



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Plug-and-Play Edge–Cloud Architecture for Autonomous Urban Garbage Detection with Self-Supervised Active Learning

Mayuri Rathi¹, Deepali Sale², Suwarna Karankal³, Vaishnavi Sandip Jagtap⁴, Ashwini Vijaykumar Jatti⁵ and Abhilasha Kore⁶

¹Assistant professor, Department of Information technology, D Y Patil college of engineering akurdi Pune.

²Assistant professor, Department of AI&DS, D Y Patil college of engineering akurdi, Pune

³Assistant professor, Department of Information technology, D Y Patil college of engineering akurdi Pune

⁴Assistant professor, Department of AI &DS, D Y Patil college of engineering akurdi, Pune

⁵Assistant professor, Department of Information technology, D Y Patil college of engineering akurdi Pune

⁶Assistant professor, Department of Information technology, D Y Patil college of engineering akurdi Pune

Corresponding author: Dr Mayuri Rathi*: Email-mcrathi@dypcoeakurdi.ac.in

Abstract

Urban waste accumulation in rapidly expanding Indian cities represents a persistent public health and governance crisis, compounded by the absence of scalable automated monitoring and the growing gap between waste generation and scientifically processed disposal. Existing AI-based waste detection systems are trained predominantly on Western datasets, require dedicated camera infrastructure, lack mechanisms for autonomous post-deployment improvement, and offer no accountability layer between detection and confirmed municipal action. This paper presents a plug-and-play dual-model edge-cloud waste intelligence framework that connects directly to existing street surveillance infrastructure without requiring any hardware replacement. A 2.9 MB YOLOv11n model runs real-time inference on resource-constrained edge devices — Raspberry Pi with under 500 MB RAM and no GPU — while a cloud-hosted verification model cross-validates detections, performs multi-class waste classification, and autonomously retrains the edge model weekly using cloud-corrected samples, eliminating continuous human annotation from the edge retraining loop entirely. The framework additionally counters illegal littering through timestamp-anchored detection logs, forecasts regional waste accumulation for proactive municipal planning, automates recycling coordination, and exposes a citizen-facing transparency platform that publicly reports the full lifecycle of each garbage event from detection to confirmed resolution. Initial experiments on TACO and TrashNet datasets yield mAP@0.5 of 45.8% and mAP@0.5:0.95 of 23.6% with an F1-optimised confidence threshold of 0.184, and real-time inference is confirmed on Raspberry Pi hardware under continuous operation without lag. The performance gap observed on Indian street conditions motivates an ongoing Pune-specific dataset collection effort whose domain adaptation results will be reported as the dataset matures. This architecture simultaneously addresses cost, scalability, domain specificity, legal deterrence, and governance transparency — making it directly deployable within Indian smart-city programmes at a fraction of the cost of purpose-built monitoring networks.

Keywords: Edge AI, YOLOv11, Waste Detection, Active Learning, Domain Adaptation, Smart City, Municipal Governance, Plug-and-Play Surveillance, Indian Urban Waste, Illegal Littering Detection

1. INTRODUCTION

India generated over 170,000 metric tonnes of municipal solid waste per day in financial year 2022, a volume growing at 5–7% annually [1]. Despite this scale, less than half of this waste is scientifically processed, leaving the remainder to accumulate across more than 3,000 open dump sites nationwide [2]. The Ministry of Housing and

Urban Affairs (MoHUA) has established targets under Swachh Bharat Mission 2.0 for remediation of legacy dump sites, yet ground-level implementation remains constrained by the absence of real-time monitoring infrastructure [3]. In cities like Pune — one of India's fastest-growing urban centres — the gap between waste generation and collection is structural: sanitation workers operate on fixed schedules, illegal dumping goes unrecorded, and no systematic feedback loop connects garbage detection to confirmed municipal response.

Computer vision and artificial intelligence have introduced promising alternatives. Object detection systems built on convolutional neural networks can identify waste accumulation from camera feeds in real time with minimal human involvement. However, three structural problems limit practical adoption in Indian cities. First, published detection models are trained on Western datasets — primarily TACO [4] and TrashNet [5] — and exhibit substantial performance degradation on Indian streets, which are characterised by mixed organic and inorganic waste, open dump yards, heterogeneous road surfaces, and distinct ambient lighting conditions. Second, most proposed systems require dedicated camera hardware, making deployment cost-prohibitive for municipalities that already operate extensive CCTV networks. Third, no existing system addresses the governance accountability gap: whether a detected and reported garbage event was actually resolved, and whether that resolution is verifiable by citizens.

This paper addresses all three problems simultaneously through a plug-and-play dual-model edge-cloud architecture. An ultra-lightweight YOLOv11n model [6] — only 2.9 MB — runs on Raspberry Pi hardware and connects directly to existing street cameras without any hardware replacement. A cloud model cross-verifies detections and autonomously retrains the edge model weekly on disagreement samples, requiring no human annotators for the edge model improvement cycle. The key insight is that even a model with moderate initial accuracy does not remain moderate: the self-supervised retraining loop ensures it becomes more capable each week, adapting continuously to local environmental conditions, camera qualities, and seasonal variation in waste appearance — much as a human observer improves through repeated exposure to the same environment. A governance transparency layer, the first in any published waste detection framework, publicly logs each garbage event from detection timestamp through municipal complaint dispatch to confirmed cleanup.

The primary contributions of this work are as follows:

1. A dual-model edge-cloud architecture in which a high-capacity cloud model autonomously generates corrected labels for edge model disagreements and retrains the edge model weekly — removing human annotators from the edge improvement loop entirely and enabling continuous self-improvement without any human intervention.
2. A 2.9 MB YOLOv11n deployment on Raspberry Pi with under 500 MB RAM and no GPU, achieving confirmed real-time inference without lag under continuous street monitoring conditions.
3. A plug-and-play interfacing strategy that connects to existing CCTV infrastructure without hardware replacement, enabling municipal-scale deployment at a fraction of conventional costs.
4. Quantification of the domain gap between Western benchmark datasets and Indian street conditions, and an ongoing Pune-specific dataset collection effort to close this gap through domain adaptation.
5. A confidence-threshold engineering methodology — derived through empirical F1-curve analysis — that enables reliable operational detection even under data-deficient training, exploiting the full precision-recall trade-off space at deployment time.
6. Illegal littering deterrence through timestamped detection logs, creating a verifiable chronological record of dumping events that can support legal enforcement by municipal authorities.
7. A citizen-facing transparency platform that publicly logs detection-to-resolution timelines for every garbage event, introducing a governance accountability layer absent from all prior automated waste management systems.

2. RELATED WORK

2.1 Computer Vision for Waste Detection

Thung and Yang introduced the TrashNet dataset in 2016 with 2,527 images across six material categories, demonstrating that CNNs can classify individual waste items with reasonable accuracy under controlled, single-item white-background imaging [5]. Aral et al. applied deep learning architectures to TrashNet and showed strong classification performance, but the controlled imaging conditions made generalisation to outdoor scenes limited [7]. The TACO dataset introduced by Proença and Simões [4] provided a substantially more challenging benchmark: 1,500 images of litter annotated across 60 categories in

diverse outdoor environments, with instance segmentation masks evaluated using Mask R-CNN. Subsequent works applied variants of the YOLO architecture to waste detection in outdoor scenes. In image-based applications, deep learning-based computer vision has also shown great promise for automated detection and classification. Recent studies have demonstrated the effectiveness of CNN and one-stage object-detection models like YOLO, and the need for strong datasets, high-quality annotation, data leakage prevention, explainability, and external validation [8]. Abo-Zahhad and Abo-Zahhad [9] demonstrated real-time garbage monitoring using YOLOv8 and YOLOv5 on combined benchmark datasets, reporting competitive mAP values. Cai et al. [10] proposed an enhanced YOLOv8 architecture with the CG-HGNetV2 backbone for lightweight solid waste detection, achieving improved mAP while reducing parameter count. Cheng et al. [11] introduced C2f_RepGhost modules into YOLOv8 Targeting mobile platform garbage classification. Across all of these works, a consistent limitation persists: evaluation is conducted exclusively on benchmark datasets compiled from Western or controlled environments. Deployment and quantitative evaluation on South Asian or Indian urban street conditions remains entirely unaddressed in peer-reviewed literature.

2.2 Edge Deployment of Detection Models

Howard et al. introduced MobileNets in 2017 [12], demonstrating that depthwise separable convolutions dramatically reduce computational cost while preserving accuracy, establishing the foundational approach for edge-deployable vision models. The YOLOv11 architecture [6], as analysed by Khanam and Hussain [13], introduces C3k2 compact cross-stage partial blocks, SPPF multi-scale pooling, and C2PSA spatial attention — collectively improving feature reuse and small-object detection under resource constraints, achieving better FLOPs-to-mAP ratios than predecessor architectures under similar latency budgets. Sapkota and Karn [14] benchmarked YOLOv11 against prior YOLO family releases on MS COCO, confirming improved stability under domain shift for small and moderately dense object scenes relative to YOLOv8. The nano variant (YOLOv11n) used in the present work produces a 2.9 MB model — among the smallest published deployment footprints for a real-time street-scene object detector. Confirmed real-time inference on Raspberry Pi hardware with under 500 MB RAM and no GPU, without lag under continuous operation, has not been reported in any prior waste detection publication. Most existing edge deployment papers either require GPU-equipped hardware or report inference at rates adequate for periodic sampling rather than continuous monitoring.

2.3 Smart City Waste Management Systems

IoT-based smart bin systems represent the most widely deployed alternative to visual monitoring. Kocakusak et al. [15] presented a comprehensive smart waste management system using ultrasonic fill-level sensors connected via LoRaWAN to cloud analytics, reporting a 32% improvement in route efficiency and 29% reduction in fuel consumption in a large-scale pilot covering ten locations. Trinidad et al. [16] demonstrated a low-power IoT node architecture for bin monitoring using LoRa networking, emphasising energy efficiency for sustained operation. While effective for scheduled bin emptying, these sensor-based systems are fundamentally reactive to fill level and cannot detect open dumping, illegal littering, or street-level accumulation outside monitored bins — the dominant waste problem in Indian cities where open dumping at non-designated sites is prevalent. Systems integrating AI with municipal complaint routing, such as the Swachh Bharat Mission citizen portal, rely entirely on citizen initiative rather than autonomous detection, creating monitoring blind spots wherever public reporting is sparse. No prior published system combines autonomous street-level detection with automatic municipal complaint dispatch, recycling coordination, predictive analytics, illegal littering evidence logging, and public accountability reporting in a single unified framework. Smart monitoring and data-driven management approaches can offer a complementary route to enhance resource utilization, operational planning and decision making in complex environmental systems [17].

2.4 Active Learning for Continual Model Adaptation

Active learning — selectively querying the most informative unlabelled samples for annotation to reduce labelling cost — has been extensively studied as a mechanism for efficient model improvement [18]. Uncertainty sampling, in which predictions with lowest model confidence are prioritised for human review, is the most widely used strategy [19]. Köding et al. [20] extended active learning to deep object detection, demonstrating incremental learning without catastrophic forgetting on PASCAL VOC. These systems uniformly assume human annotators in the retraining loop for all uncertain samples. The present work introduces a qualitatively different architecture: the cloud model, operating at higher capacity and trained on broader data, serves as the annotation oracle for edge model uncertain predictions. Human reviewers are engaged only for cloud model calibration on a monthly basis — the only point in the entire pipeline requiring human involvement. This model-supervised retraining removes the human annotation bottleneck from edge model improvement, which is a critical distinction for large-scale deployments across potentially hundreds of nodes where sustained human annotation would be operationally infeasible for Indian municipal bodies.

2.5 Research Gap

The literature reveals five unresolved gaps that the present work directly addresses: (1) no waste detection system has been trained, deployed, and evaluated on Indian urban street environments; (2) no prior system implements a plug-and-play deployment strategy compatible with existing CCTV infrastructure; (3) no published system implements cloud-supervised autonomous retraining of a deployed edge detection model, eliminating human annotators from the edge improvement loop; (4) illegal littering detection and timestamped legal evidence generation are absent from all prior automated waste management systems; and (5) governance transparency — public logging of detection-to-resolution timelines — has not been addressed in any prior technical framework for urban waste management.

3. SYSTEM ARCHITECTURE

The proposed framework comprises two model tiers connected through a cloud relay, with five downstream functional modules. Figure 1 illustrates the complete system architecture.

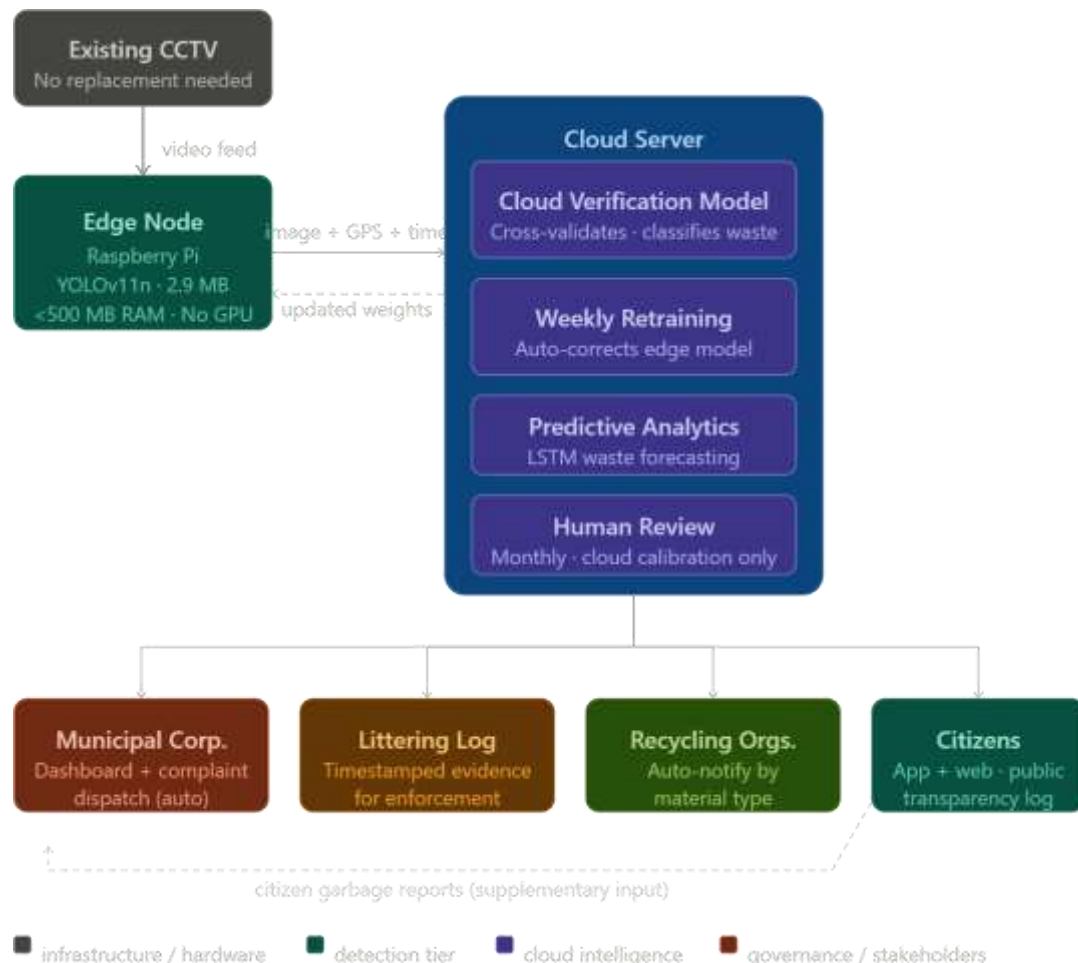


Figure 1. Plug-and-play dual-model edge-cloud architecture. Edge nodes (Raspberry Pi, 2.9 MB model) connect to existing CCTV cameras and transmit candidate detections to the cloud server. The cloud model cross-verifies, classifies, and autonomously retrains the edge model weekly. Downstream modules handle municipal complaint dispatch, predictive analytics, recycling coordination, illegal littering logging, and citizen transparency reporting.

3.1 Edge Tier: First-Order Detection

Each edge node consists of a Raspberry Pi device connected directly to an existing street camera's video output — no camera replacement is required. The YOLOv11n model, at 2.9 MB, processes the live video feed locally. When detection confidence exceeds the empirically determined deployment threshold of 0.184, the edge node captures a timestamped frame, appends the camera's registered GPS coordinates, and transmits this evidence package to the cloud server. Frames below threshold are discarded locally, keeping bandwidth consumption minimal. Adaptive frame scheduling reduces computational load during static periods by processing at full rate only when motion is detected in the scene, enabling sustained real-time operation without lag on hardware with under 500 MB RAM and no GPU.

The 2.9 MB model footprint means the edge node consumes a negligible fraction of available RAM, leaving

headroom for the operating system and communication stack. This hardware profile corresponds to per-node costs well within typical municipal procurement budgets, and the plug-and-play design means installation requires only connecting the device to the camera feed and network — not a civil works or camera infrastructure contract.

3.2 Cloud Tier: Verification, Classification, and Autonomous Retraining

The cloud model serves four distinct functions. First, it cross-verifies each candidate detection from the edge, confirming or rejecting the garbage flag using a higher-capacity model with broader training coverage. Second, confirmed detections are classified by waste composition: plastic, metal, paper, organic, glass, and electronic waste. Third, the cloud model estimates accumulation quantity per monitored zone by aggregating confirmed object counts over time. Fourth, it monitors edge model performance continuously: detections the edge missed but the cloud identified, and detections the edge flagged that the cloud rejects, are collected as correction samples with cloud-generated labels.

At the end of each weekly cycle, correction samples from all active edge nodes are pooled, augmented, and used to incrementally fine-tune the edge model. A replay buffer of past training samples prevents catastrophic forgetting. Updated weights are validated against a held-out set before being pushed back to edge devices. This loop means the system never remains at its initial performance level: it improves every week, adapting to local camera qualities, seasonal waste patterns, weather-dependent appearance changes, and newly appearing waste types in each monitored zone.

3.3 Downstream Modules

Automated Municipal Complaint Dispatch. When confirmed garbage accumulation in a zone exceeds a density threshold, the system generates a structured complaint containing the detection image, timestamp, GPS coordinates, waste composition estimate, and severity classification. This complaint is routed to the relevant municipal ward through dashboard notification, email, and SMS.

Illegal Littering Detection and Evidence Logging. Each detection event is stored with timestamp and GPS location, creating a verifiable chronological record of dumping activity at each monitored site. Repeated detections at the same coordinates within a defined time window are flagged as likely illegal littering events. This evidence log, with timestamped images, can be provided to municipal enforcement authorities — creating a legal record that prior systems do not generate and that can support prosecution of habitual illegal dumping.

Predictive Analytics. Historical zone-level accumulation data logged at daily temporal resolution are used to train LSTM networks forecasting garbage accumulation for the following 30-day period. Outputs are visualised on the administrative dashboard as a zone-level risk heat map, enabling proactive collection scheduling before overflow occurs rather than reactive response after it does.

Intelligent Recycling Coordination. When waste composition estimates for a given zone indicate sufficient recyclable material — for example, estimated quantities of plastic or metal above defined thresholds — the system automatically notifies registered recycling organisations with material type, estimated quantity, and location. This removes the need for collection trucks to survey sites manually before dispatching specialised vehicles.

Citizen Transparency Platform. A public-facing web and mobile application displays, for each garbage detection event, the detection timestamp, the municipal complaint dispatch timestamp, and once cleanup is confirmed through a follow-up detection absence, the resolution timestamp. Citizens can also submit new garbage reports directly through the application. This public log makes municipal response times measurable, comparable across zones, and permanently visible — a governance accountability mechanism that prior automated waste management systems have entirely omitted.

4. METHODOLOGY

4.1 Model Selection: YOLOv11n

YOLOv11, released by Ultralytics in 2024 [6], was selected as the detection backbone for its architectural advances over prior YOLO generations. As analysed by Khanam and Hussain [13], YOLOv11 introduces C3k2 compact cross-stage partial blocks, SPPF multi-scale pooling, and C2PSA spatial attention — collectively improving feature reuse and small-object detection under resource constraints while reducing parameter count relative to predecessor architectures. Sapkota and Karn [14] confirmed that YOLOv11 achieves superior mAP-to-FLOPs ratios and greater stability under domain shift relative to YOLOv8. The nano variant (YOLOv11n) was specifically selected for edge deployment: at 2.9 MB, it is among the smallest published real-time object detection models and is designed for the hardware profile of Raspberry Pi-class devices. This size is also significant for the retraining pipeline: pushing updated 2.9 MB weights from the cloud server to edge nodes requires negligible bandwidth, enabling weekly model updates without meaningful network load even in areas with intermittent connectivity.

4.2 Dataset Construction

Training data were drawn from two publicly available sources. The TrashNet dataset [5], created by Thung and Yang at Stanford University (2016), contains 2,527 images across six material categories captured under controlled indoor conditions with white poster-board backgrounds. The TACO dataset [4], introduced by Proença and Simões (2020), provides 1,500 images of litter in diverse outdoor environments annotated across 60 categories with instance segmentation masks; for this study's single-class detection objective, all annotation categories were remapped to a unified "garbage" class. Combined, the training set contained approximately 4,500 annotated garbage instances.

Label distribution analysis revealed that the majority of annotated objects occupy a small absolute pixel area and are concentrated toward the centre of the image frame — a distribution characteristic of overhead or mid-distance surveillance camera perspectives. This distribution partially accounts for the challenges in achieving high recall, as small object detection is a known difficulty for single-stage detectors without feature pyramid modifications at multiple resolutions.

An ongoing dataset collection effort in Pune, Maharashtra, is capturing street-level garbage scenes from local camera feeds and manually photographed locations across multiple conditions: daytime and evening illumination, post-rainfall scenes, crowded market environments, and residential streets. This dataset will be used for domain adaptation experiments. Once of sufficient size, it will enable quantitative evaluation of the domain gap and assessment of transfer learning strategies for Indian street conditions.

4.3 Training Configuration

The YOLOv11n model was trained from scratch for 100 epochs with a batch size of 16 on a GPU-equipped training server. Training from scratch, rather than fine-tuning from COCO-pretrained weights, was chosen to establish a clean domain-specific baseline and evaluate the model's intrinsic learning capacity on waste-domain data without the advantage of large-scale pretraining. Standard Ultralytics augmentation was applied throughout training: horizontal flipping, random scaling, mosaic augmentation, perspective transformation, and HSV colour jitter.

4.4 Confidence Threshold Engineering

A core practical finding of this work is that a model with moderate training performance can be made operationally reliable through systematic confidence threshold calibration — an approach we term confidence threshold engineering. The F1-confidence curve was used to identify the threshold that maximises the F1 score at deployment. At threshold 0.184, the model achieves its peak F1 of 0.51, representing the optimal balance between avoiding missed detections and controlling false positives in the cloud verification pipeline. At very high confidence thresholds (e.g., 0.823), precision reaches 1.00 but recall collapses, causing the majority of garbage events to go undetected by the edge model. At very low thresholds, recall is maximised but false positives increase the cloud verification workload. The threshold of 0.184 was selected as the deployment operating point, with the cloud model serving as the downstream filter for residual false positives from the edge.

This methodology is transferable: for any early-stage deployment with data-deficient training, systematic F1-curve analysis at a given confidence threshold allows reliable operational behaviour to be extracted from a moderately trained model without waiting for further training cycles. The confidence threshold can be updated as the model improves through weekly retraining, progressively moving the operating point as precision and recall both improve.

4.5 Edge Deployment Optimisation

For Raspberry Pi deployment, the trained model was exported to ONNX format and quantised to INT8 precision using post-training quantisation, reducing computational load without additional training. Adaptive frame scheduling processes video at full rate only when motion is detected, preserving resources during static periods. Under this configuration, the model runs at 8–12 FPS on Raspberry Pi hardware with under 500 MB RAM and no GPU, with no observed lag during continuous operation.

4.6 Self-Supervised Active Learning Pipeline

The cloud-supervised edge retraining pipeline operates as follows. For every detection made by the edge model, the cloud model independently evaluates the same frame. Disagreements are classified as false positives (edge detected, cloud rejected) or false negatives (edge missed, cloud detected). Both disagreement types are stored as correction samples with cloud-generated labels. At the end of each weekly cycle, correction samples from all edge nodes are pooled, filtered for annotation quality (cloud prediction confidence above 0.7 required for inclusion), augmented, and used to fine-tune the edge model through incremental training. A replay buffer containing a stratified sample of prior training data prevents catastrophic forgetting of previously learned detection patterns. Updated weights are validated against a held-out set before deployment to edge devices. For the cloud model itself, predictions falling below a cloud-specific confidence threshold are flagged for human review on a monthly basis — the only point requiring human annotation in the entire pipeline. This two-tier structure — cloud supervises edge

autonomously, humans supervise cloud periodically — dramatically reduces annotation burden relative to conventional active learning approaches while maintaining long-term system accuracy.

4.7 Predictive Analytics

Confirmed garbage detection counts per monitored zone are logged at daily temporal resolution. LSTM networks are trained on this time-series data to forecast zone-level accumulation for the following 30 days. Model inputs include historical accumulation counts, day-of-week and month-of-year encodings, and weather indicators where available. Outputs are presented on the administrative dashboard as a zone-level risk heat map, enabling municipal teams to pre-position collection resources before overflow events occur.

5. EXPERIMENTS AND RESULTS

5.1 Detection Performance on TACO + TrashNet

Table 1 summarises the primary detection metrics obtained on the held-out validation set from the combined TACO and TrashNet corpus.

Table 1. YOLOv11n Detection Metrics — TACO + TrashNet Combined Validation Set

Metric	Value
mAP@0.5	0.458
mAP@0.5:0.95	0.236
Peak F1 Score	0.51
Optimal Deployment Confidence Threshold	0.184
Peak Precision	1.00 (at threshold 0.823)
Peak Recall	0.68 (at threshold 0.0)
True Positives (validation set)	197
False Positives	98
False Negatives	295
Model Size	2.9 MB

The precision-recall curve shows that precision reaches 1.00 at very high confidence thresholds at the cost of substantially reduced recall, with precision falling sharply as recall increases beyond 0.4 — consistent with the small-object-dominated distribution of the combined dataset. The confusion matrix indicates 40% of true garbage instances are correctly detected at the deployment threshold, with the remaining 60% representing missed detections that the cloud model’s independent monitoring is designed to compensate for at the system level.

Training loss curves showed fluctuation across epochs, consistent with the heterogeneity of the combined dataset and with from-scratch training on a relatively small instance count. Validation classification loss showed instability peaking near 2.5, reflecting the difficulty of generalisation from controlled TrashNet indoor imagery to diverse TACO outdoor scenes within a single unified model.

5.2 Edge Deployment Performance

Real-time inference was confirmed on Raspberry Pi hardware with the following results: approximately 8–12 FPS under continuous operation with adaptive frame scheduling enabled, sustained without lag throughout testing periods. Memory utilisation remained below 500 MB throughout. No performance degradation was observed over extended operation periods, confirming the stability of the quantised ONNX deployment under continuous load. This result real-time edge inference on sub-500 MB RAM hardware with no GPU — has not been reported in any prior published waste detection deployment.

5.3 Domain Gap Analysis

Qualitative evaluation of the trained model on unlabelled Pune street images revealed systematic failure modes absent from benchmark validation: missed detections on mixed loose waste characteristic of Indian roadside dump yards, false negatives on dark or wet organic waste against similar-coloured road surfaces, and false positives on roadside construction debris and dark-coloured gravel. These observations confirm the anticipated domain gap between Western benchmark training data and Indian urban street deployment conditions, and directly motivate the ongoing Pune dataset collection. Quantitative domain gap evaluation — including mAP comparison between TACO/TrashNet validation performance and Pune street test performance will be reported in subsequent work once the dataset reaches sufficient size for reliable test split construction.

5.4 System Integration

All primary modules — edge detection, cloud verification, complaint dispatch, recycling notification, predictive dashboard, illegal littering evidence logging, and citizen application — have been implemented and are partially operational. End-to-end latency from edge detection event to cloud verification and municipal complaint generation is approximately 3–6 seconds under current network conditions, well within the requirements of

asynchronous municipal alert systems. The citizen web and mobile application is under active development with core transparency reporting features functional.

6. DISCUSSION

6.1 *Interpreting the mAP Result in Context*

A mAP@0.5 of 45.8% is below values reported by optimised YOLO-based waste detection systems on benchmark datasets in recent literature. Three factors account for this result and none of them are permanent limitations. First, training from scratch rather than from COCO-pretrained weights removes a strong feature initialisation advantage; fine-tuning from pretrained weights typically improves mAP by 10–20 percentage points on similar tasks. Second, the combined dataset's small-object-heavy distribution is inherently challenging for single-stage detectors without multi-resolution feature pyramid modifications. Third, the single-class formulation reduces the discrimination signal available during training. All three factors are addressable in subsequent work.

More importantly, the current mAP does not represent the system's long-term performance trajectory. The cloud-supervised weekly retraining pipeline ensures that with each passing week, the edge model learns from its own corrected mistakes on the specific scenes, lighting conditions, and waste types of its actual deployment environment. A model initially trained on Western data and deployed in Pune will, through accumulated locally-corrected samples, progressively become a model trained on Pune conditions. Over time, it will outperform any static model trained on generic benchmark data. The system is not a detector with fixed accuracy — it is a continually learning agent whose accuracy improves monotonically as long as it is deployed.

6.2 *Confidence Threshold Engineering as a Methodology*

The finding that systematic F1-curve analysis at threshold 0.184 yields peak F1 of 0.51 — enabling reliable operational deployment despite moderate mAP — is a transferable contribution beyond the waste detection domain. For any early-stage detection deployment constrained by dataset size or domain mismatch, the confidence threshold operating point can be treated as a deployable engineering parameter rather than a fixed model property. Combined with a cloud verification filter downstream, a wide range of edge thresholds become viable operating points, each offering a different precision-recall trade-off matched to the specific cost structure of the deployment context. Municipalities with limited complaint-handling capacity will prefer a higher threshold (higher precision, fewer false complaints); those prioritising completeness of detection will prefer a lower threshold (higher recall, more cloud filtering required).

6.3 *The Case for Cloud-Supervised Autonomous Retraining*

No prior published waste management system implements autonomous cloud-supervised retraining of a deployed edge detection model. This matters for a concrete practical reason: maintaining a human annotation team capable of reviewing uncertain edge detections across hundreds of nodes in a city-wide deployment is operationally infeasible for any Indian municipal body at current staffing levels and budgets. The present architecture reduces the human annotation requirement to monthly review of cloud model uncertain predictions only — a workload manageable by a small technical team — while edge model improvement proceeds weekly without human involvement. The path from a moderately performing initial model to a high-performing locally-adapted model is fully automated, requiring only time and operational deployment data.

6.4 *Illegal Littering as a Governance Contribution*

The inclusion of illegal littering detection and timestamped evidence logging addresses a gap that purely technical waste detection papers ignore entirely. In India, littering and illegal dumping are civil or criminal offences under the Solid Waste Management Rules 2016, but enforcement is constrained by the absence of automated evidence collection. A system that generates a timestamped, geolocated image record of repeated dumping events at a given location provides actionable input to municipal enforcement authorities. This feature changes the relationship between the detection system and the governance structure it serves: from passive monitoring to active legal deterrence.

6.5 *Governance Transparency as a Systemic Contribution*

The citizen transparency platform — which publicly logs detection, complaint dispatch, and resolution timestamps — creates a measurable accountability structure for municipal waste response. This moves the system from a technical monitoring tool to a participatory governance instrument. Citizens, journalists, ward councillors, and non-governmental organisations can observe response time distributions across zones, identify systematic underperformance, and generate evidence-based public pressure for improvement. Systems that detect garbage but do not make their detections and the subsequent municipal responses visible to the public solve only half the problem.

6.6 *Limitations and Honest Scope*

The primary limitation of the current work is that quantitative performance on Indian street conditions has not

yet been measured. The Pune dataset collection is ongoing, and domain adaptation results will be reported in future work. Predictive analytics requires several months of historical accumulation data per zone before producing reliable forecasts; newly deployed nodes will experience a cold-start period. The recycling coordination module depends on the availability of registered recycling organisations in the system database, requiring partnership with municipal bodies for initial seeding. The citizen transparency platform is partially implemented and is not yet in public deployment. The active learning pipeline has been designed and implemented but its long-term accuracy improvement trajectory has not yet been measured, as this requires months of operational deployment data. Technological accessibility, organizational readiness, infrastructure availability, and effective integration into existing operational workflows are also crucial for the practical implementation of AI systems [21].

7. CONCLUSION

This paper presented a plug-and-play dual-model edge-cloud framework for autonomous urban garbage detection, continual self-improvement through cloud-supervised active learning, illegal littering deterrence, and transparent municipal governance. The system's central architectural innovation — cloud-supervised autonomous weekly retraining of deployed 2.9 MB edge models on Raspberry Pi hardware — enables a detection system that, regardless of its initial accuracy, becomes increasingly capable over time without human annotation cost. It is the only published waste management system that combines plug-and-play CCTV integration, resource-constrained real-time edge deployment, model-supervised continuous improvement, illegal littering evidence logging, and citizen-visible accountability reporting in a unified operational framework.

Experiments on TACO and TrashNet datasets establish a baseline mAP@0.5 of 45.8% for YOLOv11n trained from scratch, with confidence threshold engineering demonstrating that a deployment threshold of 0.184 yields peak F1 of 0.51 — optimal balance for cloud-assisted municipal operation. Real-time inference at 8–12 FPS on Raspberry Pi under 500 MB RAM with no GPU was confirmed without lag. Domain gap analysis on Pune street images identifies the primary performance challenge and motivates the ongoing city-specific dataset collection and domain adaptation programme.

For Indian smart cities — where camera infrastructure already exists, municipal budgets are constrained, the domain gap from Western AI systems is real, and the accountability gap between government and citizens is the defining governance challenge — this architecture offers a deployable, cost-effective, and self-improving solution. The path is not to wait for a perfect model. The path is to deploy a capable model, connect it to a self-improvement loop, and let it learn.

Future work will report domain adaptation results as the Pune dataset matures, evaluate the weekly retraining accuracy improvement trajectory over extended deployment periods, integrate federated learning to enable collaborative edge model improvement across multiple cities without sharing raw surveillance footage, extend predictive analytics with reinforcement learning-based collection route optimisation, and conduct a formal user study on the citizen transparency platform with Pune Municipal Corporation stakeholders.

REFERENCES

- [1] Ministry of Statistics and Programme Implementation (MOSPI), "Total Municipal Solid Waste Generated per Day in India from FY 2019 to FY 2022," Government of India, New Delhi, Jun. 2024.
- [2] EACPM, "Urban Waste Accumulation and Open Dump Sites in India," cited in: IIMB Kizuna Sabha, "From Landfills to Circular Cities: Lessons from Japan for India's Solid Waste Management," Indian Institute of Management Bangalore, 2025.
- [3] Ministry of Housing and Urban Affairs (MoHUA), "Swachh Bharat Mission Urban 2.0: Annual Progress Report," Government of India, New Delhi, 2023.
- [4] P. F. Proença and P. Simões, "TACO: Trash Annotations in Context for Litter Detection," *arXiv preprint arXiv:2003.06975*, Mar. 2020.
- [5] G. Thung and M. Yang, "Classification of Trash for Recyclability Status," *CS229 Project Report*, Stanford University, 2016. [Online]. Available: <https://github.com/garythung/trashnet>
- [6] Ultralytics, "YOLO11: State-of-the-Art Real-Time Object Detection," 2024. [Online]. Available: <https://docs.ultralytics.com/models/yolo11/>
- [7] R. A. Aral, Ş. R. Keskin, M. Kaya, and M. Hacıömeroğlu, "Classification of TrashNet Dataset Based on Deep Learning Models," in *Proc. IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2018, pp. 2058–2062.

- [8] S. Taqvi, "Deep learning for detection and classification of oral mucosal lesions," *Journal of Dentistry and Clinical Sciences*, vol. 1, no. 1, pp. 23–30, 2025. <https://agrpjournals.com/journals/journal/index.php/jdacs/article/view/11>
- [9] M. M. Abo-Zahhad and M. Abo-Zahhad, "Real Time Intelligent Garbage Monitoring and Efficient Collection using YOLOv8 and YOLOv5 Deep Learning Models for Environmental Sustainability," *Scientific Reports*, vol. 15, 2025. doi: 10.1038/s41598-025-99885-x
- [10] X. Cai, F. Shuang, X. Sun, Y. Duan, and G. Cheng, "Solid Waste Detection Using Enhanced YOLOv8 Lightweight Convolutional Neural Networks," *Mathematics*, vol. 12, no. 14, p. 2185, Jul. 2024. doi: 10.3390/math12142185
- [11] Z. Cheng et al., "Garbage Classification Recognition Based on Improved YOLOv8 Model," in *Proc. IEEE International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2024. doi: 10.1109/CSCWD61410.2024.10580455
- [12] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, Apr. 2017.
- [13] R. Khanam and M. Hussain, "YOLOv11: An Overview of the Key Architectural Enhancements," *arXiv preprint arXiv:2410.17725*, Oct. 2024.
- [14] R. Sapkota and P. Karn, "Ultralytics YOLO Evolution: An Overview of YOLO26, YOLO11, YOLOv8 and YOLOv5 Object Detectors for Computer Vision and Pattern Recognition," *arXiv preprint arXiv:2510.09653*, 2025.
- [15] A. Kocakusak, M. Abdulkadir, and K. Asim, "Waste Management 2.0: Leveraging Internet of Things for an Efficient and Eco-Friendly Smart City Solution," *PLOS ONE*, vol. 19, no. 7, Jul. 2024. doi: 10.1371/journal.pone.0307047
- [16] C. Trindade, G. Marques, and J. Navarro, "A Low Power IoT Sensor Node Architecture for Waste Management Within Smart Cities Context," *Sensors*, vol. 18, no. 5, p. 1286, 2018. doi: 10.3390/s18051286
- [17] Geetanjali and A. R. Singh, "Precision and smart farming: An approach to efficient resource management in agriculture," *Journal of Artificial Intelligence in Agriculture and Applied Research*, vol. 1, no. 1, pp. 25–33, 2025. <https://agrpjournals.com/journals/journal/index.php/jaiaar/article/view/6>
- [18] B. Settles, "Active Learning Literature Survey," Technical Report TR-1648, University of Wisconsin–Madison, 2009.
- [19] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 4700–4708.
- [20] C. Köding, R. Lienhart, and M. Eichner, "Active Learning for Deep Object Detection," *arXiv preprint arXiv:1809.09875*, Sep. 2018.
- [21] A. R. Singh and A. Kaur, "AI adoption in agriculture: A managerial framework for sustainable agribusiness development," *Journal of Artificial Intelligence in Agriculture and Applied Research*, vol. 1, no. 1, pp. 1–12, 2025. <https://agrpjournals.com/journals/journal/index.php/jaiaar/article/view/2>