IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SWACHH AI – Public Spitting Detection System

Dr. Girish S. Thakare
Computer Science and Engineering
Sipna college of engineering and technology
Amravati, India

Anay B. Pund
Computer Science and Engineering
Sipna college of engineering and technology
Amravati, India

Swanandi G. Ambekar
Computer Science and Engineering
Sipna college of engineering and technology
Amravati, India

Atharva S. Rakhonde
Computer Science and Engineering
Sipna college of engineering and technology
Amravati, India

Rushikesh M. Charjan
Computer Science and Engineering
Sipna college of engineering and technology
Amravati, India

Aishwarya A. Dhole
Computer Science and Engineering
Sipna college of engineering and technology
Amravati, India

Abstract—Public spitting of chewable substances such as Gutkha and Paan Masala poses persistent hygiene, health, and aesthetic challenges in Indian cities, imposing substantial cleaning costs. Existing manual enforcement under the Swachh Bharat Abhiyan remains inefficient and unsustainable. This paper proposes Swachh AI, an AI-based real-time surveillance and enforcement system for detecting public spitting incidents and identifying offenders. The system integrates (i) a YOLOv11-based instance segmentation model to detect spitting actions with pixel-level accuracy, and (ii) an OCR-based license plate recognition module for offender identification. Together, these modules enable automated evidence capture, offender tracking, and data logging for civic enforcement.

Swachh AI is designed for modular integration, scalability, and privacy compliance, suitable for deployment in smart cities and public infrastructure. Experimental evaluations demonstrate high detection accuracy in varied outdoor conditions, affirming its potential for scalable urban hygiene management.

Keywords—Spitting detection, YOLOv11, instance segmentation, OCR, license plate recognition, public hygiene, AI surveillance, smart cities.

I. INTRODUCTION

Public hygiene in India faces severe challenges due to habits like spitting paan masala, gutkha, and tobacco-related products in public spaces. These unhygienic practices not only tarnish the aesthetic value of streets, walls, transport hubs, and government premises but also contribute to the spread of communicable diseases. Despite cleanliness campaigns like the Swachh Bharat Abhiyan, municipal authorities struggle to enforce hygiene regulations effectively. Gutkha spitting, in particular, leaves deep red stains on infrastructure, necessitating large-scale cleaning expenditures and posing serious health risks. The lack of scalable enforcement mechanisms further exacerbates the issue, as manual monitoring, cautionary signboards, and sporadic fines fail to deter habitual offenders.

Addressing these gaps, Swachh AI is designed as an AI-driven surveillance system capable of detecting spitting incidents in real time and identifying offenders through vehicle license plate recognition. The system integrates YOLOv11 for precise spit detection and PaddleOCR for license plate extraction, enabling authorities to automate penalties and create evidence logs for legal or administrative actions. Real-time alerts allow municipal bodies to respond efficiently, significantly improving hygiene enforcement in densely populated areas. Although limitations such as dependency on CCTV clarity and internet stability exist, the AI-powered solution aims to foster cleaner urban environments by reducing municipal maintenance costs and encouraging civic responsibility through deterrence-based technology.

II. LITERATURE REVIEW

Public spitting, especially involving gutkha and paan masala, remains a severe urban hygiene challenge in India, leading to significant health, economic, and social consequences. Studies by Dhamija et al. (2021) and Behera et al. (2020) highlight the role of spit-transmitted pathogens in spreading communicable diseases, including tuberculosis and COVID-19, particularly in densely populated environments. Stanistreet (2018) emphasizes the cultural normalization of spitting in developing nations, making enforcement difficult despite its harmful impact. Tobacco consumption trends, as outlined in the Tobacco Prevention & Cessation Journal (2021) and WHO guidelines (Yadav, 2016), show high gutkha use across various Indian states, directly correlating with oral health issues like submucous fibrosis and cancer. Financially, reports such as Zee News (2021) and the Ministry of Housing and Urban Affairs (2022) reveal massive municipal expenses—Indian Railways alone spends ₹12,000 crore annually on cleaning gutkha stains. Traditional enforcement methods, including manual monitoring and spot fines (Times of India, 2023), have shown limited effectiveness due to the absence of real-time offender identification. Recent AI advancements, as demonstrated by Sheikh et al. (2023), IEEE Xplore (2024), and IJRPR (2023), successfully deploy deep learning models for automated violation detection, particularly in traffic monitoring. Swachh AI integrates similar technology, utilizing YOLOv11 for spit detection and PaddleOCR for offender identification through vehicle license plate recognition (Redmon & Farhadi, 2018; Smith, 2007). The system enables real-time alerts and automated incident documentation, significantly enhancing enforcement capabilities. While limitations such as

dependence on CCTV clarity exist, Swachh AI presents a scalable, AI-driven solution to mitigate hygiene violations, reduce municipal expenses, and promote civic responsibility in urban spaces.

III. METHEDOLOGY

3.1 Research Design and Data Collection:

This study employs an Applied/Experimental research design with a descriptive component. The applied aspect focuses on developing and testing an AI-powered spitting detection system integrated with license plate recognition, aimed at tackling urban hygiene issues. The descriptive element examines public perception, prevalence, and social responses to spitting behaviors.

To frame the study, secondary data collection involved reviewing research papers, public health reports, civic regulations, and news articles, establishing the social, health, and legal context of public spitting. Primary data collection included field observations and community surveys, capturing 390 images of public spitting incidents using mobile and stationary cameras. The dataset ensures diverse environmental conditions such as varying lighting, backgrounds, and vehicle types, enabling robust AI model training. Public spitting remains a widespread urban hygiene concern in India, with surveys indicating that even habitual offenders perceive the act as unpleasant and expect civic authorities to intervene. This contradiction reflects the necessity for a structured enforcement mechanism to address hygiene violations effectively. To tackle this issue, a custom AI dataset was developed, given the absence of open-source datasets for spitting detection. The data set captures real-world scenarios across urban environments with varying lighting and backgrounds, ensuring robustness in detection. The annotation process employed the Computer Vision Annotation Tool (CVAT) with polygon masks for higher precision in instance segmentation.

- Social Behavior Insight: Even habitual offenders acknowledge public spitting as unpleasant yet continue the practice, necessitating enforcement.
- Custom AI Dataset: Due to the lack of open-source surveillance datasets, real-world data was manually collected and processed.
- Annotation Precision: CVAT was used to generate polygon masks rather than bounding boxes, improving object identification accuracy.

Defined Annotation Classes:

- Spit: Includes individual's head and spitting trajectory.
- No Spit: Individuals present without engaging in spitting.
- Number Plate: Visible vehicle registration plates for offender identification.
- Motorcycle: Entire vehicle and rider, excluding the spitting action.

Data Augmentation: Enhancing model adaptability through flipping, rotation, cropping, exposure adjustments, and noise addition.

Real-Time Surveillance Goal: The refined dataset forms the foundation for AI-powered detection and automated enforcement, assisting municipal authorities in curbing public spitting efficiently.

This AI-driven approach enables precise monitoring, offender identification, and scalable urban hygiene enforcement, ensuring cleaner public spaces.

Data Augmentation Techniques To enhance model robustness and counter data scarcity, the following augmentation techniques were applied:

• Horizontal Flip
• 90° and Random Rotations
• Random Cropping
• Random Shearing
• Blurring
• Exposure Adjustments
• Addition of Random Noise

This resulted in an enriched dataset better suited for training deep learning models under diverse real-world conditions.

3.2 Model Selection

YOLO11, the latest in the Ultralytics YOLO series, enhances real-time object detection with improved accuracy, speed, and efficiency. It features an upgraded backbone and neck architecture for better feature extraction, optimized training pipelines for faster processing, and achieves higher mAP with 22% fewer parameters than YOLOv8m. Designed for adaptability, it supports deployment on edge devices, cloud platforms, and NVIDIA GPUs. YOLO11 excels across multiple computer vision tasks, including object detection, instance segmentation, image classification, pose estimation, and oriented object detection (OBB).

- Optimized Efficiency: Refined architecture and optimized training pipelines enable faster processing while maintaining accuracy.
- Higher Accuracy: YOLO11m achieves improved mean Average Precision (mAP) with 22% fewer parameters than YOLOv8m.
- Adaptability: Supports deployment across edge devices, cloud platforms, and NVIDIA GPUs.
- Versatile Applications: Handles object detection, instance segmentation, image classification, pose estimation, and oriented object detection (OBB).

YOLO11 improves accuracy with fewer parameters using optimized model design, achieving higher mean Average Precision (mAP) while being computationally efficient. Its architecture enhances feature extraction, making it suitable for resource-constrained devices. Additionally, YOLO11 offers adaptability across various environments, including edge devices, cloud platforms, and NVIDIA GPU-based systems, supporting diverse applications such as real-time detection and segmentation tasks.

3.3 PaddleOCR

PaddleOCR is a multilingual Optical Character Recognition (OCR) framework developed by Baidu's PaddlePaddle team, offering a modular, end-to-end pipeline for text detection, direction classification, and recognition. It supports over 80 languages, including Latin, Chinese, Arabic, and Devanagari scripts, leveraging deep learning models for high-accuracy text extraction. The toolkit includes advanced algorithms such as PP-OCR, SRN, and NRTR, ensuring adaptability for both research and industrial applications. PaddleOCR is optimized for deployment across cloud and edge environments, providing lightweight models for low-resource applications and heavyweight server models for enhanced accuracy. Its flexibility allows users to refine individual components independently, making it suitable for real-time tasks such as document digitization, mobile scanning, and intelligent surveillance.

PaddleOCR utilizes a structured pipeline comprising text detection, recognition, and angle classification to enhance OCR accuracy. Its Differentiable Binarization (DB) model employs segmentation-based techniques to localize text precisely, overcoming challenges posed by irregular shapes, varying orientations, and cluttered backgrounds. The CRNN-based text recognition model integrates CNNs for feature extraction, RNNs for sequential processing, and CTC loss for alignment-free sequence prediction, enabling robust handling of distorted and multi-line text. Additionally, angle classification and post-processing techniques detect and correct misaligned text, improving recognition accuracy for scanned documents, license plates, and images with rotated text, making PaddleOCR adaptable for diverse applications.

Technical aspects of PaddleOCR's key features and advantages:

1. Multilingual Support

PaddleOCR supports over 80 languages, including complex scripts like Chinese, Arabic, and Indic languages, ensuring bidirectional text recognition. It enhances accuracy using language-specific character sets and lexicons while accommodating both Right-to-Left (RTL) and Left-to-Right (LTR) scripts, making it highly suitable for international OCR applications.

2. High Accuracy and Speed

PaddleOCR optimizes precision and speed using PP-OCRv3, a lightweight architecture featuring PP-LCNet for efficient processing, Differentiable Binarization (DB) for robust text detection, and CRNN with CTC loss for alignment-free sequence recognition. PaddleSlim quantization further reduces computational load while maintaining accuracy, ensuring effective OCR performance for various text types.

3. Mobile-Optimized Models

For edge deployment, PaddleOCR integrates MobileNetV3, a highly efficient CNN designed for low-latency inference on mobile and embedded devices. It also incorporates:

- Post-training Quantization (PTQ) and Quantization-aware Training (QAT) methods, reducing model size and inference time.
- Model pruning and distillation, ensuring adaptability for real-time applications.

3.3.2 Use Cases and applications

PaddleOCR is a versatile OCR framework designed for diverse industry applications, leveraging deep learning for text extraction. It facilitates document digitization, enabling automated conversion of scanned and printed materials into searchable formats. In financial processing, it extracts structured data from invoices, bank statements, and receipts for automation. The system supports retail and logistics by recognizing barcodes and printed text in supply chain operations. License Plate Recognition (LPR) capabilities allow for character identification in parking and traffic monitoring. Additionally, PaddleOCR enhances text-in-image search, enabling metadata extraction for media assets. In healthcare, it digitizes handwritten prescriptions and medical forms. With multilingual support and optimized models for embedded systems, PaddleOCR is well-suited for real-time deployment across mobile and cloud environments.

3.3.3 Advanced Features in PP-OCRv3

PP-OCRv3 introduces several improvements to enhance performance and efficiency:

• Improved Backbones: Uses MobileNetV3 for better speed and parameter efficiency.
• Knowledge Distillation: Employs teacher-student learning to transfer knowledge from large models to smaller ones.
• CTC Loss & Sequence Modeling: Improves recognition of long or connected text lines, even in noisy images.
• Data Augmentation Techniques: Includes synthetic data generation, perspective distortion, and character-level randomization to improve robustness.
• Refined Post-Processing: Enhances reading order and formatting, particularly in tabular and multi-column documents

3.2 3. Performance Analysis Using OpenCV for real-world applications reveals several strengths and challenges: • Strengths: OpenCV excels in real-time processing due to optimized algorithms.
• Challenges: Handling complex scenes (e.g., low-light conditions) can sometimes reduce accuracy.

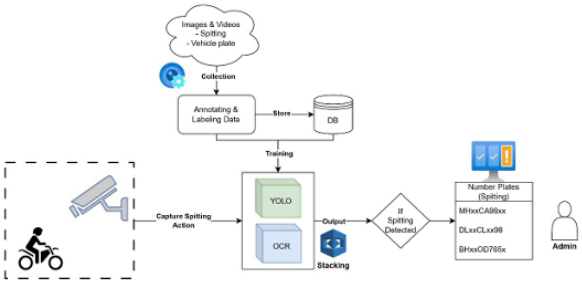
Performance Optimization:

- Hardware Acceleration: Utilize GPUs or TPU support.
- Multi-threading: Leverage multi-core processors for parallel computation.

3.3 System Architecture

The proposed system is engineered to detect public spitting incidents through real-time video analysis, leveraging advanced computer vision techniques. The architecture comprises the following components:

• Data Acquisition: Utilizing mobile cameras capable of 1080p resolution at 60fps for initial data collection.
• Model Training Environment: Training conducted on Google Colab's GPU environment, specifically utilizing a 12GB Tesla T4 GPU.
• Inference Engine: Deployment of the trained model for real-time inference on video feeds.
• OCR Module: Integration of PaddleOCR for license plate recognition.
• Alert System: Real-time notifications sent via Meta's WhatsApp Cloud API upon detection of spitting incidents.
• User Interface: A Flask-based web dashboard for monitoring and reviewing incidents.



3.3.4 Workflow Explanation :

The operational workflow of the system is as follows:

1. Frame Extraction: Continuous extraction of frames from live video feeds.
2. Pre-processing: Application of image enhancement techniques to improve detection accuracy.
3. Instance Segmentation: Utilization of the YOLOv11 instance segmentation model to identify and segment objects of interest.
4. OCR Processing: Application of PaddleOCR to extract text from detected license plates.
5. Alert Generation: Triggering of alerts through WhatsApp API when a spitting incident is detected.

6. Data Logging: Storage of incident data for future analysis and reporting.

3.3.5 Model Trainind Process
Model Selection and Configuration
• Model: YOLOv11n-seg (nano version with segmentation capabilities).
• Pre-trained Weights: Initialization with weights pre-trained on the COCO dataset.
• Training Parameters:
- o Epochs: 200
- o Batch Size: 16
- o Image Size: 640x640 pixels
- o Optimizer: AdamW (automatically selected for optimal performance)
- o Momentum: 0.9
- o Loss Functions: Combination of box loss, segmentation loss, classification loss, and distribution focal loss.
- o Learning Rate: 0.00125

- Training Environment
  - o Platform: Google Colab
  - o Hardware: Tesla T4 GPU with 12GB VRAM
  - o Software: Python 3.11.11, PyTorch 2.6.0+cu124

3.4.1 Hyperparameter Tuning
Hyperparameters were fine-tuned to optimize model performance:
• Learning Rate: Set to 0.00125 for stable convergence.
• Optimizer: AdamW selected for its adaptive learning capabilities and weight decay regularization.
• Batch Size: Determined as 16 to balance computational efficiency and model accuracy.
• Data Augmentation: Implemented techniques such as flipping, rotation, cropping, shearing, blurring, exposure adjustment, and noise addition to enhance model robustness.

3.4.2 Loss Curve Analysis
Training and validation loss curves were monitored to assess model learning:
• Initial Epochs: High loss values indicating model learning from scratch.
• Mid Training: Steady decline in loss values, suggesting effective learning.
• Final Epochs: Loss values plateaued, indicating convergence. Note: Include graphical representations of loss curves here for visual analysis.

3.4.3 Detection Pipeline

Confidence and IoU Thresholds
• Confidence Threshold: Set to 0.5 to balance precision and recall.
• IoU Threshold: Set to 0.7 to ensure accurate overlap detection. Handling Overlapping Detections
• Overlapping detections, such as a motorcycle rider spitting, were managed by associating the spitting action within the bounding box of the motorcycle, ensuring accurate attribution.

3.4.4 Real-Time Video Feed Handling
Frame Extraction and Pre-processing
• Frames extracted at the original video feed rate (60fps) to maintain temporal consistency.
• Pre-processing steps included resizing, normalization, and augmentation to enhance detection accuracy. Inference Performance
• The system achieved real-time processing with inference times per 60 frames being less than a second, resulting in no noticeable lag.

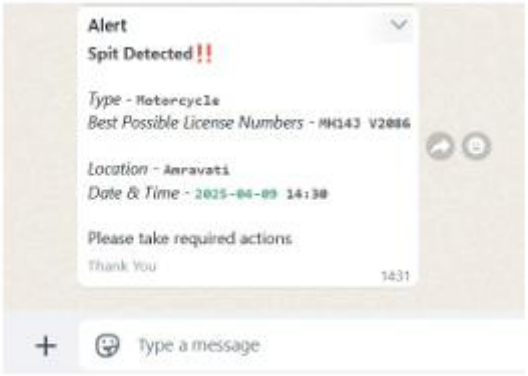
3.4.5 Number Plate OCR Pipeline
OCR Tool Selection
• Tool: PaddleOCR selected for its high accuracy and support for complex scripts.

Pre-processing Techniques:
• Conversion to grayscale.
• Contrast and sharpness enhancement.
• Dynamic exposure adjustment.
• Skew correction to handle angled plates.

Performance Metrics:
• Accuracy Range: 78% to 98%, with lower accuracy in cases of blurred or occluded plates.

3.4.6 Integration with Alert and Logging System
Alert Mechanism
• Platform: Meta's WhatsApp Cloud API.
• Trigger Condition: On successful detection of a spitting action and a readable number plate, the system automatically sends a WhatsApp alert to a pre-configured administrator number.
• Message Format: o Incident ID
- o Timestamp
- o Extracted vehicle number
- o Location coordinates (if available from video metadata or CCTV system)
- o Snapshot image of the incident frame
- o Optional: video snippet or link to footage stored securely on a server.

Example Alert:



3.4.6Incident Logging and Database Design
A structured database is essential for recording, managing, and analyzing incident data for administrative and legal follow-up.
Database: MongoDB Atlas Primary Collections:
• incidents o incident_id (unique) o date_time
- o location
- o vehicle_number
- o image_url
- o status (New, In Review, Action Taken)

• notifications
- o message_id
- o incident_id (foreign key)
- o receiver_contact
- o message_text

o delivery_status

This NoSQL structure was chosen for its flexibility, horizontal scalability, and ease of integration with real-time applications.

3.4.7 Challenges and Mitigation Strategies

Key Challenges Encountered:

• Detecting spitting actions in low-light conditions.
• Handling blurred, skewed, and partially occluded number plates.
• Distinguishing spitting action from similar gestures like coughing or drinking.
• Managing overlapping object detections (e.g. spitter behind a vehicle).

Mitigation Approaches:

• Applied frame-level brightness correction and dynamic gamma adjustments for dark frames.
• Enhanced image pre-processing in the OCR pipeline (contrast sharpening, skew correction).
• Trained the YOLO model with a larger and more diverse spitting action dataset including edge cases.
• Implemented bounding box hierarchy logic to manage overlapping detections.

3.5 User Interface and Visualization:

A lightweight, mobile-friendly Flask dashboard was created for:

• Live Feed Monitoring: View of real-time detections.
• Incident History: Searchable and filterable incident log.
• Alert Logs: Record of WhatsApp alerts dispatched.
• Data Analytics: Visualizations such as pie charts for violation distribution by location, time of day, and vehicle type.

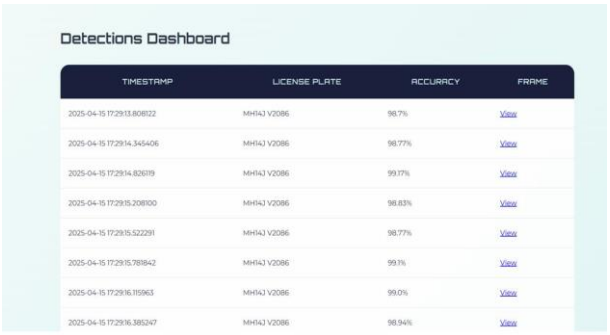
Technologies Used:

• HTML5/CSS3
• Bootstrap
• Chart.js for dynamic graphs
• Flask REST API backend

IV. IMPLEMENTATON

Detection Performance Analysis The detection system was evaluated on a held-out test set of annotated video frames representing diverse environmental conditions. The primary metrics considered were **mean Average Precision (mAP), F1-score, Precision,** and **Recall.**

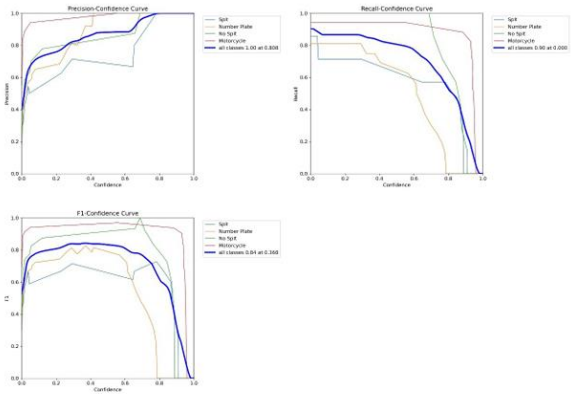
.



At a confidence threshold of 0.5, the detection system achieved:

• mAP: 88.36%

- F1-score: 88.92%
- Precision: 85.71%
- Recall: 83.21%



4. 1.1 **Normalized Confusion Matrix**

Table 6.1 presents the normalized confusion matrix for the detection model, with the predicted labels on the Y-axis and actual labels on the X-axis.

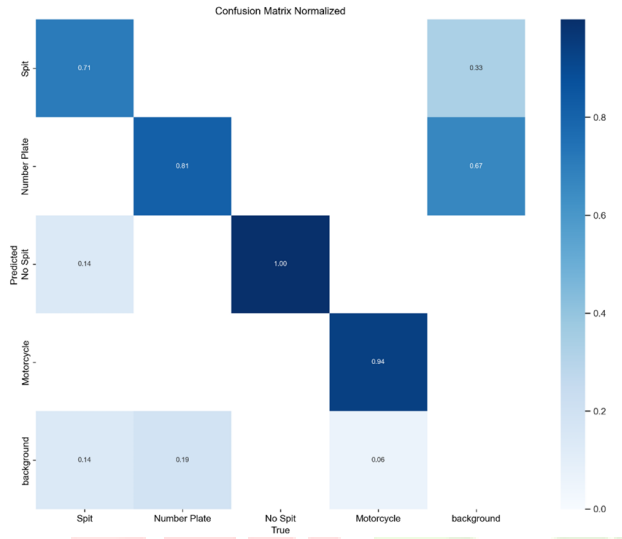


Table 6.1: Normalized Confusion Matrix

Interpretation:

• The model performs well on Spit (71%) and Number Plate (81%) classes. • Background confusion appears occasionally, with background being classified as Number Plate and Spit in a few instances.

• The No Spit and Motorcycle classes have high classification accuracies (1.00 and 0.94 respectively).

References

[1] S. Dhamija, et al., “Public spitting: A health hazard ignored,” National Library of Medicine, 2021.

[2] D. Stanistreet, “Public Spitting in 'Developing' Nations of the Global South: Harmless Embedded Practice or Disgusting, Harmful and Deviant?,” ResearchGate, 2018.

[3] A. Yadav, “Article 8 Implementation Guidelines,” WHO Framework Convention on Tobacco Control (FCTC).

[4] D. Behera, et al., “COVID-19 and Public Spitting: A Potential Health Hazard,” National Library of Medicine, 2020.

[5] R. Sheikh, et al., “Helmet and Number Plate Detection using AI,” ResearchGate, 2023.

[6] IEEEXplore, “Automatic Violation Detection System for Public Safety,” IEEE, 2024.

[7] IJRPR, “Helmet and Number Plate Recognition for Traffic Monitoring,” International Journal of Research Publication and Reviews, vol. 4, no. 6, 2023.

[8] Tobacco Prevention & Cessation Journal, “Tobacco Use Among Indian States: NFHS 5 (2019–20) & GATS (2016–17) Data,” Tobacco Prevention & Cessation, 2021.

[9] Times of India, “Pune Civic Body Struggles with Public Spitting Fines,” 2023.

[10] Times of India, “PMC Collects Fines for Spitting in Public,” 2023.

[11] J. Redmon and A. Farhadi, “YOLO: Real-Time Object Detection,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2018.

[12] R. Smith, “An Overview of the Tesseract OCR Engine,” in Proc. 9th Int. Conf. Document Analysis and Recognition (ICDAR), 2007.

[13] OpenCV, “Computer Vision and Machine Learning Techniques,” OpenCV Documentation, 2023.

[14] Ministry of Housing and Urban Affairs, Indian Urban Development Report, Government of India, 2022.

[15] Zee News, “Railways Spend Rs 12,000 Crore a Year to Clean Gutkha Stains, Come Up with New Plan,” 2021.

[16] N. Chaudhari, “Swachh AI: Real-time Spitting Detection using Camera,” 2022.