Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 39 (June 2025) PP. 169-184
DOI: 10.35631/JISTM.1039011


JOURNAL OF INFORMATION SYSTEM AND TECHNOLOGY MANAGEMENT (JISTM)
www.jistm.com

GLOBAL ACADEMIC EXCELLENCE

# AUTOMATED SMOKING DETECTION IN RESTRICTED ZONES: A YOLOv8-BASED APPROACH WITH GENDER IDENTIFICATION

Muhammad Ismail[1], Ahmad Shakaff Ali Yeon[2*], Syed Muhammad Mamduh Syed Zakaria[3], Ammar Zakaria[4], Latifah Munirah Kamarudin[5], Retnam Visvanathan[6], Muhammad Reza Zainal Abidin[7], Norbazlan Mohd Yusof[8], Muhammad Azri A Baharom[9], Ahmad Helmi M Amin[10], Hafizal Khalid[11]

[1] Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis
Email: muhdismail@studentmail.unimap.edu.my
[2] Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis
Email: ahmadshakaff@unimap.edu.my
[3] Faculty of Electronic Engineering & Technology, Universiti Malaysia Perlis
Email: smmamduh@unimap.edu.my
[4] Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis
Email: ammarzakaria@unimap.edu.my
[5] Faculty of Electronic Engineering & Technology, Universiti Malaysia Perlis
Email: latifahmunirah@unimap.edu.my
[6] Ideria Sdn. Bhd.
Email: retnam@ideria.co
[7] Innovation & Centre of Excellence PLUS Berhad
Email: reza.zainalabidin@plus.com.my
[8] Innovation & Centre of Excellence PLUS Berhad
Email: bazlan@plus.com.my
[9] Project Delivery and Technology, PETRONAS
Email: muhammadazri@petronas.com
[10] Project Excellence, PETRONAS
Email: helmi.mamin@petronas.com
[11] Project Excellence, PETRONAS
Email: hafizal_khalid@petronas.com
[*] Corresponding Author

**Article Info:**

**Article history:**
Received date: 31.03.2025
Revised date: 27.04.2025
Accepted date: 29.05.2025

**Abstract:**

This study investigates the application of YOLOv8 for smoking detection, utilizing a custom dataset and evaluating different optimization parameters, including optimizers AdamW, Adam, and SGD, with learning rates of 0.1, 0.01, and 0.001. The evaluation highlights that the Adam optimizer with a learning rate of 0.001 delivers the best overall performance, achieving an mAP@50-95 of 0.713, along with high precision (93.1%) and recall (85.9%).

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 39 (June 2025) PP. 169-184
DOI: 10.35631/JISTM.1039011

Published date: 20.06.2025

**To cite this document:**

Ismail, M., Yeon, A. S. A., Syed Zakaria, S. M. M., Zakaria, A., Kamarudin, L. M., Visvanathan, R., Abidin, M. R. Z., Yusof, N. M., Baharom, M. A. A., Amin, A. H. M., & Khalid, H. (2025). Automated Smoking Detection In Restricted Zones: A YOLOv8-Based Approach With Gender Identification. *Journal of Information System and Technology Management, 10* (39), 169-184.

**DOI:** 10.35631/JISTM.1039011



In comparison, AdamW at the same learning rate shows slightly higher precision (93.9%) but marginally lower recall (86.4%), resulting in similar mAP values. On the other hand, SGD performs well in some cases but falls short overall, particularly at higher learning rates, which lead to unstable convergence and reduced accuracy. While the model exhibits robust detection capabilities in close-range, low-angle scenarios, it struggles with high-angle or long-distance conditions due to dataset diversity limitations. This research contributes to restricted-zone monitoring systems by offering insights into real-time behavioral tracking and informing future applications in public health and policy enforcement. The findings emphasize Adam's superiority in balancing precision, recall, and stability, providing an effective framework for real-time smoking detection systems. Future work should focus on diversifying and augmenting the dataset to enhance generalization across challenging conditions.

**Keywords:**

YOLO, Object Detection, Smoking Detection, Demographic

## Introduction

In high-risk environments such as oil and gas facilities and public spaces, the enforcement of smoking restriction zones is essential to ensure both safety and regulatory compliance. Traditional monitoring techniques—often reliant on manual surveillance—are inefficient and prone to errors, leading to inconsistent enforcement of no-smoking regulations. As smoking in restricted areas remains a significant public health and safety concern, especially in densely populated or hazardous zones, the need for automated, intelligent monitoring systems is more urgent than ever (Li & Liu, 2024; Ren, 2024).

Recent advances in deep learning and computer vision, particularly using You Only Look Once (YOLO) frameworks, have enabled the development of real-time, accurate smoking detection systems. These systems are capable of recognizing smoking activities and individuals with high precision, offering promising alternatives to manual monitoring (Zutshi et al., 2024; Indra et al., 2024). Despite these innovations, challenges remain, such as demographic profiling, smoke source disambiguation, and real-time data integration (Harshavardhan & Upender, 2025; Ma et al., 2022).

Studies have shown that finetuned YOLOv8-based models can achieve robust detection of cigarettes and smoking gestures in complex environments (Li & Liu, 2024). In urban deployments, hybrid systems combining gas sensors and computer vision have been implemented to enhance detection accuracy and trigger responsive alarms in real-time (Harshavardhan & Upender, 2025). Furthermore, the integration of demographic analysis—such as gender classification—within smoking detection frameworks has improved the granularity of behavioral insights, aiding enforcement and health policy development (Pandey et al., 2024).

This study introduces a custom YOLOv8-based automated system for smoking detection, integrating object recognition, demographic classification, and activity logging. It detects not only cigarettes and smoking-related gestures but also tracks individuals over time using advanced object tracking algorithms. Metadata such as timestamps and detected behaviors are

Journal of Information System and Technology Management JISTM
EISSN: 0128-1666
Volume 10 Issue 39 (June 2025) PP. 169-184
DOI: 10.35631/JISTM.1039011
recorded for regulatory analysis. To avoid data redundancy, a cooldown mechanism is incorporated, preventing excessive logging from the same individual. The proposed system addresses both public health and industrial safety needs by enhancing surveillance capabilities in smoke-free zones, thereby supporting smarter urban management and occupational safety compliance.

## Literature Review

### *Smoking Detection Challenges In Outdoor Environments*
Outdoor environments pose significant challenges for smoking detection due to fluctuating lighting, complex backgrounds, and adverse weather conditions. For example, Lavu et al. (2024) emphasized the transient nature of smoke in real-world settings and proposed dual spectral fusion with YOLO-based models to enhance visibility and classification accuracy. These preprocessing techniques are crucial for boosting system robustness under dynamic outdoor scenarios.

### *Camera-Base Smoking Detection*
Camera-based systems are increasingly used for detecting smoking behaviors, employing various modalities such as RGB and infrared sensors. Indra et al. (2024) developed a YOLO-driven RGB camera system to differentiate smoking gestures from other hand movements in diverse urban settings. Building on these advances, Wang et al. (2023) proposed a YOLOv8-MNC framework that enhances detection precision in visually complex environments. Additionally, Wang et al. (2024) introduced a specialized YOLOv8 model tailored for safety monitoring in chemical plants, demonstrating effectiveness in detecting smoking behavior under challenging industrial conditions.

### *Deep Learning-Based Smoking Detection*
Deep learning models have demonstrated remarkable accuracy in detecting smoking activities, leveraging their ability to learn intricate patterns in data.

#### *Convolutional Neural Networks (CNNs)*
CNNs are widely employed in image-based smoking detection due to their strong feature extraction capabilities. Ma et al. (2022) introduced a CNN-enhanced YOLO framework that demonstrated high precision in detecting cigarettes in cluttered outdoor scenes, underscoring the adaptability of deep learning in real-time applications.

#### *Recurrent Neural Networks (RNNs)*
Temporal modeling with RNNs, particularly Long Short-Term Memory (LSTM) networks, has also shown promise. Pandey et al. (2024) leveraged an LSTM-based system integrated with YOLO-NAS to capture time-dependent smoking cues such as lighting a cigarette or prolonged gestures, resulting in improved contextual accuracy.

#### *YOLO Framework*
YOLO-based models, known for their efficiency and real-time detection capabilities, have proven valuable for identifying smoking behaviors in surveillance environments. Salau et al. (2024) integrated YOLOv5 in a monitoring system to detect smoking and violent acts in hostel environments, demonstrating robust accuracy in crowded conditions. Advancing this, Wang et al. (2023) introduced the YOLOv8-MNC framework, which improved feature representation

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 39 (June 2025) PP. 169-184
DOI: 10.35631/JISTM.1039011
and enhanced detection performance in visually cluttered scenes. Wang et al. (2024) further adapted YOLOv8 to industrial contexts, offering specialized detection for workplace smoking violations.

### *YOLOv8 in Smoking Detection*
YOLOv8's architectural improvements—such as its anchor-free detection mechanism and deeper feature extraction—make it highly suitable for detecting complex human behaviors like smoking. Wang et al. (2023) demonstrated the effectiveness of the YOLOv8-MNC model in identifying smokers in visually cluttered environments. Zhang and Jiang (2025) introduced DAHD-YOLO, an optimized YOLOv8-based framework designed for robust real-time smoking detection in public surveillance systems, showcasing improvements in multi-class classification and reduced false positives.

### *Tracking Algorithms in Smoking Detection*
Tracking algorithms play a vital role in following smoking activities over time, ensuring continuous monitoring and data analysis.

#### *BotSort Algorithm*
Although no study directly evaluated BotSort in smoking-specific contexts, recent YOLOv8 tracking implementations favor DeepSORT or Transformer-based alternatives. For example, Li et al. (2024) proposed a novel neck-enhanced YOLOv8 variant, paired with multi-class deep SORT, to detect smoking and phone use behaviors in traffic monitoring, confirming the growing importance of robust multi-object tracking in behavior classification systems.

#### *Simple Online and Realtime Tracking (SORT)*
SORT provides lightweight, fast tracking for real-time systems but lacks the ability to differentiate between multiple classes. As highlighted by Li and Liu (2024), this limitation constrains its effectiveness in multi-class environments where classification-specific monitoring is required.

This review highlights the advancements in automated smoking detection, emphasizing the challenges of outdoor detection and the effectiveness of camera-based systems. The evolution of deep learning techniques, particularly YOLOv8, has significantly enhanced detection accuracy. Tracking algorithms such as BotSort and SORT provide valuable tools for continuous monitoring but face limitations in class-specific tracking. Integrating YOLOv8 with BotSort offers a promising solution for robust smoking detection and tracking in real-world applications.

## Methodology

### *System Design Overview*
The proposed smoking detection system integrates both hardware and software components into a unified architecture designed for real-time surveillance, behavioral analysis, and data logging. The system runs on a high-performance hardware configuration that includes an NVIDIA GeForce RTX 2080 Ti GPU, an AMD Ryzen 5 7500F processor, and 32 GB of RAM. It operates on Windows 11 Home and uses Python 3.9 for executing the deep learning pipeline. This setup was selected to meet the computational demands of object detection and tracking models during continuous video processing. The surveillance network comprises seven CCTV

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 39 (June 2025) PP. 169-184
DOI: 10.35631/JISTM.1039011
cameras that stream video using the RTSP protocol. These video feeds are received by a central server, which processes the data in real time, identifying activities related to smoking behavior and tracking individuals across frames. The system incorporates a YOLOv8-based detection model to identify objects of interest, and BotSort as a multi-object tracking algorithm to ensure that detected individuals are persistently followed throughout the video sequence. This configuration allows for the seamless capture of not only spatial data but also temporal behavioral patterns, providing a scalable and efficient framework for monitoring environments such as public areas or industrial sites where smoking is prohibited.

From a software standpoint, the system leverages a custom-trained YOLOv8 model tailored specifically for detecting smoking-related behaviors such as hand-to-mouth movements and visible cigarettes. The BotSort tracking algorithm further enhances this by providing continuity across frames, enabling the system to distinguish between individuals and follow their behavior over time. This integrated approach to object detection and tracking makes it possible to generate rich metadata, including frame timestamps and object IDs, which are logged for compliance verification and future analysis. The software design ensures both real-time responsiveness and data archival, supporting both immediate enforcement actions and retrospective auditing.

***YOLOv8 Training***

The training of the YOLOv8 model began with the initialization of pre-trained weights from the yolov8n.pt checkpoint, selected for its balance between model size and inference speed. The model architecture was defined using the yolov8n.yaml configuration file, which specifies the network layers and connection schema. Transfer learning was applied to adapt the general object detection capabilities of YOLOv8 to the domain-specific task of smoking detection. The dataset used for training consisted of manually and publicly sourced images annotated with bounding boxes for cigarettes, hand positions, and related features. The training process was configured to run for 70 epochs with a batch size of 32, using two worker threads for data loading to maintain high throughput. To optimize performance, the training routine experimented with several optimization algorithms, including Stochastic Gradient Descent (SGD), Adam, and AdamW. Three learning rates—0.001, 0.01, and 0.1—were also tested to evaluate the model's sensitivity to learning rate adjustments. This combination of hyperparameter tuning and pre-trained weight initialization allowed the model to converge efficiently, minimizing training time while maximizing accuracy.

Once trained, the model was deployed in conjunction with BotSort for live video inference. BotSort, which merges the best features of ByteTrack and SORT, was selected for its ability to handle object occlusions and maintain identity tracking in crowded scenes. Although BotSort does not natively support class-specific tracking, its class-agnostic framework was sufficient for the study’s goal of monitoring smoking behavior in public settings. As noted by Zang et al. (2024), such integration provides a comprehensive solution for surveillance applications requiring both detection accuracy and persistent object tracking. Furthermore, the combination of YOLOv8 and BotSort has proven effective in real-world deployments such as industrial safety and urban health monitoring (Wang et al., 2024).

***BotSort Tracking Algorithm***

BoT-SORT (Bidirectional One-Shot Tracker with Camera Motion Compensation and Re-Identification) represents a significant advancement in Multi-Object Tracking (MOT), offering improvements especially relevant to domains like autonomous systems, urban surveillance, and behavioral analysis. Traditionally, MOT systems have adhered to the tracking-by-detection paradigm, where objects are first detected in each frame and subsequently linked across frames by tracking algorithms. A common approach in classical MOT pipelines has been the use of Kalman Filters for motion prediction, coupled with simple heuristics like the Hungarian algorithm for data association (Alikhanov & Kim, 2023). While efficient, such approaches often falter under challenging conditions such as occlusion, fast motion, or camera displacement.

In recent years, the field has shifted toward deep learning-based models that incorporate Convolutional Neural Networks (CNNs) for robust appearance modeling. These models extract discriminative features from detected objects to match identities across frames using similarity metrics, thereby improving data association (Huang et al., 2024). Nevertheless, systems such as SORT and DeepSORT still suffer from performance drops in complex environments, particularly due to their limited capability to handle rapid object motion or inconsistencies introduced by moving cameras (Tu et al., 2024).

BoT-SORT addresses these limitations through three core innovations. First, it introduces Camera Motion Compensation (CMC) to correct for background shifts caused by moving cameras, significantly enhancing the accuracy of bounding box localization. This is particularly valuable for mobile surveillance systems and drone-based monitoring (Mahdian et al., 2024). Second, BoT-SORT incorporates Re-Identification (ReID) modules using CNNs to encode appearance features, allowing for more precise identity tracking even under occlusion or re-entry scenarios. Lastly, it fuses traditional Intersection over Union (IoU) metrics with ReID embeddings to optimize data association, achieving a balance between Multiple Object Tracking Accuracy (MOTA) and Identity F1 Score (IDF1), both crucial indicators of performance in MOT benchmarks (Zhu et al., 2025).

These architectural improvements make BoT-SORT well-suited for applications requiring continuous and accurate identity tracking in cluttered and dynamic environments. Although its class-agnostic nature can present limitations when class-specific tracking is desired, it remains a powerful framework that outperforms many existing systems under real-world constraints (Rathore et al., 2024). Its modular design also supports seamless integration with object detection systems such as YOLOv8, enhancing its practical utility across diverse surveillance scenarios.

***Data Collection***

To ensure robustness and real-world applicability, the dataset used in this study was constructed through a hybrid approach that combined publicly available annotated datasets with manually curated image data. Public datasets were sourced from platforms such as Kaggle and Roboflow, which provide well-structured, labeled image datasets in formats including YOLOv3, COCO JSON, and Pascal VOC XML. These repositories offered a reliable foundation for training and benchmarking deep learning models in object detection tasks (Ahmad et al., 2021).

To increase variability in environmental conditions and subject appearances, additional video footage was acquired from internal archives and publicly accessible online sources—most notably, YouTube channels with content licensed for reuse. The video streams were processed using OpenCV in conjunction with FFmpeg, enabling efficient frame extraction while preserving temporal continuity This process allowed the creation of a high-resolution image

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 39 (June 2025) PP. 169-184
DOI: 10.35631/JISTM.1039011
set that captured diverse lighting, background, and behavioral conditions essential for robust model training.

Following extraction, all image frames were annotated using LabelImg, a widely adopted open-source annotation tool that facilitates bounding box creation and multi-class labeling (Tzutalin, 2015; Lavu et al., 2024). Annotation focused on key visual features such as cigarette presence, facial visibility for identity analysis, and visual gender indicators to support demographic profiling. These annotations were critical for training object detection models capable of real-time behavioral analysis and demographic classification.

The resulting dataset was partitioned into training (70%), validation (20%), and testing (10%) sets, ensuring representative distribution across sources and classes. This stratified sampling strategy aimed to reduce overfitting and enhance generalizability on unseen data—a common best practice in computer vision model development (Sivaraman et al., 2023). The diversity of this dataset, combined with rigorous annotation and balanced partitioning, forms a comprehensive foundation for evaluating smoking detection systems under realistic conditions. The final composition of the dataset is presented in Table 1, reflecting both the heterogeneity of the data sources and the rigor of the annotation process.

**Table 1: Distribution Of Class Labels And Their Occurrences**

| Class Label | Occurrences |
|---|---|
| Cigarette | 3195 |
| Face_F | 1507 |
| Smoking | 1824 |
| Female | 1541 |
| Face_M | 2591 |
| Male | 2637 |

***System Workflow***

The system employs a YOLOv8 custom model trained to detect multiple classes, including persons, smoking-related objects such as cigarettes, and gender-specific features. Each detected object is assigned a unique ID, with bounding boxes drawn around them to classify objects and their respective activities. To maintain continuity in identifying and analyzing subjects across frames, the BotSort tracking algorithm assigns a persistent ID to each tracked individual. To ensure reliable detections, the system incorporates a stability threshold requiring multiple consistent detections, typically three within one second, to avoid false positives caused by transient detections such as momentary occlusions.

Smoking activity is detected by analyzing overlaps between a person's bounding box and smoking-related objects. If the system confirms a stable detection, both face and full-body images are captured, accompanied by metadata such as timestamps, gender, and activity status. The system also captures images of individuals in non-smoking states to create a balanced dataset for further analysis or training. To minimize redundancy, a cooldown period of one second is applied between successive captures of the same individual in a specific state, ensuring efficient storage usage while avoiding duplicate captures. Gender classification is

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 39 (June 2025) PP. 169-184
DOI: 10.35631/JISTM.1039011
determined by achieving at least 70% agreement among recent frames, while smoking activity is validated similarly to ensure reliability.

Image capture conditions include consistent overlaps between smoking-related objects and a person's bounding box for at least three consecutive detections, with at least 70% of recent detections confirming the activity. For non-smoking individuals, the system captures images only if the cooldown period has elapsed since the last capture. Faces detected within a person's bounding box are also captured alongside full-body images, providing comprehensive documentation. The system processes captured images in real-time, annotating them with details such as track ID, gender, activity status, and timestamps for contextual analysis.

To enhance accuracy, the system dynamically adjusts bounding box sizes for specific classes and filters detections outside the stability window, minimizing redundant processing. Captured images are resized and annotated for clarity, ensuring they meet minimum resolution requirements. By integrating YOLOv8 and BotSort, the system offers a robust framework for detecting smoking activities and analyzing gender in public spaces. The inclusion of stability checks and cooldown mechanisms ensures reliability while optimizing performance, making it an effective tool for behavioural analysis and regulatory compliance.

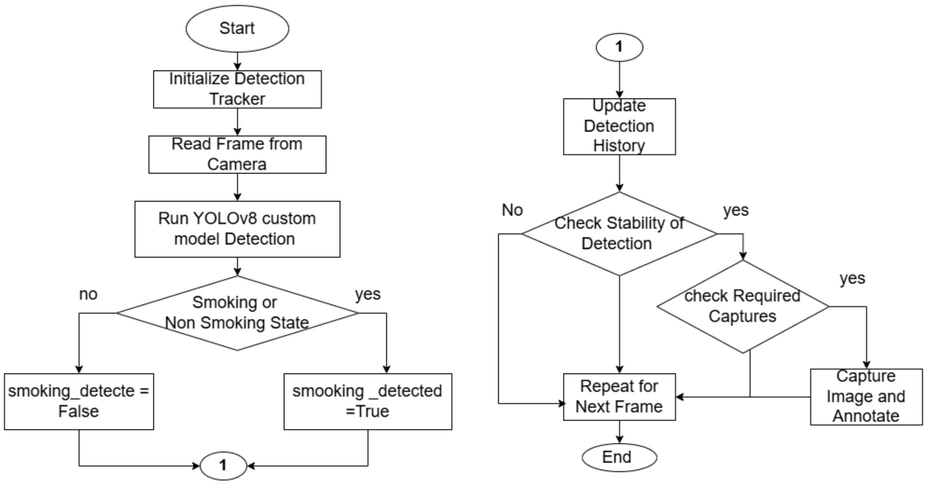


**Figure 1: Flowchart System Work**

***Result and Discussion***

The experiments aimed to identify the optimal training configuration for a system designed to detect and capture individuals smoking. The evaluation involved varying learning rates ([0.001, 0.01, 0.1]) and optimization algorithms (SGD, Adam, AdamW), with metrics including mean Average Precision (mAP@50, mAP@50-95), precision, recall, and training time. The results are summarized in table 2, which provides insights into the performance of these configurations

**Table 2: Experimental Results For Different Learning Rates And Optimizers.**

| Learning_rate | Optimizer | mAP50 | mAP50-95 | precision | recall | training_time(min) |
|---|---|---|---|---|---|---|
| 0.001 | SGD | 0.892 | 0.668 | 0.915 | 0.830 | 23.320 |
| 0.001 | Adam | 0.922 | 0.713 | 0.931 | 0.859 | 23.743 |
| 0.001 | AdamW | 0.923 | 0.714 | 0.939 | 0.864 | 23.815 |
| 0.01 | SGD | 0.920 | 0.712 | 0.924 | 0.862 | 23.741 |

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 39 (June 2025) PP. 169-184
DOI: 10.35631/JISTM.1039011

| | | | | | | |
|---|---|---|---|---|---|---|
| 0.01 | Adam | 0.901 | 0.675 | 0.925 | 0.827 | 23.849 |
| 0.01 | AdamW | 0.902 | 0.679 | 0.913 | 0.849 | 23.798 |
| 0.1 | SGD | 0.907 | 0.688 | 0.918 | 0.846 | 23.813 |
| 0.1 | Adam | 0.702 | 0.456 | 0.704 | 0.649 | 24.307 |
| 0.1 | AdamW | 0.284 | 0.138 | 0.334 | 0.357 | 23.707 |

At a learning rate of 0.001, the AdamW optimizer exhibited the best performance, achieving a high mAP@50 of 0.923, mAP@50-95 of 0.714, precision of 0.939, and recall of 0.864. This makes it the optimal setting for the smoking detection system. Training time for this configuration was comparable to other settings, making it efficient and accurate. By contrast, the SGD optimizer, although relatively efficient, lagged slightly behind in mAP metrics with a score of 0.892 for mAP@50.

Increasing the learning rate to 0.01 showed consistent performance for the SGD optimizer, maintaining a respectable mAP@50 of 0.920, along with strong precision and recall values. However, the performance of Adam and AdamW dropped slightly in comparison to their results at 0.001. The results suggest that both optimizers are sensitive to higher learning rates. At the highest learning rate of 0.1, all optimizers experienced degraded performance, particularly Adam and AdamW. For instance, the mAP@50 for Adam dropped significantly to 0.702, and AdamW demonstrated the worst results overall, with an mAP@50 of only 0.284 and an mAP@50-95 of 0.138. These findings highlight that higher learning rates can lead to unstable training or ineffective optimization, especially for adaptive algorithms.

The analysis indicates that the AdamW optimizer with a learning rate of 0.001 provides the best balance between precision, recall, and mAP metrics, making it the most suitable choice for deploying the smoking detection system. On the other hand, configurations using AdamW at a learning rate of 0.1 yielded the poorest results and should be avoided. Future implementations should consider the effects of learning rate and optimizer choice on system stability and performance.

The AdamW optimizer with a learning rate of 0.001 has proven to be the best-performing configuration for the smoking detection system, as evident from both the numerical metrics and the confusion matrix analysis as shown in figure 2. This setup achieves high accuracy across most categories, with notable strengths in detecting male and female classifications at 96% and 94% accuracy, respectively. Misclassifications in these categories are minimal, indicating the model's robust ability to distinguish gender. Similarly, the system performs well in face detection, achieving 94% accuracy for male faces (Face_M) and 92% for female faces (Face_F). However, some misclassification between these categories and the background suggests minor room for improvement in distinguishing faces from other elements in the scene.

In smoking detection, the model achieves an accuracy of 84%, demonstrating its effectiveness in identifying smoking behaviors. However, there is notable confusion with background elements (20%) and other categories such as male (14%), indicating the system's sensitivity to noise and overlapping features. Cigarette detection, on the other hand, presents more significant challenges, with a lower accuracy of 72% and a high rate of misclassification as background (44%). This suggests that the system struggles to detect small or partially visible objects like cigarettes, potentially due to limited feature representation or insufficient data diversity.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 39 (June 2025) PP. 169-184
DOI: 10.35631/JISTM.1039011

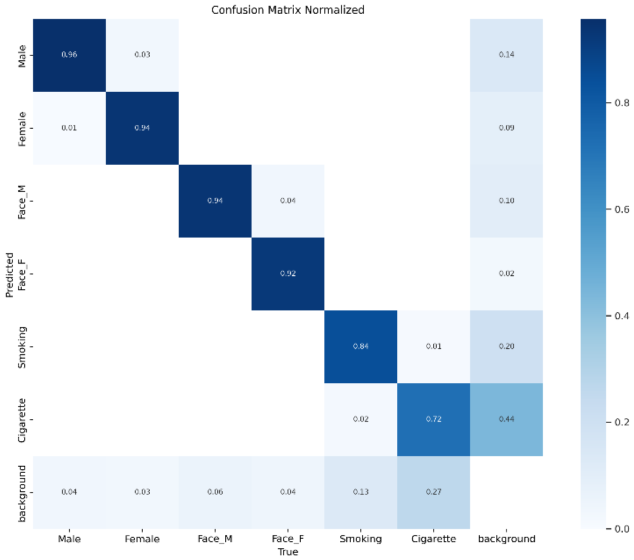


**Figure 2: Confusion Matrix Normalized For Training And Validation Dataset Optimizer Adamw Learning Rate 0.001**

Figure 3 illustrates the training and validation performance of the YOLOv8 model optimized with AdamW and a learning rate of 0.001 using a custom dataset. The training curves for box loss, classification loss, and DFL loss demonstrate a consistent downward trend over 70 epochs, indicating effective convergence during training. Notably, the box loss starts around 1.4 and decreases steadily, reaching values below 0.8 by the end of training, suggesting that the model effectively improves its ability to localize bounding boxes. Similarly, classification loss and DFL loss show parallel declines, underscoring the model's growing proficiency in categorizing and refining object predictions.

The metrics for precision, recall, mAP@50, and mAP@50-95 reveal progressive improvements across epochs. Precision and recall begin at moderate levels and rise steadily, stabilizing above 0.9 and 0.85, respectively, by the end of training. This reflects the model's ability to both accurately predict positive cases and minimize false negatives. The mAP@50 metric sees a sharp increase in early epochs, stabilizing around 0.92, while mAP@50-95 follows a similar trajectory, reaching approximately 0.71. These trends validate the model's robustness and generalization capabilities when applied to diverse data distributions within the custom dataset.

Overall, the AdamW optimizer with a 0.001 learning rate ensures smooth and stable optimization, allowing the YOLOv8 model to achieve high accuracy and reliability. The alignment of training and validation losses further confirms the absence of significant overfitting, making this configuration an ideal choice for deploying the smoking detection system in real-world scenarios. Future improvements could focus on further reducing losses and enhancing precision in challenging categories like cigarette detection.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 39 (June 2025) PP. 169-184
DOI: 10.35631/JISTM.1039011

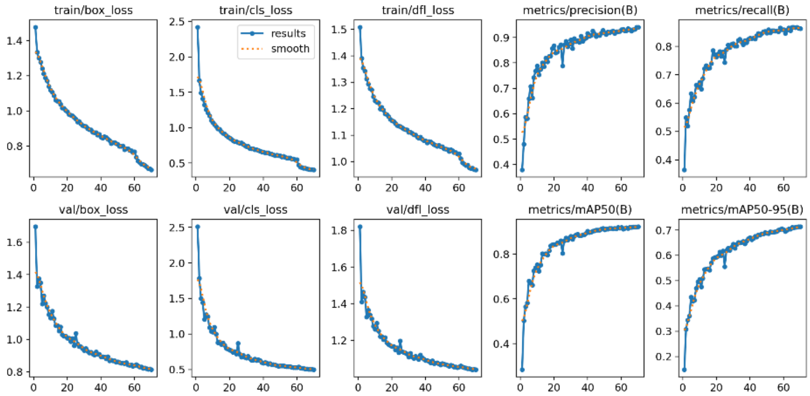


**Figure 3: Result Obtain After Training Yolov8 Optimizer Adamw Learning Rate 0.001 Using Custom Dataset**

Figure 4 displays the ground truth labels in the training dataset, including "Male" and "Female" bounding boxes (blue and teal), "Face_M" labels (white), and a lack of explicit annotations for "Cigarette" or "Smoking," suggesting these behaviors were less emphasized in labeling. Figure 5 shows the detection results from the trained model, where "Male" bounding boxes are detected with confidence levels of 0.5–0.6. Smoking-related features such as "Cigarette" and "Smoking" are detected with low confidence (0.3) in one or two frames, indicating weak predictions. Some "Face_M" detections are missing or weak, likely due to challenges in detecting faces in complex environments. The model excels at consistently identifying individuals, particularly "Male," with moderate confidence but struggles with smoking-related features and face detection. The low confidence for smoking features and missing detections highlight weak generalization and potential issues with lighting, occlusion, or small object sizes. Additionally, complex environments with cluttered backgrounds hinder the model's object detection and classification performance.

| | |
|---|---|
|  |  |
| **Figure 4: Sample Label In Training Dataset .** | **Figure 5: Detection Result Using The Trained Model Of Training Dataset** |

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 39 (June 2025) PP. 169-184
DOI: 10.35631/JISTM.1039011

Figure 6 presents the evaluation results of the YOLOv8 model, optimized with AdamW and a learning rate of 0.001, on the test dataset. The confusion matrix reveals the model's ability to classify various categories, including male, female, smoking, cigarette, face-related labels (Face_M and Face_F), and background. Performance metrics are influenced by the distribution of occurrences for each class, as noted: Cigarette (312), Face_F (155), Female (166), Face_M (259), Smoking (188), and Male (260).

The model exhibits high accuracy for gender classification, with 97% and 96% for male and female categories, respectively. Minimal misclassification is observed, with a small percentage of male instances predicted as female (1%) and vice versa (2%). Similarly, face detection performs well, achieving 94% for Face_M and 92% for Face_F. However, some confusion exists with background, where Face_F and Face_M instances are misclassified as background (5%).

For smoking and cigarette detection, the model demonstrates moderate performance. Smoking instances achieve an 85% accuracy rate, but 14% of instances are incorrectly predicted as background, indicating challenges in separating smoking-related actions from noisy environments. Cigarette detection is less accurate, with 74%, and a significant 24% of cigarette occurrences being classified as background. This suggests the model struggles with smaller or less distinct cigarette features compared to other classes.

The background class exhibits notable confusion, with 17% of smoking instances and 50% of cigarette occurrences misclassified as background. This misclassification underscores the need for better feature extraction and enhanced training data to improve the model’s capacity to differentiate subtle or overlapping elements.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 39 (June 2025) PP. 169-184
DOI: 10.35631/JISTM.1039011

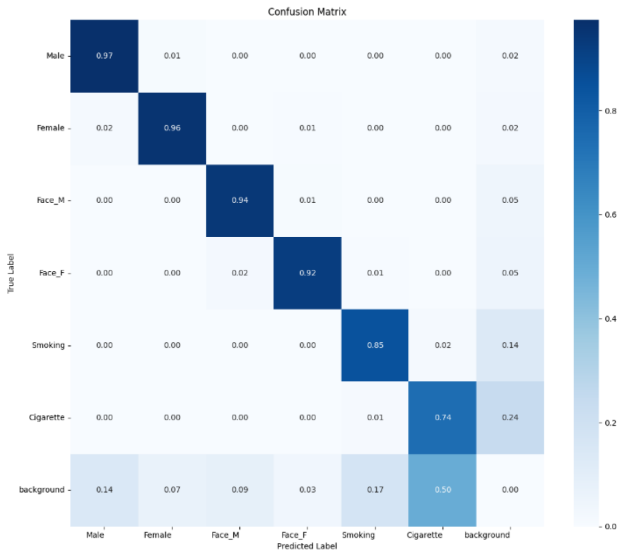


**Figure 6: Confusion Matrix Normalized For Testing Dataset**

Figure 7 shows the ground truth labels in the testing dataset, with bounding boxes for "Male" (blue), "Face_M" (white), and "Cigarette" (pink), representing the expected outcomes. Figure 8 presents the model's detection results, where the "Male" bounding box is detected reliably, with varying confidence levels for "Face_M" (e.g., 0.6, 0.8). However, the "Cigarette" detection is missing in the model's predictions, indicating a significant gap in its ability to detect smoking behavior. This may be due to insufficient representation of "Cigarette" in the training dataset, its small size, or subtle features that complicate detection. The inconsistent confidence for "Face_M" suggests potential issues with feature extraction or limited training data for faces.


**Figure 7: sample images with labels in testing dataset**


**Figure 8: Detection Result Using The Trained Model Of Training Dataset**

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 39 (June 2025) PP. 169-184
DOI: 10.35631/JISTM.1039011

The custom YOLOv8 model exhibits a notable limitation in detecting smoking behavior from high angles and at greater distances. This issue stems from the composition of the training dataset, which predominantly features close-up images with a limited number of labeled instances of smoking from CCTV footage. As a result, the model is well-tuned to recognize smoking in scenarios where the camera angle is low and the subject is close to the lens, ash shown in figure 9.

In contrast, when the camera angle is elevated or the subject is farther away, the model's performance declines significantly, as illustrated in figure 10. This discrepancy highlights the importance of a diverse and representative training dataset. In the current dataset, the lack of varied perspectives and distances means the model has not learned to generalize well to these conditions.

To improve the model's detection capabilities, it is crucial to augment the training dataset with more images and videos that capture smoking behavior from a variety of angles and distances. This could involve collecting additional CCTV footage that includes high-angle views and distant subjects, as well as manually labeling more instances of smoking in these contexts. By doing so, the model can learn to recognize smoking behavior more accurately across different scenarios, leading to more robust and reliable detection in real-world applications.

| | |
|---|---|
|  |  |
| **Figure 11: Detection of Smoking at Close and Low Angles** | **Figure 10: Failure to Detect Smoking at High Angles and Distances** |

**Conclusion And Future Work**

In conclusion, the custom YOLOv8 model integrated with BotSort tracking has demonstrated reliable performance in detecting smoking and non-smoking behaviors in controlled conditions, achieving high accuracy with the Adam optimizer at a learning rate of 0.001. This combination has proven to be the most effective, with high precision and recall, particularly in close-range, low-angle scenarios. However, the model's performance is limited when handling high camera angles or distant subjects, which reveals the dataset's lack of diversity. The current dataset's focus on close-up, low-angle conditions restricts the model's ability to generalize to more varied real-world environments. Future efforts will focus on diversifying and expanding the dataset, improving detection accuracy under different perspectives and distances, and further optimizing the model. Advanced training techniques and real-world testing will be employed to enhance robustness, while integration with edge computing for real-time

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 39 (June 2025) PP. 169-184
DOI: 10.35631/JISTM.1039011

processing and expanding the detection capabilities to analyze additional behaviors will ensure that the system adapts to diverse real-world conditions.

**Acknowledgement**

The authors would like to express their sincere gratitude to Prof. Madya Dr. Syed Muhammad Mamduh bin Syed Zakaria, the lead supervisor, for his invaluable guidance, constructive feedback, and continuous support throughout the development of this research. Deep appreciation is also extended to Dr. Ahmad Shakaff bin Ali Yeon, the co-supervisor, for his insightful comments and encouragement that significantly contributed to the quality of this work. The authors further acknowledge Ikram Mohsin for his assistance in the dataset collection process, which was essential for the success of this study.

**References**

Ahmad, M., Ali, S., & Yousaf, A. (2021). Benchmarking and comparative analysis of computer vision datasets in real-time detection tasks. Multimedia Tools and Applications, 80(4), 5055–5079. https://doi.org/10.1007/s11042-020-09988-3

Alikhanov, J., & Kim, H. (2023). Online action detection in surveillance scenarios: A comprehensive review and comparative study of state-of-the-art multi-object tracking methods. IEEE Access. https://ieeexplore.ieee.org/document/10173520

Harshavardhan, S., & Upender, P. (2025). SmokerBeacon: Hybrid Real-Time Smoking Detection Using Object Detection and Gas Sensor. IEEE. https://ieeexplore.ieee.org/document/10941558

Huang, S. K., Hsu, C. C., & Wang, W. Y. (2024). Multiple object tracking incorporating a person re-identification using polynomial cross entropy loss. IEEE Access. https://ieeexplore.ieee.org/document/10669031

Indra, Z., Hatmi, R. A., & Al Idrus, S. I. (2024). Smoking Violation Detection System Using YOLO in Non-Smoking Areas in Medan City. Jurnal Komputer. https://journalinstal.cattleyadf.org/index.php/Instal/article/view/313

Lavu, A. A. N. M., Zhang, H., Jonayed, M. D. A. I., & Zhao, H. (2024). Indoor Smoking Detection Method Based on Dual Spectral Fusion Image and YOLO Framework. LC International Journal of STEM. http://www.lcjstem.com/index.php/jstem/article/view/221

Li, M., & Liu, F. (2024). A Novel Finetuned YOLOv8 Transfer Learning Model for Smoking Behavior Detection. IEEE ICICML. https://ieeexplore.ieee.org/document/10957847

Li, R., Yu, C., Qin, X., An, X., Zhao, J., & Chuai, W. (2024). YOLO-SGC: A dangerous driving behavior detection method with multiscale spatial-channel feature aggregation. *IEEE Sensors Journal*. https://ieeexplore.ieee.org/abstract/document/10680977/

Ma, Y., Yang, J., Li, Z., & Ma, Z. (2022). YOLO-cigarette: An effective YOLO network for outdoor smoking real-time object detection. IEEE. https://ieeexplore.ieee.org/document/9816219

Mahdian, N., Jani, M., & Enayati, A. M. S. (2024). Ego-motion aware target prediction module for robust multi-object tracking. University of Victoria Repository. https://dspace.library.uvic.ca/bitstreams/c7e39aab-5687-4ed0-88ae-ca661d5ba261/download

Pandey, N. N., Pati, A., & Maurya, R. (2024). DriSm_YNet: a breakthrough in real-time recognition of driver smoking behavior using YOLO-NAS. Neural Computing and Applications. https://www.researchgate.net/publication/382596454

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 39 (June 2025) PP. 169-184
DOI: 10.35631/JISTM.1039011
Rathore, P. S., Balana, A., & Bhaumik, R. (2024). Benchmarking object detection and tracking for UAVs: An algorithmic comparison. IEEE International Conference on Vehicular Electronics and Safety. https://ieeexplore.ieee.org/document/10928092

Ren, Z. (2024). A Novel Feature Fusion-Based and Complex Contextual Model for Smoking Detection. IEEE CISC. https://ieeexplore.ieee.org/document/10653351

Sivaraman, G., Arora, R., & Mehta, S. (2023). Data augmentation and partitioning strategies in image classification: A comparative study. Journal of Computational Vision and Robotics, 5(2), 145–158. https://doi.org/10.1007/s00521-023-07845-1

Tu, S., Cao, Y., Liang, Y., Zeng, Z., Ou, H., & Du, J. (2024). Tracking and automatic behavioral analysis of group-housed pigs based on YOLOX+BoT-SORT-Slim. Smart Agricultural Technology, 4, 100171. https://www.sciencedirect.com/science/article/pii/S2772375524001710

Tzutalin. (2015). LabelImg: Labeling tool for object detection. GitHub repository. https://github.com/tzutalin/labelImg

Wang, Z., Lei, L., & Shi, P. (2023). Smoking behavior detection algorithm based on YOLOv8-MNC. *Frontiers in Computational Neuroscience*. https://www.frontiersin.org/articles/10.3389/fncom.2023.1243779/full

Wang, C., Zhao, J., Yu, Z., Xie, S., Ji, X., & Wan, Z. (2022). Real-time foreign object and production status detection of tobacco cabinets based on deep learning. Applied Sciences, 12(20), 10347. https://www.mdpi.com/2076-3417/12/20/10347

Wang, Z., Liu, Y., Lei, L., & Shi, P. (2024). Smoking-YOLOv8: A novel smoking detection algorithm for chemical plant personnel. Pattern Analysis and Applications. https://link.springer.com/article/10.1007/s10044-024-01288-7

Zang, H., Hu, Y., & Wang, Y. (2024). Research on Smoke and Fire Target Detection with Tracking Algorithm Based on YOLOv8 and DeepSORT. 2024 39th Youth Academic Annual Conference of Chinese Association of Automation. https://ieeexplore.ieee.org/document/10598422

Zhang, J., & Jiang, C. (2025). DAHD-YOLO: A new high robustness and real-time method for smoking detection. Sensors (Basel). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11902441/

Zhu, Y., Liu, Z., Li, K., Li, Y., Qi, X., & Ding, N. (2025). Research on multi-objective pedestrian tracking algorithm based on full-size feature fusion. IEEE Access. https://ieeexplore.ieee.org/document/10908239

Zutshi, T., Yeola, R., Kantroo, D., & Apte, R. (2024). Image-based Addiction Detection: An Automated Approach. IEEE ICIP. https://ieeexplore.ieee.org/document/10660823Li, M., & Liu, F. (2024). A Novel Finetuned YOLOv8 Transfer Learning Model for Smoking Behavior Detection. IEEE ICICML. https://ieeexplore.ieee.org/document/10957847