25077/chipset.7.02.140-148.2026

Dataset split	80:10:10
Classes	5

Inference Implementation and Validation Logic

During inference, the webcam produces video at a resolution of 640 × 480 pixels, and each frame is processed by the model at an input size of 640 × 640 pixels. The software implementation uses Python and OpenCV for image acquisition and Ultralytics YOLO for inference. GPU acceleration is provided through CUDA, while FP16, cuDNN benchmark, layer fusion, MJPEG format, and a one-frame camera buffer are used to reduce processing load and latency.

The system uses five detection classes: mask, labcoat, glove, no_glove, and no_labcoat. However, compliance evaluation uses only three mandatory attributes: mask, laboratory coat, and glove. The glove and no_glove classes are used as triggers so that evaluation is not activated merely because a person passes in front of the camera. Once the trigger is detected, each attribute is validated consistently for at least three seconds. If an attribute is absent for more than 1.2 seconds, its validation timer is reset. If the PPE remains incomplete after eight seconds, the system assigns a violation status.

The final detection results are displayed as bounding boxes, attribute status, FPS, and violation count on the LCD. When a violation occurs, the data are first placed in the offline queue. If the network is available, the data are sent to Firebase; when the network is disconnected, the data continue to be processed locally and are transmitted again once the connection is restored. Audio playback and data transmission use separate threads so that they do not interfere with the main inference process.

Testing Procedure

Testing was divided into five main groups: (1) system functionality, (2) camera capability, (3) computational performance, (4) notification and monitoring, and (5) overall system testing. Functional testing used 40 samples with various PPE combinations, and each sample was tested three times. Camera testing covered distances of 50–300 cm and illumination levels of 20–131 lux. Performance testing observed FPS, CPU, GPU, GPU clock, RAM, temperature, and power. Notification testing measured audio delay and speech clarity. Monitoring integration testing measured successful synchronization between Firebase and the PWA.

RESULTS AND DISCUSSION

Model Performance after Training

Model evaluation results show that YOLOv8s was able to learn the five classes well. The confusion matrix shows that most predictions are located on the main diagonal. The glove, labcoat, mask, no_glove, and no_labcoat classes each have predominantly correct predictions, while errors mainly arise from undetected objects and a small number of glove–no_glove and labcoat–no_labcoat class confusions. The model accuracy calculated from the evaluation results is 89%.

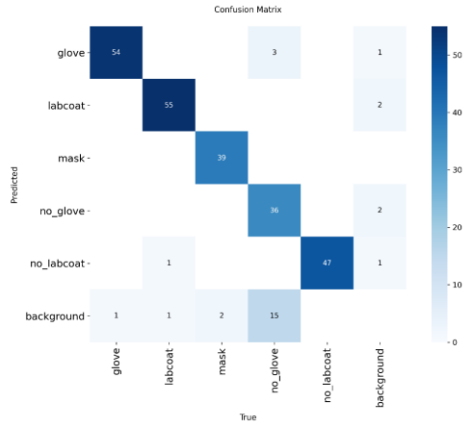

Figure 3. Confusion matrix of the YOLOv8s model for the five PPE classes.

The training graph shows a decrease in both training loss and validation loss through the end of 100 epochs. At the end of training, precision was approximately 97%, recall approximately 93%, mAP@0.5 approximately 97%, and mAP@0.5–0.95 approximately 80%. These results indicate that the model reached convergence and was sufficiently stable for deployment on the edge device.

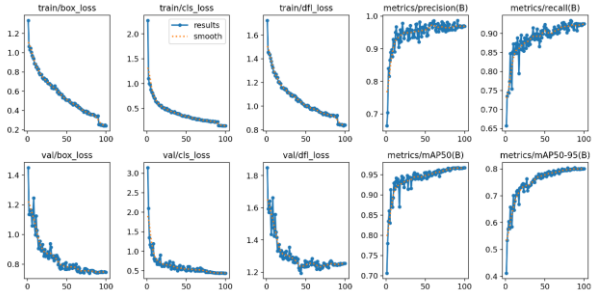

Figure 4. Training and validation curves of the YOLOv8s model over 100 epochs.

Precision–Recall analysis yielded AP@0.5 values of 0.989 for glove, 0.987 for labcoat, 0.985 for mask, 0.875 for no_glove, and 0.989 for no_labcoat. The overall mAP@0.5 reached 0.965. The no_glove class had the lowest performance, indicating that distinguishing bare hands from gloved hands remains more challenging than the other classes.

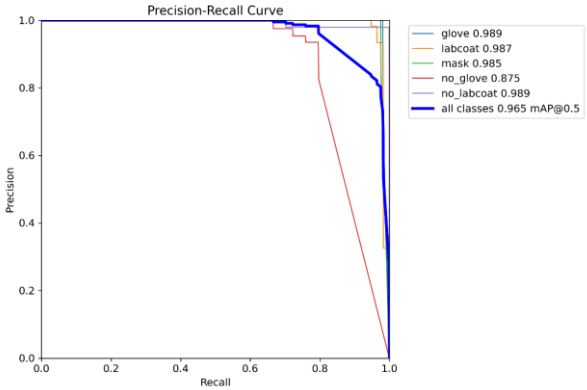

Figure 5. Precision–Recall curve of the YOLOv8s model.

Table 4. Model Evaluation Results

Metric	Result
Accuracy	89%
Precision	≈97%
Recall	≈93%

mAP@0.5	96,5%
mAP@0.5–0.95	≈80%
Maximum F1-score	0,94
Lowest AP@0.5	No_glove = 0,875

Real-Time PPE Detection Testing

Among 40 samples representing various PPE combinations, 38 samples were correctly recognized according to the ground truth and two samples were misclassified, resulting in a system success rate of 95%. Each sample was tested three times to evaluate consistency. The main errors occurred when distinguishing glove from no_glove and labcoat from no_labcoat under less optimal lighting. White shirts were a particular challenge because their visual characteristics are similar to laboratory coats.

Table 5. Functional Detection Results

Aspect	Result
Number of samples	40
Repetitions per sample	3 times
Correct detections	38
Incorrect detections	2
Detection success rate	95%
Most challenging classes	No_glove; no_labcoat
Evaluation trigger	Glove / no_glove
Consistent validation	>= 3 seconds

8-Hour Operational Stability

Continuous testing showed that the system was able to operate for eight hours without lag or crashes. FPS ranged from 14–16 FPS. CPU temperature increased from 49.59°C at the beginning of the test to a maximum of 62.94°C, while GPU temperature increased from 47.94°C to a maximum of 61.16°C. Initial power consumption was 5.7 W and remained within 8.1–9.2 W after the system reached its operating condition, with an average of approximately 8.49 W after the initial phase. No continuous increase in temperature or power was observed, indicating relatively stable operating conditions.

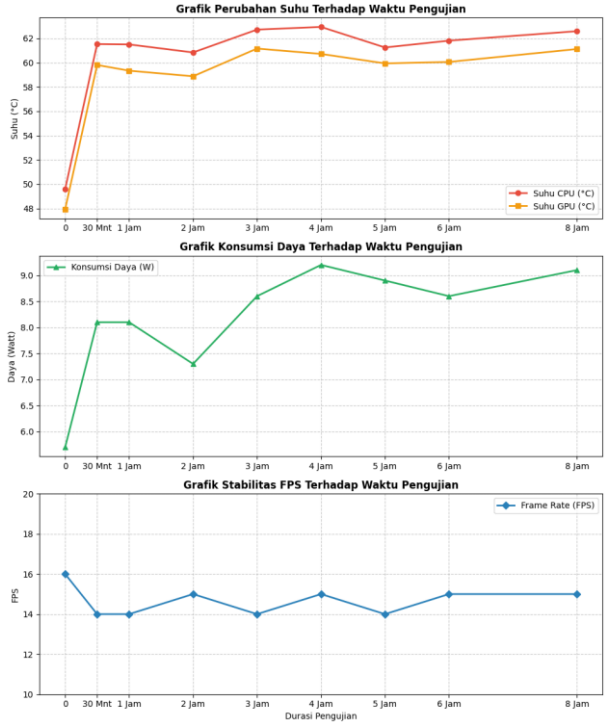

Figure 6. Temperature, power consumption, and FPS variation during the 8-hour test

Table 6. Stability Test Summary

Parameter	Result
Operation duration	8 hours
FPS	14–16 FPS
CPU temperature	49.59–62.94 °C
GPU temperature	47.94–61.16 °C
Power	5.7–9.2 W
Operational status	Normal; no lag/crash

Effect of Object Count on Computational Load

Increasing the number of objects increased GPU load more noticeably than CPU and RAM usage. With one object, GPU utilization was 27.6% and FPS was 14.80. When the number of objects increased to two, GPU utilization rose to 40% with an FPS of 14. At three objects, GPU utilization reached 80.6% and FPS decreased to 13.50. CPU usage remained around 22% and RAM usage ranged from 7.7–7.9 GB. These results show that the Jetson Orin NX was still capable of real-time detection, but multi-object usage significantly increased accelerator load.

Table 7. Computational Load versus Number of Objects

Objects	FPS	CPU (%)	GPU (%)	RAM (GB)
1	14,80	22,4	27,6	7,9
2	14,00	22,0	40,0	7,7
3	13,50	22,2	80,6	7,8

Based on these results, the final implementation was configured to inspect one user at a time at the entrance area. This approach reduces the influence of other users who merely pass by and keeps the validation, notification, and data transmission processes focused on the subject being inspected.

Camera Capability under Different Distances and Lighting Conditions

Distance testing showed the best performance at 100–150 cm. At 100–200 cm, all PPE attributes could still be detected well, whereas at 50 cm the laboratory coat was not captured completely because the user was too close to the camera. At 250 cm, the mask and gloves became difficult to detect, and at 300 cm the system primarily recognized large objects such as the laboratory coat. Based on these results, the camera was installed approximately 80–100 cm from the user.

Table 8. Detection by Object Distance

Distance	Observation
50 cm	Laboratory coat not fully captured; mask and glove still detected
100 cm	All attributes detected stably
150 cm	All attributes detected stably
200 cm	All attributes still detected
250 cm	Mask and glove begin to fail to detect
300 cm	Mainly large objects are detected

In illumination testing, the system successfully detected PPE across 20–131 lux. At 20 lux, the image was darker but detection still succeeded. Above 60 lux, the results became more stable, while laboratory conditions of 96 and 131 lux produced the most consistent performance. These findings support installation in laboratory areas that generally have illumination around 100–300 lux.

Table 9. Detection by Illumination

Illumination	Result
20 lux	Detection successful, but the image is darker
61–64 lux	Detection successful and more stable
96 lux	Detection successful and consistent in the laboratory
131 lux	Detection successful and consistent in the laboratory

https://doi.org/10.25077/chipset.7.02.140-148.2026

Audio Notification and PWA Monitoring

Audio notification was tested under eight conditions, ranging from complete PPE to all attributes being incomplete. All audio outputs matched the results shown on the LCD. Playback delay ranged from 0.26–0.41 seconds, with an average of approximately 0.31 seconds. This indicates that audio information can be delivered immediately after the PPE status is determined.

Table 10. Audio Notification Test

Parameter	Result
Scenario	8 PPE condition combinations
Minimum delay	0.26 seconds
Maximum delay	0.41 seconds
Average delay	≈0.31 seconds
Audio-LCD consistency	100% in test scenarios

In terms of coverage, sound intensity was recorded at 83 dB at 0.5 m, 77 dB at 1 m, 73 dB at 1.5 m, 71 dB at 2 m, 69 dB at 2.5 m, and 64 dB at 3 m. All notifications remained clearly audible under the test-room conditions, indicating that the speaker successfully served as an additional feedback channel.



Figure 7. Example of PPE detection results displayed on the LCD



Figure 8. PPE monitoring page on the Progressive Web Application

Integration between Firebase Realtime Database and the PWA was successfully implemented using real-time synchronization. When the network was connected, violation data were displayed with a delay of approximately 58–70 ms. When the connection was interrupted, the system continued detection and temporarily stored data through the offline queue. Once the connection was restored, the pending data were sent automatically and displayed on the PWA. This mechanism improves system resilience to network disruptions without stopping local detection.

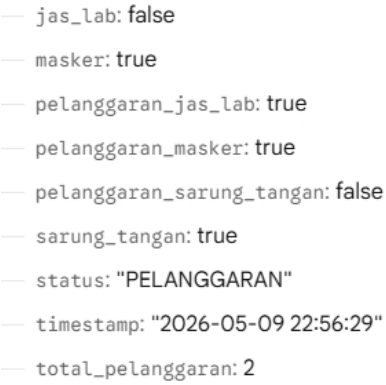

Figure 9. Violation data structure in the Firebase Realtime Database.

Overall System Testing and Discussion

During end-to-end testing, the entire functional chain operated according to the intended flow: video acquisition, YOLOv8s inference, compliance validation, LCD display, audio notification, data transmission to Firebase, and PWA updating. When PPE was complete, the system displayed the appropriate status and played the entry-permission audio. When a violation occurred, the system displayed the incomplete attributes, played the corresponding audio, and recorded the event in the PWA monitoring interface.

The test results show that model performance and system performance should be distinguished. The 89% model accuracy reflects classification capability on the evaluation data, whereas the 95% system success rate on 40 samples reflects successful decisions in the tested operational scenario. This difference indicates that temporal validation logic, the glove/no_glove trigger, and input-output integration help transform frame-level inference results into more stable system decisions.

The addition of the no_glove and no_labcoat classes was the most important change from the initial design. These additional classes did not merely increase the number of labels; they changed the problem from simply detecting the presence of PPE to distinguishing between PPE-wearing and non-PPE-wearing conditions. Testing showed that these additions helped reduce classification errors, although no_glove remained the class with the lowest AP@0.5. Because the remaining errors were related to lighting and visual similarity, expanding the negative-class dataset and maintaining consistent camera settings are more relevant improvements than simply increasing the confidence threshold.

From an edge-computing perspective, the Jetson Orin NX provided adequate computational margin for one user at a time. When the number of objects increased to three, GPU utilization rose sharply even though FPS remained around 13.5. This confirms that the system operating limit is determined not only by the model but also by the usage scenario. For a laboratory entrance, restricting the primary inspection to one subject is consistent with the system objective and keeps the decision process focused.

The main research limitations are sensitivity to distance and lighting, small object size, and the performance of the no_glove

class. In addition, direct PWA integration still depends on network connectivity, although the offline queue reduces its impact. Future research may investigate multi-camera configurations, model optimization such as quantization or pruning, and user identification mechanisms so that violation histories can be linked to practitioner identities [3].

This study successfully realized a computer vision-based PPE compliance detection system in a pharmaceutical laboratory using YOLOv8s and the NVIDIA Jetson Orin NX. The final implementation prioritized the actual requirements identified in Chapter IV of the thesis, namely a five-class detection system with the addition of no_glove and no_labcoat, temporal validation, audio notification, and monitoring through the Firebase Realtime Database and PWA. The model achieved 89% accuracy and 96.5% mAP@0.5, while operational testing produced a 95% detection success rate on 40 samples. The system operated for eight hours at 14–16 FPS without lag or crashes, with CPU temperature of 49.59–62.94°C, GPU temperature of 47.94–61.16°C, and power consumption of 5.7–9.2 W. The most effective camera working range was approximately 100–150 cm with illumination above 60 lux, while an actual placement of approximately 80–100 cm was selected for the entrance area. Audio notification had an average delay of approximately 0.31 seconds, and data synchronization to the PWA was 58–70 ms. Thus, the system meets the main functional requirements as a supporting medium for real-time PPE compliance monitoring in a pharmaceutical laboratory.

Confidence Analysis and Detection Thresholds

In addition to model evaluation metrics, the system was tested using confidence scores under different PPE-use conditions. Testing was performed while keeping the subject and camera positions fixed, followed by observation of the average FPS and confidence values of each successfully detected class. Average FPS was calculated from the latest FPS values stored by the system so that transient fluctuations had less influence on the result. This approach provides a more representative view of inference speed under operational conditions.

Under complete PPE conditions, the recorded confidence scores were 0.90 for mask, 0.91 for laboratory coat, and 0.87 for glove, with an average FPS of 14.06. When the laboratory coat and gloves were incomplete, FPS reached 14.83 with confidence values of 0.86 for mask, 0.85 for no_labcoat, and 0.80 for no_glove. When the mask and gloves were not used, FPS was 13.93 with confidence values of 0.94 for labcoat and 0.90 for no_glove. When the mask and laboratory coat were not used, FPS reached 14.90 with confidence values of 0.88 for no_labcoat and 0.88 for glove. These variations show that changes in PPE-attribute combinations did not cause a substantial decrease in FPS.

Table 11. FPS and Confidence on Selected PPE Conditions

Condition	Avg. FPS	Confidence	Interpretation
Complete PPE	14,06	Mask 0,90; Labcoat 0,91; Glove 0,87	All attributes detected
No glove + no labcoat	14,83	Mask 0,86; No_labcoat 0,85; No_glove 0,80	Negative classes detected
No mask + no glove	13,93	Labcoat 0,94; No_glove 0,90	Mask not detected as expected
No mask + no labcoat	14,90	No_labcoat 0,88; Glove 0,88	Violation combination detected

These results also illustrate the trade-off between confidence and the ability to capture objects. The F1-confidence curve shows a maximum F1 score of 0.94 at a confidence of 0.146 for all classes. Meanwhile, precision increases as the confidence threshold increases, while recall decreases. Because the operational system uses different confidence thresholds for different classes rather than relying on a single global value, the configuration is intended to preserve detection of small objects such as gloves while reducing false positives for laboratory-coat attributes.

The implication is that the threshold should not be selected solely based on the highest precision value. In a safety system, missing detection of PPE that is or is not worn can affect compliance decisions. Therefore, the final configuration considers evaluation-curve results, object size, visual characteristics, and field-test results. Temporal validation then serves as a second layer after confidence thresholding so that the final decision is not made from a single frame alone.

State Machine and Temporal Validation Analysis

The final implementation converts frame-by-frame model outputs into system decisions through a simple state machine. The system remains in an idle state when there is no user candidate requiring inspection. Glove or no_glove is used as the trigger because these attributes are located in the hand region and provide a sufficiently specific indication that a person is in the inspection position. After the trigger is activated, the system enters the evaluation phase, continuously monitors mask, labcoat, and glove, and waits for a stable condition before making the final decision.

A minimum three-second validation period is used to reduce the effect of transient predictions. In real-time video, a single frame may lose an object because of blur, minor occlusion, or position changes. If an attribute is not visible for more than 1.2 seconds, its timer is reset so that the system does not immediately classify the situation as a violation. The eight-second limit acts as a timeout so that the system does not wait indefinitely when a person does not satisfy the detection condition. The combination of the trigger, three-second validation, 1.2-second reset, and eight-second timeout forms a decision flow that is more appropriate for an entrance scenario than frame-by-frame decisions.

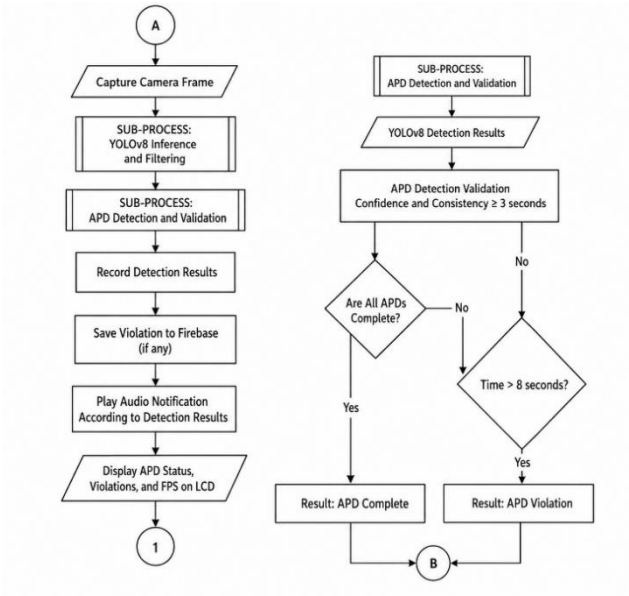

Figure 10. System Flowchart

From a software-implementation perspective, audio processing and database transmission run on separate threads. This approach prevents audio playback or network latency from stopping the inference loop. The offline queue also provides tolerance to network disruptions. Thus, local detection remains the primary function, while PWA synchronization is treated as a supporting function that may be delayed without changing the PPE decision currently being processed.

https://doi.org/10.25077/chipset.7.02.140-148.2026

Data Structure and Progressive Web Application Monitoring

The PWA is used not only to display the latest results but also as a historical evaluation interface. The monitoring page provides Today and All History modes. The main information includes total violations, violations on the current day, details of violated PPE types, violation trends, and event details. This structure allows personnel to review events without having to remain in front of the detection device.

Table 12. Data Fields Stored in Firebase

Field	Meaning	Use in PWA
masker	Mask use status	Compliance summary
jas_lab	Laboratory coat use status	Compliance summary
sarung_tangan	Glove use status	Compliance summary
pelanggaran_masker	Mask violation boolean	Violation-type classification
pelanggaran_jas_lab	Laboratory coat violation boolean	Violation-type classification
pelanggaran_sarung_tangan	Glove violation boolean	Violation-type classification
status	Final detection status	Event filtering
total_pelanggaran	Number of incomplete attributes	Statistics
timestamp	Event time	History and charts

Under normal network conditions, data from the Jetson Orin NX are updated to the PWA within 58–70 ms. This value is much smaller than the PPE decision interval and therefore does not become a bottleneck in monitoring. When the connection is interrupted, the offline queue retains the local data. Once the network becomes available again, the queue is transmitted automatically. This mechanism is important because the safety function should not stop simply because the remote monitoring channel is unavailable.

The structure also supports separation of detection and visualization functions. The Jetson Orin NX acts as an edge node that makes local decisions, while Firebase and the PWA function as logging and monitoring layers. This architecture reduces dependence of real-time decisions on the server, so network delays do not directly change the PPE status displayed on the local device.

Audio Notification Integration as a Human–Machine Interface

Audio notification was added after evaluation showed that an LCD-only display required users to focus their attention on the screen. The final implementation provides eight voice messages representing all relevant PPE completeness combinations, ranging from complete PPE to three incomplete attributes. Each audio file contains instructions without background music so that the information can be received clearly.

Table 13. Implemented Audio Scenarios

Condition	Audio Response
Complete PPE	"Complete PPE, please enter."
Mask incomplete	Instruction to wear a mask
Laboratory coat incomplete	Instruction to wear a laboratory coat
Glove incomplete	Instruction to wear gloves
Mask + laboratory coat incomplete	Instruction to wear a mask and laboratory coat
Mask + glove incomplete	Instruction to wear a mask and gloves
Laboratory coat + glove incomplete	Instruction to wear gloves and a laboratory coat
All PPE incomplete	Instruction to wear all three PPE items

Testing showed an audio delay of 0.26–0.41 seconds with an average of approximately 0.31 seconds. In addition to being fast, the notifications remained audible over a range of 0.5–3 m. Sound intensity decreased from 83 dB at 0.5 m to 64 dB at 3 m, but clarity was still considered adequate under the test-room conditions. Therefore, the audio channel provides useful redundant information when the user is not looking at the LCD.

Implementation Changes Compared with the Initial Design

The implementation changes are an important aspect that distinguishes this article from the design presented in the initial thesis chapters. In the initial design, the system was planned as a Smart Station integrating four functions: PPE detection, health screening, environmental monitoring, and drug-dimension measurement. In the final implementation, the latter two functions were developed as separate devices, while the system discussed in this article focuses on PPE detection and still shares the application environment with the drug-dimension measurement module through a mode-switching mechanism.

These changes directly affected the selection of the processing unit. The Jetson Orin NX was used because the final workload came not only from YOLOv8s but also from camera acquisition, LCD display, audio, and database communication. The numeric keypad replaced several separate buttons to reduce wiring and provide a mode-switching interface. The speaker was added to address the limitations of visual notification. The dataset changes also responded to implementation findings, particularly the need to distinguish white shirts from laboratory coats and bare hands from gloved hands.

Comparison with Previous Studies

Study [3] used YOLOv5 for PPE compliance detection in hazardous areas. This study follows a similar objective, namely the automation of PPE inspection, but the presented implementation uses YOLOv8s on the NVIDIA Jetson Orin NX edge device and explicitly adds negative classes. The no_glove and no_labcoat classes allow the model not only to search for the presence of positive PPE attributes but also to learn conditions in which PPE attributes are not worn.

Study [4] demonstrated the use of deep learning for real-time image processing. In this study, the real-time requirement is evaluated not only by the model's ability to detect objects but also by the stability of the entire system over eight hours, CPU/GPU/RAM usage, notifications, and data synchronization. This approach is important for laboratory devices because the system must operate as an actual instrument rather than merely as a model demonstration.

In the pharmaceutical field, machine-vision studies such as [5]–[7] show that visual inspection can support automation and quality control. The contribution of this study is more specific to personnel safety by placing PPE detection at the laboratory entrance and linking it with feedback mechanisms and violation logging. Thus, the practical novelty lies in integrating edge inference, temporal decision logic, audio feedback, and PWA monitoring in an educational pharmaceutical-laboratory context.

Although the test results show good performance, this study does not provide a direct comparison with YOLOv5, YOLOv7, or other YOLO models on the same dataset. Therefore, the study does not claim that YOLOv8s is universally superior; instead, its selection is justified by its suitability for the implementation requirements and the system test results obtained. Direct cross-model comparison is a potential future study that could provide a more experimental justification for model selection.

Research Limitations and Development Implications

The first limitation is dependence on the camera field of view. At 50 cm, the user is too close for the laboratory coat to be captured completely; at 250–300 cm, small objects such as masks and gloves lose detail. The second limitation is lighting. Although the system still succeeds at 20 lux, image quality is better above 60 lux and is more consistent at 96–131 lux. This indicates that physical camera installation and the environment remain part of the system design rather than merely secondary parameters.

The third limitation is the performance of the no_glove class. Its AP@0.5 value of 0.875 is lower than that of the other classes, and the confusion matrix shows that some false negatives occur in this class. This may be influenced by the relatively small size of glove objects and their visual similarity to bare hands. A larger dataset covering hand conditions, pose variations, glove colors, and mild occlusion could improve accuracy in future research.

For further development, quantization or pruning can be applied to reduce computational load and then evaluated on lower-specification devices such as the Jetson Nano or other edge accelerators. Multiple cameras can be used to reduce blind spots and cover more than one entrance direction. User identification through RFID, QR Code, or other methods can also be added so that violations can be directly associated with practitioner identities. However, these developments should preserve the principle that the primary PPE decision can still be made locally so that the system does not depend on Internet connectivity.

Table 14. Comparison of Initial Design and Final System Implementation

Aspect	Initial Design	Final Implementation	Reason
Processor	Jetson Nano	Jetson Orin NX	Inference, video, audio, and Firebase load
Audio output	Visual only	LCD + speaker	Feedback without requiring the user to look at the screen
Input mode	Several push buttons	Mini numeric keypad	Simpler and more flexible wiring
Dataset	3 classes, then expanded	5 final classes	Reduces ambiguity between no_glove/no_lab coat
Detection logic	Basic PPE detection	Trigger + temporal validation	Reduces decisions made when a user is merely passing by
Data monitoring	Local storage	Firebase + PWA + offline queue	Remote monitoring and network tolerance

System Acceptance Criteria

In addition to model-metric testing, the implementation results should be viewed in terms of satisfying system acceptance criteria. Each major function was not only tested separately but also evaluated in relation to the operational objectives at the laboratory entrance. This approach ensures that research success does not stop at the model-accuracy value but includes the device's ability to execute the decision flow, provide feedback, and retain detection results.

Table 15. Mapping of Final Requirements and Test Results

Requirement	Acceptance Indicator	Final Result
Image acquisition and detection	Real-time frames; five classes can be processed	Met; YOLOv8s runs at 14–16 FPS
PPE status determination	Complete/violation decision matches ground truth	Met; 38/40 samples correct
Visual output	LCD displays bounding boxes and status	Met; detection results are displayed in real time
Audio output	Message matches condition and responds quickly	Met; 8 scenarios, average delay 0.31 s
Continuous operation	8 hours without failure	Met; no lag/crash
Remote monitoring	Data synchronized to PWA	Met; 58–70 ms, offline queue active

This consistency shows that the system is an integration of several mutually supporting layers. YOLOv8s produces object information, the state machine converts object information into decisions, the local interface communicates those decisions, and Firebase/PWA extends them into monitorable information. If any one layer fails, the practical value of the system may decrease even when the model itself has high individual metrics.

From an implementation perspective, the tests also show the importance of defining operational limits. The system is not intended to handle every position and every distance, but is designed for a user standing in front of the camera at approximately 80–100 cm. This limit is close to the test range that produced stable results, namely 100–150 cm, while providing room for all attributes to remain within the field of view during inspection. Establishing such operational limits is a form of systems engineering that converts condition-dependent model performance into a controllable usage procedure.

Lighting conditions should also be treated as part of the installation criteria. The model can still detect PPE at around 20 lux, but results are more consistent above 60 lux. Therefore, system installation in the laboratory should consider light sources and camera position to avoid backlighting, excessive shadows, or overly dark objects. The 96–131 lux results measured in the laboratory provide empirical evidence that the actual implementation environment supports the detection process.

Finally, the system has two success paths: a local path and a monitoring path. The local path must continue operating whether or not an Internet connection is available, while the monitoring path may be delayed without stopping detection. This separation was demonstrated through network testing, in which data could still be recorded temporarily when the connection was lost and synchronized when the network returned. Therefore, the edge-first architecture applied in this study is consistent with the needs of safety devices that must preserve core functionality even when network services are disrupted.

CONCLUSIONS

Based on the implementation and testing results, the computer vision-based PPE compliance detection system using YOLOv8s on the NVIDIA Jetson Orin NX successfully detects masks, laboratory coats, and gloves in real time. The addition of the no_glove and no_labcoat classes enables the system to distinguish PPE-use conditions from visually similar objects. The model achieved 89% accuracy and 96.5% mAP@0.5, while testing on 40 samples achieved a 95% detection success rate. The system also successfully determines PPE completeness and provides notifications through the LCD and speaker according to the detection results.

Operationally, the system can run continuously for 8 hours without lag or crashes, with FPS of 14–16, CPU temperature of 49.59–62.94 °C, GPU temperature of 47.94–61.16 °C, and power consumption of 5.7–9.2 W. Integration with the Firebase Realtime Database and Progressive Web Application (PWA) was also successfully implemented, with data update times of approximately 58–70 ms and temporary storage when the Internet connection is interrupted. These results indicate that the system meets the requirements for real-time PPE detection and monitoring and is suitable as a supporting medium for PPE compliance supervision in a pharmaceutical laboratory.

ACKNOWLEDGMENT

The authors would like to thank the Department of Computer Engineering, Faculty of Information Technology, Andalas University, as well as the supervisors who provided guidance throughout the system development and evaluation. The authors

https://doi.org/10.25077/chipset.7.02.140-148.2026

also thank the Pharmaceutical Laboratory for supporting the system implementation testing process.

REFERENCES

[1] World Health Organization, Laboratory Biosafety Manual, 4th ed. Geneva, Switzerland: WHO, 2020.
[2] M. Arisandi, "Peran Pengetahuan dan Sikap terhadap Kepatuhan Penggunaan Alat Pelindung Diri pada Mahasiswa Farmasi: Studi Cross-Sectional," Babul Ilmi J. Ilm. Multi Sci. Kesehatan, vol. 17, no. 2, 2025.
[3] A. R. Pratama, D. P. Kartikasari, and A. Setiawan, "Implementasi Deteksi Kelengkapan Alat Pelindung Diri (APD) di Area Berbahaya Menggunakan Metode YOLOv5," J. Elektronika, Kendali dan Industri (ELKOLIND), vol. 10, no. 2, pp. 85–94, 2024.
[4] P. A. Khairunnisa, "Sistem Pengolahan Citra Digital Untuk Mendeteksi Ekspresi Wajah Secara Real-Time Menggunakan Deep Learning YOLOv5," J. Ilm. Res. Student, vol. 2, no. 1, pp. 594–607, 2025.
[5] A. Diószegi et al., "Automated tablet defect detection and the prediction of disintegration time using machine vision," Int. J. Pharmaceutics, vol. 645, art. 123420, 2024.
[6] M. S. Narlawar et al., "Precision measurement of tablet characteristics: A digital device approach," SSRG Int. J. Electr. Electron. Eng., vol. 11, no. 7, pp. 283–291, 2024.
[7] A. Vijayakumar et al., "Real-time visual intelligence for defect detection in pharmaceutical packaging using deep learning," Sci. Rep., vol. 14, art. 15832, 2024.
[8] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, vol. 25, no. 11, pp. 120–125, 2000.
[9] Fifine Microphone, "K420 2K Webcam User Manual & Specifications," Fifine Technical Support, 2022.
[10] Badan Standardisasi Nasional, SNI 03-6575-2001: Tata Cara Perancangan Sistem Pencahayaan Buatan pada Bangunan Gedung. Jakarta, Indonesia: BSN, 2020.

AUTHOR(S) BIOGRAPHY



Abil Rahman
is a student in the Department of Computer Engineering, Faculty of Information Technology, Andalas University, Padang, Indonesia. His research interests include embedded systems, Internet of Things (IoT), and computer vision. His research focuses on the application of computer vision and embedded systems to support automation, monitoring, and safety improvement in laboratory environments.