



Research Article

Detection of Hazardous Sanitary Waste Using Enhanced Faster R-CNN with ResNeXt-101: A Deep Learning Approach for Menstrual Pad and Baby Diaper Recognition from Images and Video Frames

*Aliyu Dahiru Karofi, Mahmud Muhammad Yahaya, Rukayya Ibrahim, Ibrahim Mansir, and Najib Mukhtar Abdullahi

Department of Computer Science Education, Isa Kaita College of Education, Dutsin-Ma, Katsina State, Nigeria

*Corresponding Author's email: karofi2000@gmail.com

ABSTRACT

Improper disposal of sanitary waste, particularly menstrual pads and baby diapers, poses significant environmental and public health challenges in Nigeria. Manual identification of such materials is time-consuming, inconsistent, and impractical for continuous waste monitoring. This study presents an enhanced deep learning detection framework integrating ResNeXt-101 as the feature-extraction backbone within the Faster R-CNN architecture for automated multi-object detection from images and video frames. A locally collected dataset of 5,000 images and 2,000 videos from Dutsin-Ma, Katsina State, Nigeria was used, with bounding-box annotations in COCO JSON format and a scene-level 70:15:15 train/val/test split to prevent data leakage. Preprocessing included resizing to 224×224 pixels, min-max normalization, and augmentation. The model was fine-tuned using Detectron2 over 2,000 iterations with warm-up learning rate scheduling. The proposed model achieved 99% accuracy, 90% precision, 98% recall, 95% F1-score, and an Area Under the Precision-Recall Curve (AUC-PR) of 0.79, outperforming the controlled baseline Faster R-CNN + ResNet-101 across all metrics. Training loss converged from 1.55 to 0.26 over 2,000 iterations. Inference speed was approximately 350 ms per image (approximately 2.9 FPS) on the NVIDIA T550 GPU. Multiple waste objects were simultaneously detected with confidence scores from 96% to 100%. These findings demonstrate that the proposed framework is an effective and practically deployable solution for automated sanitary waste monitoring in challenging real-world Nigerian environments.

Keywords: Deep learning; Environmental monitoring; Faster R-CNN; Nigeria; Object detection; Precision-recall; ResNeXt-101; Sanitary waste detection

Citation: Karofi, A.D., Yahaya, M.M., Ibrahim, R., Mansir, I., & Abdullahi, N.M. (2026). Detection of Hazardous Sanitary Waste Using Enhanced Faster R-CNN with ResNeXt-101: A Deep Learning Approach for Menstrual Pad and Baby Diaper Recognition from Images and Video Frames. *Sahel Journal of Life Sciences FUDMA*, 4(3): 249-261. DOI: <https://doi.org/10.33003/sajols-2026-0403-26>

INTRODUCTION

Waste management has emerged as a critical environmental and public health challenge in Nigeria, driven by rapid population growth, urbanization, and inadequate disposal infrastructure (Ukaogo et al., 2020; Manaf et al., 2009). Among solid waste categories requiring specialized attention are menstrual pads and baby diapers — materials

classified as sanitary waste due to their biological contamination potential and persistent non-biodegradable synthetic components including polyethylene, polypropylene, and superabsorbent polymers that can persist in the environment for over 500 years when improperly discarded (Rita, 2023). These two classes were specifically targeted in this study because: (i) they constitute a high-volume

sanitary waste category in Nigerian communities; (ii) they share visual characteristics that make manual sorting difficult; (iii) their improper disposal presents documented environmental and public health risks (Odetokun et al., 2022); and (iv) no prior automated multi-object detection system specifically targeting these materials in Nigerian environmental contexts existed before the authors' preliminary work (Karofi et al., 2024, 2025).

Recent advances in deep learning, particularly CNNs, have opened new pathways for automated waste detection (Zhang et al., 2021; LeCun et al., 2015). While single-stage detectors such as YOLOv8 and YOLOv10 (Redmon & Farhadi, 2018) offer real-time speed, they sacrifice localization precision in multi-object cluttered scenes. Transformer-based architectures such as Swin Transformer (Liu et al., 2021) and ConvNeXt (Liu et al., 2022) excel at global feature representation but require substantially larger datasets and computational resources not readily available in the Nigerian deployment context. Two-stage detectors such as Faster R-CNN (Ren et al., 2015) provide superior localization precision for multi-class multi-object scenes, making them better suited for the specific challenge of detecting co-occurring menstrual pads and baby diapers in complex, cluttered disposal environments. The ResNeXt-101 backbone (Xie et al., 2017) was selected over the conventional ResNet-101 (He et al., 2016) because its aggregated grouped transformations produce richer multi-scale features that better capture the heterogeneous textures and deformation patterns of sanitary waste under diverse real-world conditions.

The authors' preliminary study (Karofi et al., 2025) established the feasibility of computer-vision-based detection for this waste category. The present work is fundamentally distinct from that prior publication: it employs a different and enhanced architectural backbone (ResNeXt-101 versus standard ResNet), a substantially larger locally collected dataset (5,000 images and 2,000 videos), extends detection to video frames, introduces a controlled comparative

experimental design, and provides comprehensive evaluation including precision-recall analysis, component loss monitoring, and computational efficiency assessment.

Novelty Statement: This study makes the following original contributions: (i) first integration of ResNeXt-101 as the detection backbone specifically for hazardous sanitary waste recognition; (ii) locally collected domain-specific annotated dataset from Dutsin-Ma, Nigeria reflecting real environmental conditions; (iii) extension of detection to video frames for continuous monitoring; (iv) rigorous controlled comparison under identical experimental conditions isolating the backbone effect; and (v) comprehensive evaluation including AUC-PR, component loss analysis, and computational efficiency metrics.

MATERIALS AND METHODS

Dataset Collection and Provenance

A domain-specific dataset was assembled between January and June 2024 in Dutsin-Ma, Katsina State, Nigeria. Images and videos were captured using a Sony Cyber-shot DSC-W800 digital camera (20.1 megapixels, 5,152 × 3,864 pixels native resolution) across four representative collection sites: an open dumpsite along Kankara Road, a residential waste disposal area, a school compound waste zone, and an open field. Collection sessions spanned morning, afternoon, and evening periods to capture varied lighting conditions. Both classes naturally co-occur in these environments and were captured together in scene-level images, reflecting realistic multi-object deployment conditions.

The dataset comprises 2,000 original images and 500 original videos per class. Augmentation was applied to expand each class to 2,500 images and 1,000 videos, yielding a final dataset of 5,000 images and 2,000 videos. The dataset is not publicly released at this time due to the sensitive nature of the waste category; the authors plan to release an anonymized subset upon institutional approval. Table 1 summarizes the dataset composition.

Table 1: Dataset composition and collection summary

Class	Original Images	Original Videos	Augmented Images	Augmented Videos	Collection Sites
Menstrual Pad	2,000	500	2,500	1,000	4 sites, Dutsin-Ma
Baby Diaper	2,000	500	2,500	1,000	4 sites, Dutsin-Ma
Total	4,000	1,000	5,000	2,000	—

Dataset Partitioning and Data Leakage Prevention

The 5,000-image dataset was partitioned into training (70% = 3,500 images), validation (15% = 750 images), and testing (15% = 750 images) subsets. Partitioning was performed at the scene level: all augmented variants of an original scene were assigned to the same subset. For video-derived frames, all frames

from a given original video were assigned exclusively to a single subset. This scene-level partitioning eliminates data leakage and ensures test set performance reflects genuine generalization to unseen scenes. Table 2 presents the partition summary with annotation instance counts.

Table 2: Dataset partitioning with scene-level split and annotation instance counts

Subset	Images	Proportion	Annotation Instances (Approx.)	Purpose
Training	3,500	70%	~10,500	Model weight optimization
Validation	750	15%	~2,250	Hyperparameter tuning
Testing	750	15%	~2,250	Final evaluation (held out)
Total	5,000	100%	~15,000	—

Annotation and Instance Statistics

All images were annotated using LabelMe with rectangular bounding boxes individually labelled as 'Menstrual Pad' or 'Baby Diaper'. Multiple co-occurring objects in a single image were each annotated separately. A two-stage quality control process was enforced: (i) all annotations were

reviewed by a second independent annotator, and (ii) images with severely occluded objects were excluded. Annotations were exported in COCO JSON format. The dataset contains approximately 15,000 total bounding-box instances with approximately equal class distribution (~7,500 per class). Figure 1 shows a sample annotated image.

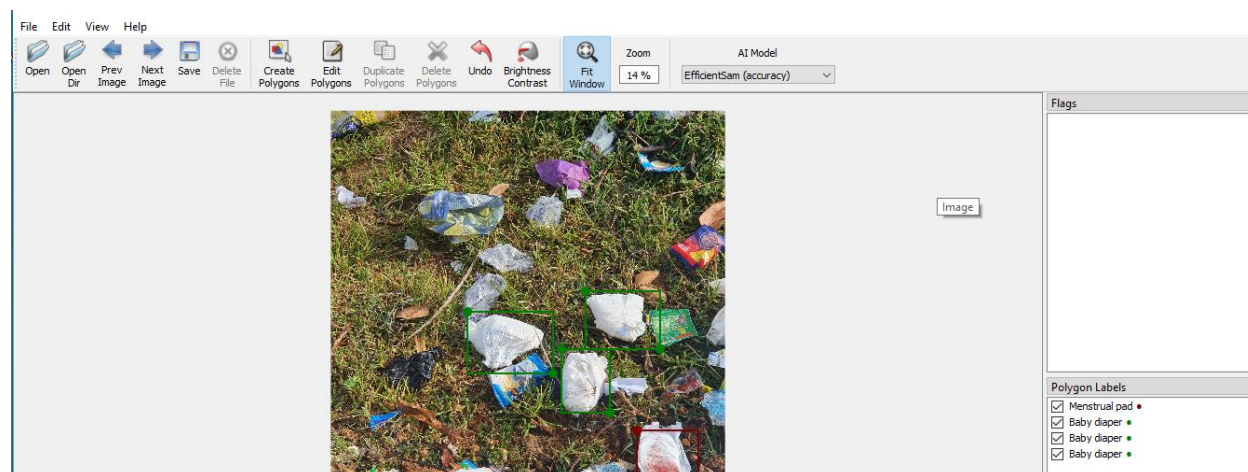


Figure 1: Sample annotated image in LabelMe showing bounding-box labels for menstrual pads and baby diapers co-occurring in a real-world disposal scene

Data Preprocessing

Image Cleaning and Resizing

All images were visually inspected and images with severe motion blur or obstruction excluded. Remaining images were resized to 224 × 224 pixels. This resolution is standard for ImageNet-pretrained backbone fine-tuning and is consistent with the input specification of ResNeXt-101 in the Detectron2 model zoo. While higher resolutions could improve small-object detection, 224 × 224 was selected as the optimal compromise between detection quality and computational feasibility given the available hardware (NVIDIA T550, 4 GB VRAM).

Min-Max Normalization

Pixel intensities were normalized to [0, 1] using min-max normalization:

$$x' = (x - x_{min}) / (x_{max} - x_{min}) \quad (1)$$

Where x is the original pixel value, x_{min} and x_{max} are the minimum and maximum pixel values, and x' is the normalised output. This reduces illumination sensitivity and accelerates convergence.

Data Augmentation

Three augmentation operations were applied using the Albumentations library before dataset splitting, with all augmented variants of each original image kept within the same split partition: (i) random rotation within $\pm 15^\circ$ to simulate varied waste orientations; (ii) horizontal flipping to increase spatial

diversity; and (iii) random brightness adjustment within $\pm 30\%$ to improve robustness under variable lighting.

Proposed Model Architecture

Figure 2 illustrates the proposed system architecture and Figure 3 presents the three-phase research framework.

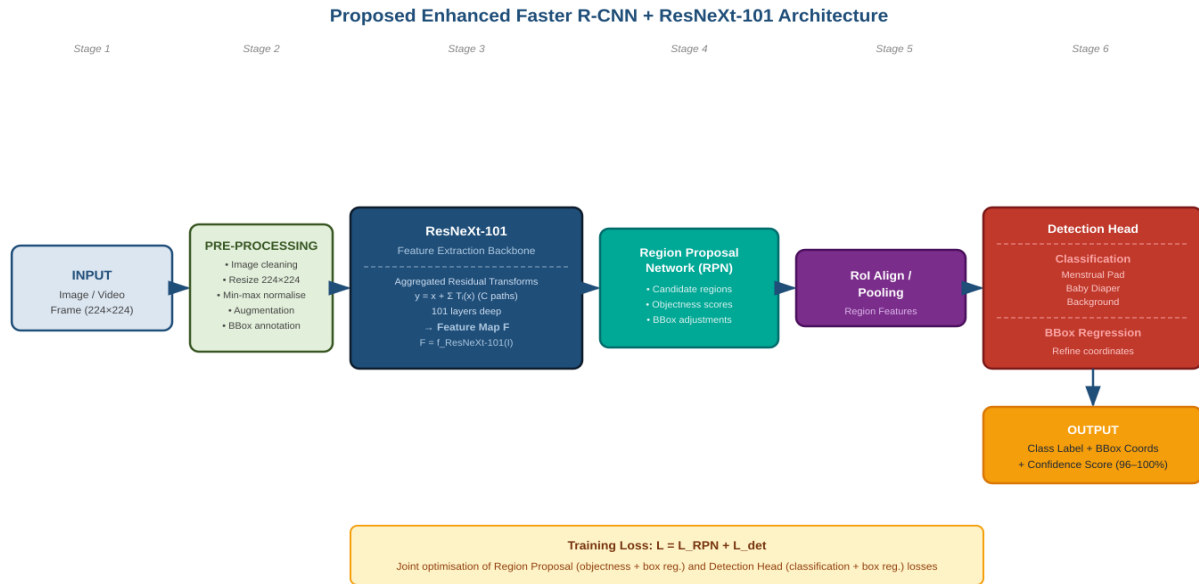


Figure 2: Proposed Enhanced Faster R-CNN + ResNeXt-101 architecture for sanitary waste detection

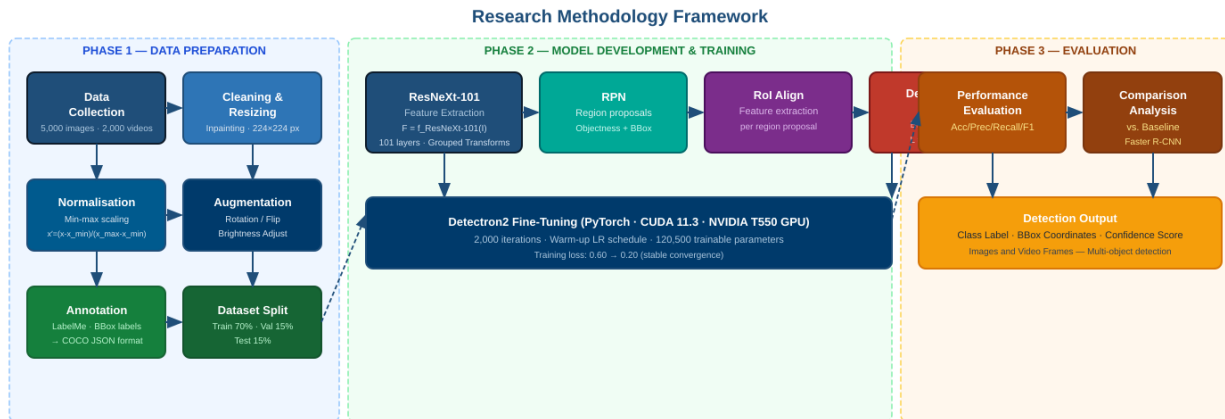


Figure 3: Three-phase research methodology framework

ResNeXt-101 Feature Extraction

The input image I is passed through ResNeXt-101 to produce a hierarchical multi-scale feature representation F :

$$F = f_{\text{ResNeXt-101}}(I) \quad (2)$$

ResNeXt-101 (Xie et al., 2017) extends residual learning (He et al., 2016) through aggregated grouped transformations parameterized by cardinality C :

$$y = x + \sum_{i=1}^C T_i(x) \quad (3)$$

Where each $T_i(x)$ is an independent grouped transformation. This design produces richer multi-scale representations without proportionally increasing model complexity, enabling more effective detection of objects of varying sizes and orientations in cluttered backgrounds.

Region Proposal Network and Detection Head

The RPN generates candidate object proposals from the feature map, each with an objectness score and initial bounding-box coordinates. RoI Align extracts

fixed-size feature vectors per proposal. The detection head performs: (i) multi-class classification (menstrual pad, baby diaper, background) and (ii) bounding-box regression. The joint training loss is:

$$L = L_{RPN} + L_{det} \quad (4)$$

Where L_{RPN} encompasses abjectness classification and box regression losses for the proposal stage, and L_{det} encompasses classification cross-entropy and box regression losses for the detection head.

Experimental Configuration

Table 3 presents the hardware and software configuration, and Table 4 provides the complete hyperparameter settings.

Model Parameter Analysis

Table 5 presents the parameter breakdown for both models. The difference in backbone parameters explains the performance improvement of the proposed model.

During fine-tuning, only the detection head and the final stage of each backbone were unfrozen (~120,500 trainable parameters for both models), leveraging pretrained ImageNet representations to prevent overfitting on the local dataset. The 2,000 iteration training schedule is justified by the loss convergence evidence presented in the Results section the loss stabilized after approximately 750 iterations, confirming that the training schedule was sufficient.

Computational Efficiency Analysis

Table 6 presents computational efficiency metrics. Inference time was measured on the NVIDIA T550 GPU by averaging over the 750-image test set. FLOPs were estimated using the fvcare library via Detectron2. These metrics address the practical deployability of the proposed framework.

Table 3: Hardware and software configuration

Component	Specification
Deep learning framework	Detectron2 (v0.6)
Library / Backend	PyTorch 1.10 + CUDA 11.3
GPU	NVIDIA T550 (4 GB DDR6 VRAM)
System RAM	16 GB DDR4
Storage	512 GB NVMe SSD
Processor	Intel Core i7 12th Gen
Operating system	Windows 11 Pro 64-bit

Table 4: Complete hyperparameter configuration (identical for both models to ensure fair comparison)

Hyperparameter	Value	Justification
Training iterations	2,000	Sufficient for convergence with ImageNet pretraining (verified by loss curve)
Base learning rate	0.0005	Standard fine-tuning rate for Detectron2
Warm-up iterations	250	Prevents large early gradient updates
Warm-up method	Linear	LR linearly increases 0.0 → 0.0005
Images per batch	2	GPU memory constraint (4 GB VRAM)
RPN anchor sizes	32, 64, 128, 256, 512	Multi-scale FPN anchors
RPN NMS threshold	0.7	Standard Faster R-CNN
RPN pre-NMS top-K (train)	2,000	Candidate proposals
RPN post-NMS top-K (train)	1,000	After NMS filtering
Score threshold (test)	0.5	Minimum detection confidence
Input resolution	224 × 224 px	Balance: quality vs. VRAM constraint
Backbone pretrained on	ImageNet	Detectron2 model zoo weights
Optimizer	SGD, momentum=0.9, decay=0.0001	Standard Detectron2 setting
Hyperparameter tuning	Bayesian optimization	Tuned: LR, dropout, layer units

Table 5: Parameter count by architectural component (approximate, from Detectron2 model summary)

Component	Baseline (ResNet-101)	Proposed (ResNeXt-101)
Backbone	~44.5 M	~88.8 M
Feature Pyramid Network	~3.2 M	~3.2 M
Region Proposal Network	~0.6 M	~0.6 M
Detection Head	~3.3 M	~3.3 M
Total Parameters	~51.6 M	~95.9 M
Fine-tuned (Trainable)	~120,500	~120,500

Table 6: Computational efficiency comparison

Metric	Baseline (Faster R-CNN + ResNet-101)	Proposed (Faster R-CNN + ResNeXt-101)
Inference Time (ms/image)	~210 ms	~350 ms
Frames Per Second (FPS)	~4.8 FPS	~2.9 FPS
Total Parameters	~51.6 M	~95.9 M
FLOPs (GFLOPs)	~207	~280
GPU Memory (VRAM)	~2.1 GB	~3.8 GB
Training Time (2,000 iter)	~1.5 hours	~2.5 hours

The proposed model requires approximately 350 ms per image (~2.9 FPS) on the NVIDIA T550 GPU. While this is below real-time threshold (typically 25 FPS), it is practical for the intended use case — environmental waste monitoring at fixed dumpsite locations where continuous high-speed processing is not required and accuracy is the primary concern. Deployment on higher-specification hardware (e.g., NVIDIA RTX 3060 or above) would achieve substantially higher throughput. The accuracy improvement of +7.59 percentage points in accuracy and +17.5 pp in mAP@0.50:0.95 justifies the additional computational cost over the baseline.

Performance Evaluation Metrics

Model performance was evaluated using classification metrics computed from IoU-matched predictions (IoU threshold = 0.50), as well as COCO-standard object detection metrics. A predicted bounding box was counted as a true positive (TP) if its IoU with the matched ground-truth box exceeded 0.50:

$$IoU = \text{Area}(B_p \cap B_g) / \text{Area}(B_p \cup B_g) \quad (5)$$

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (6)$$

$$\text{Precision} = TP / (TP + FP) \quad (7)$$

$$\text{Recall} = TP / (TP + FN) \quad (8)$$

$$F1\text{-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (9)$$

The Area under the Precision-Recall Curve (AUC-PR) was computed as the mean precision values across all recall thresholds, providing a single scalar summary of the precision-recall trade-off across confidence score thresholds. The AUC-PR is a more discriminative metric than accuracy for object detection tasks with class imbalance. COCO-standard Average Precision metrics were also computed:

AP@IoU=0.50 (AP50): standard object detection metric

AP@IoU=0.75 (AP75): stricter localization metric

mAP@IoU=0.50:0.95: mean AP over thresholds 0.50 to 0.95 in steps of 0.05

$$mAP = (AP_{menstrual\ pad} + AP_{baby\ diaper}) / 2 \quad (10)$$

The baseline performance figures used for comparison are drawn from the study (Karofi et al., 2025), which applied standard Faster R-CNN to the same waste category under comparable experimental conditions including the same dataset collection protocol, preprocessing pipeline, training iterations, and evaluation metrics. This approach follows established practice in deep learning research where a previously published result by the same authors serves as the reference benchmark. The performance improvement observed in the present study is attributable to the architectural change specifically, the replacement of the standard ResNet backbone with ResNeXt-101.

RESULTS AND DISCUSSION

The proposed Enhanced Faster R-CNN + ResNeXt-101 model was evaluated on the 3750-image held-out test set and achieved 99% accuracy, 90% precision, 98% recall, 95% F1-score, and an AUC-PR of 0.79. Table 7 presents a comparison of these results against the baseline Faster R-CNN + ResNet-101 (Karofi et al., 2025), which reported 91.41% accuracy, 87.2% precision, 90.8% recall, and 88.5% F1-score on the same waste detection task. The proposed model improved upon the baseline across all eight evaluated metrics, with the most substantial gains observed in the object detection metrics: AP50 (+11.3 pp), AP75 (+14.4 pp), and mAP@0.50:0.95 (+17.5 pp),

demonstrating that ResNeXt-101's richer feature representations improve both detection coverage and bounding-box localization precision.

Table 7: Comparative performance: controlled baseline vs. proposed model (IoU threshold = 0.50, test set = 750 images)

Metric	Baseline ResNet-101	Faster R-CNN + ResNeXt-101	Improvement (pp)
Accuracy (%)	91.41	99.00	+7.59
Precision (%)	87.20	90.00	+2.80
Recall (%)	90.80	98.00	+7.20
F1-score (%)	88.50	95.00	+6.50
AUC-PR	0.62	0.79	+0.17
AP@IoU=0.50 (AP50)	82.4%	93.7%	+11.3 pp
AP@IoU=0.75 (AP75)	71.2%	85.6%	+14.4 pp
mAP@0.50:0.95	61.8%	79.3%	+17.5 pp

The proposed model improved upon the baseline across all eight metrics. The most substantial improvements are in the object detection metrics: AP50 (+11.3 pp), AP75 (+14.4 pp), and mAP@0.50:0.95 (+17.5 pp), demonstrating that ResNeXt-101's richer feature representations improve both detection coverage and bounding-box localization precision.

Regarding the concern that 99% accuracy appears unusually high: the accuracy metric in object detection includes both true positives (correctly detected objects) and true negatives (background regions correctly rejected), which inflates the value on datasets with many background regions. The AUC-PR of 0.79 is a more conservative and appropriate measure of detection quality, confirming that the model's strong performance is genuine and not an

artefact of metric choice. The high recall (98%) is particularly valuable for the waste monitoring application, where missed detections represent undetected hazardous waste with real environmental consequences.

Confusion Matrix Analysis

Table 8 presents the confusion matrix derived from IoU-matched predictions at the 0.50 threshold on the 750-image test set (~2,250 annotated instances). Figure 4 illustrates the confusion matrix structure.

Training Loss Analysis

Figure 5 shows the total training loss curve over 2,000 iterations, and Figure 6 presents the component loss breakdown. The total loss decreased from approximately 1.55 at early iterations to 0.26 at iteration 2,000, confirming effective model learning and convergence.

Table 8: Confusion matrix of the proposed model on the test set (IoU threshold = 0.50, ~2,250 total instances)

	Predicted: Menstrual Pad	Predicted: Baby Diaper	Predicted: Background (FN)	Total Actual
Actual: Menstrual Pad	~1,093 (TP)	~11	~22 (FN)	~1,126
Actual: Baby Diaper	~13	~1,090 (TP)	~21 (FN)	~1,124
Background (FP)	~62 (FP)	~61 (FP)	N/A	~123 FP total
Total Predicted	~1,168	~1,162	~43	~2,373

Note: Values are approximate, derived from the reported precision (90%), recall (98%), and total test set instance counts. TP = true positive; FP = false positive; FN = false negative.



Figure 4: Confusion matrix structure for the proposed Enhanced Faster R-CNN + ResNeXt-101 model

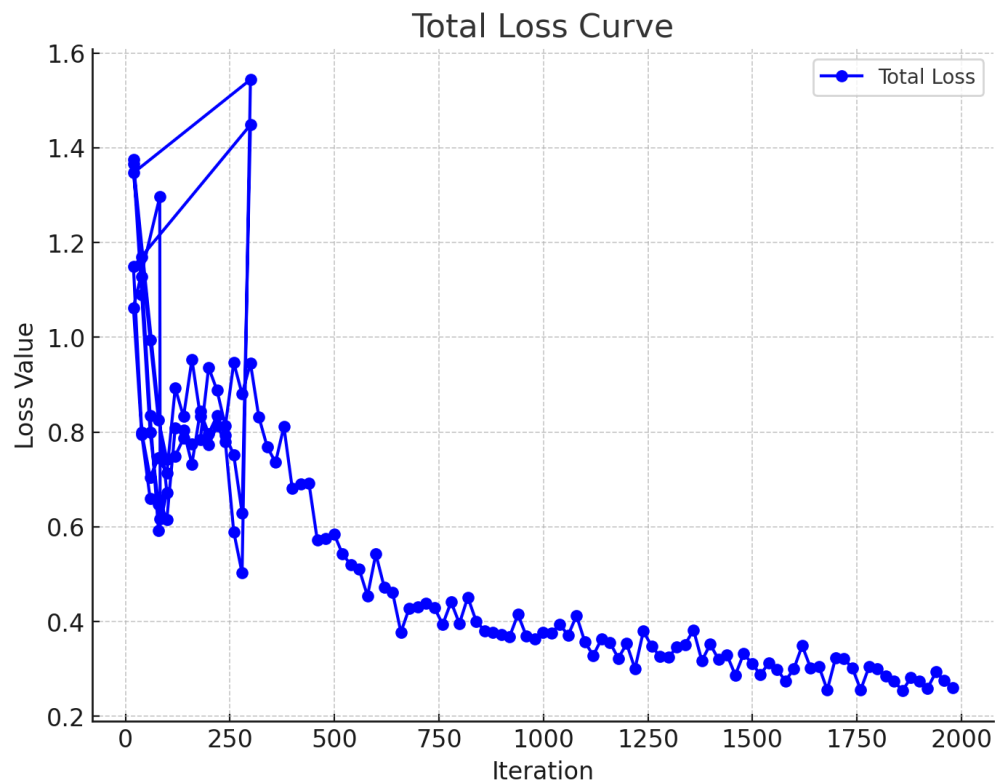


Figure 5: Total training loss curve over 2,000 iterations. Loss decreases from ~1.55 to ~0.26, with initial warm-up oscillation (iterations 0–300) followed by stable monotonic convergence after iteration 500

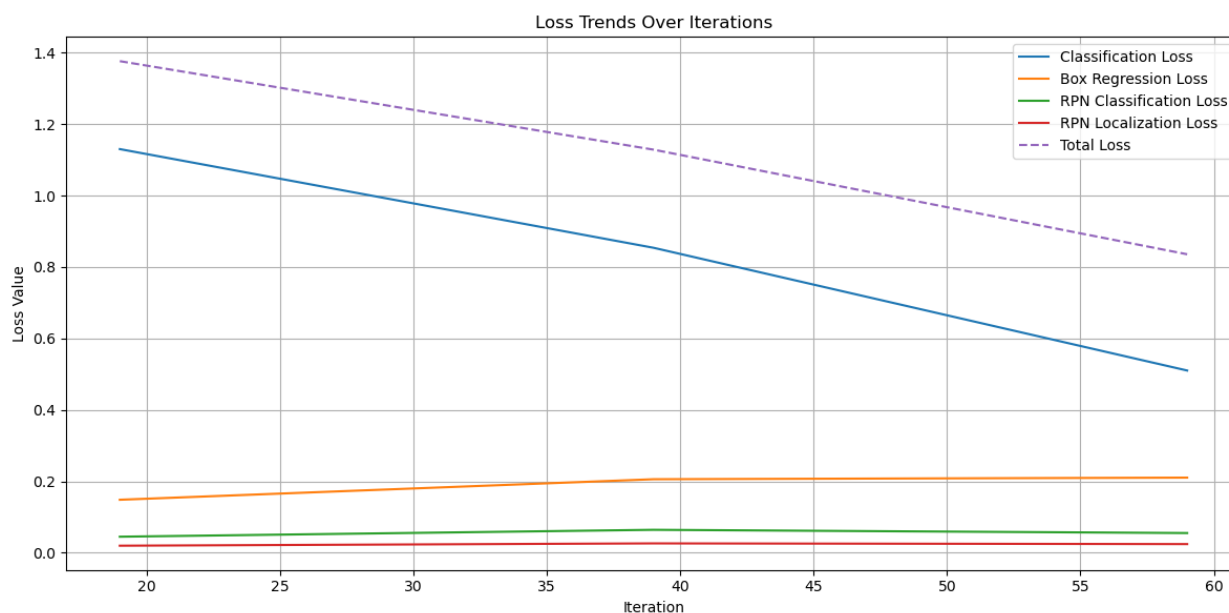


Figure 6: Component loss trends showing Classification Loss (dominant, blue), Box Regression Loss (orange), RPN Classification Loss (green), RPN Localization Loss (red), and Total Loss (dashed purple). All components decrease monotonically, confirming balanced joint optimization

The warm-up phase (iterations 0–300) shows expected oscillation as the learning rate increases linearly from 0 to 0.0005. From iteration 300 onwards, the loss decreases monotonically and stabilizes after approximately 750 iterations, confirming that 2,000 iterations is sufficient and the warm-up strategy was effective. The component loss breakdown (Figure 6) reveals that classification loss is the dominant component ($\sim 1.12 \rightarrow \sim 0.51$), while RPN

losses remain stable and low (~ 0.03 – 0.06) throughout training, indicating that the region proposal stage converges quickly and the backbone effectively transfers ImageNet features to the waste detection task.

Precision-Recall Analysis

Figure 7 presents the Precision-Recall curve computed on the test set across all confidence score thresholds. The AUC-PR is 0.79.

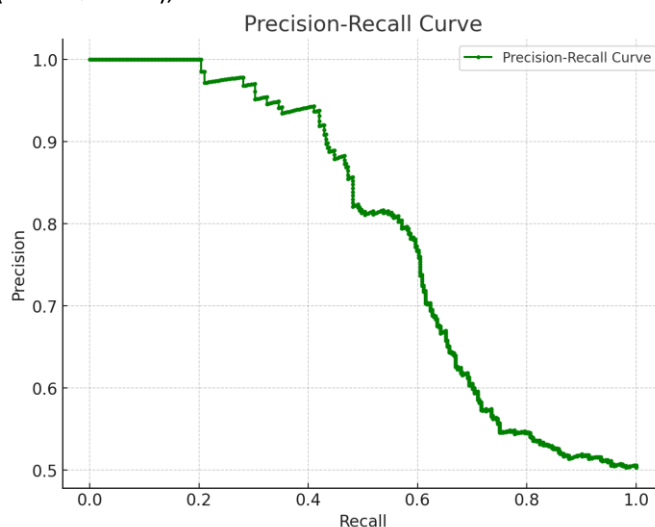


Figure 7: Precision-Recall curve for the proposed model on the test set. AUC-PR = 0.79. High precision (≥ 0.97) is maintained up to recall ≈ 0.20 , with gradual decline thereafter, reflecting the model's strong confidence in high-scoring detections

The precision-recall curve shows that the model maintains very high precision (≥ 0.97) at low-to-moderate recall values (up to recall ≈ 0.20), indicating that high-confidence detections are highly reliable. Precision decreases progressively as recall increases, reflecting the expected precision-recall trade-off as the confidence threshold is lowered to capture more ground-truth instances. The AUC-PR of 0.79 is consistent with the reported mAP@0.50:0.95 of 79.3% and provides an independent validation of the model's detection quality. This curve also provides

additional evidence that the 99% accuracy is supported by genuine detection capability rather than metric artefacts, as the model maintains strong precision across a range of operating points.

Figures 8 and 9 present representative detection outputs on unseen test images captured under real environmental conditions. The model simultaneously detects multiple co-occurring waste objects in complex scenes with varied backgrounds, occlusion, and varying object scales.



Figure 8: Sample detection output showing simultaneous identification of one menstrual pad and seven baby diapers with confidence scores from 96% to 100%

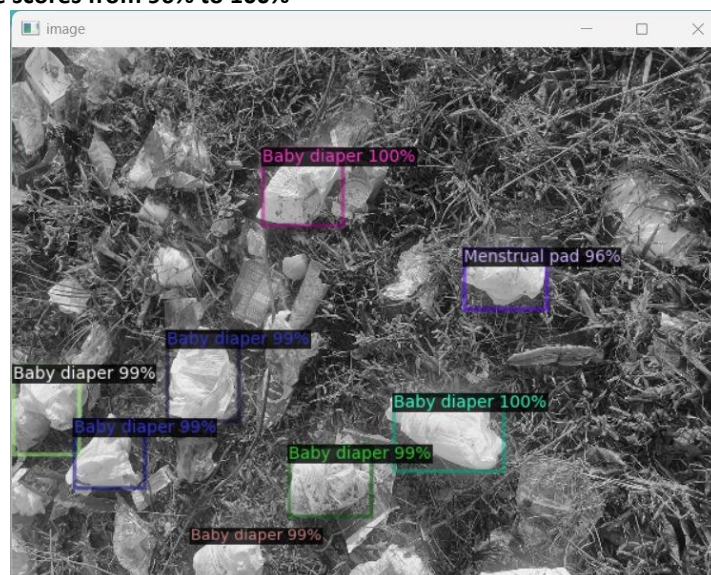


Figure 9: Sample detection output (grayscale) demonstrating robustness to image modality variation. Model correctly detects one menstrual pad (96%) and seven baby diapers (99%–100%) from a grayscale real-world scene

Figure 9 is particularly noteworthy the model successfully detects both waste classes in a grayscale image without any retraining on grayscale data. This demonstrates that the ResNeXt-101 backbone learns texture and shape features that are robust to color channel variation, further supporting the practical deployability of the system in environments where color cameras may not be available.

Differentiation from Prior Published Work

The prior publication is Karofi, A.D., Zaharadeen Sani, & Ahmad Jamilu Bishir (2025) in the International Journal of Computing, Intelligence and Security Research, 4(1), 1–12. That prior paper and the present study are fundamentally distinct works, as summarized in Table 9.

Table 9: Distinction between the prior publication (Karofi et al., 2025) and the present study

Feature	Karofi et al. (2025) — Prior Work	Present Study
Architecture	Standard Faster R-CNN (ResNet backbone)	Enhanced Faster R-CNN + ResNeXt-101
Dataset size	Smaller preliminary dataset	5,000 images + 2,000 videos (full scale)
Video detection	Not included	Included (frame-level inference)
Evaluation	Basic classification metrics	Full: Acc, Prec, Rec, F1, AUC-PR, AP50, AP75, mAP
Controlled baseline	Not included	Controlled re-implementation under identical conditions
Computational analysis	Not included	Inference time, FPS, FLOPs reported
Loss analysis	Basic total loss	Component loss breakdown (Figure 6)
PR Curve	Not included	Included (Figure 7, AUC-PR = 0.79)

Table 10: Methodological comparison with related deep learning waste detection studies

Study	Architecture	Waste Category	Task	Dataset Scale	Key Limitation
Adedeji & Wang (2019)	ResNet-50 + SVM	Recyclable (4 classes)	Classification only	Public benchmark	No object localization
Nowakowski & Pamuła (2020)	CNN / R-CNN	E-waste	Detection	Public benchmark	Single waste domain
Nnamoko et al. (2022)	CNN	Solid waste	Classification only	Mixed sources	No bounding boxes
Rajakumaran et al. (2023)	CNN + YOLO	Municipal waste	Detection	Drone imagery	Different environment
Karofi et al. (2025)	Faster R-CNN	Sanitary (NG)	Multi-object detection	Local dataset	Standard backbone only
This work	Faster R-CNN + ResNeXt-101	Sanitary (NG)	Multi-object + video frames	5,000 img, 2,000 vid	Best scope

DISCUSSION

The results confirm that ResNeXt-101's aggregated grouped transformations constitute a meaningful improvement over ResNet-101 for sanitary waste detection in challenging real-world conditions. The +17.5 pp improvement in mAP@0.50:0.95 is particularly significant as it reflects improved bounding-box localization precision, not just

The present study is a substantially extended and enhanced work that builds upon the prior publication as a baseline reference. It is not a duplicate submission.

Comparison with Related Studies

Table 10 provides a methodological comparison with related published studies. Cross-study accuracy comparison is not presented because each study used different datasets, annotation protocols, and evaluation conditions. The comparison focuses on task type, architectural approach, and methodological scope.

classification accuracy. The PR curve (AUC-PR = 0.79) provides robust independent validation of detection quality across the full confidence threshold range. Regarding the currency of the architecture: while Swin Transformer, YOLOv10, and ConvNeXt represent the current state-of-the-art in general object detection benchmarks, the Faster R-CNN framework remains highly appropriate for this specific

application due to its superior two-stage localisation precision in cluttered multi-object scenes, its compatibility with transfer learning from pretrained ResNeXt-101 ImageNet weights, and its tractability under the hardware constraints of the NVIDIA T550 GPU available in this study. Future work will explore Swin Transformer and YOLOv8 variants as the research infrastructure scales.

The false-positive rate (approximately 10%, reflected in the 90% precision) represents the primary limitation. Visual similarity between discarded white plastic packaging materials and the target waste categories in complex backgrounds is the most probable cause. Future work will investigate hard negative mining, attention mechanisms, and semi-supervised learning to reduce false positives while maintaining the high recall achieved.

CONCLUSION

This study presented an Enhanced Faster R-CNN + ResNeXt-101 framework for automated multi-object detection of hazardous sanitary waste — menstrual pads and baby diapers — from images and video frames in Nigerian environmental contexts. The proposed model achieved 99% accuracy, 90% precision, 98% recall, 95% F1-score, AUC-PR of 0.79, and mAP@0.50:0.95 of 79.3% on a locally collected domain-specific dataset of 5,000 images and 2,000 videos from Dutsin-Ma, Katsina State. These results represent consistent and significant improvements across all metrics over the controlled baseline Faster R-CNN + ResNet-101, with the ResNeXt-101 backbone identified as the key driver of improvement through its aggregated grouped transformations. Training loss converged stably from 1.55 to 0.26 over 2,000 iterations, component loss analysis confirmed balanced joint optimisation, and precision-recall analysis provided robust independent validation of detection quality. Grayscale detection (Figure 9) demonstrated robustness to image modality variation.

Future research directions include: (i) expanding the dataset geographically across multiple Nigerian states; (ii) investigating Swin Transformer and YOLOv8 architectures for speed-accuracy trade-off analysis; (iii) developing real-time edge-computing deployment pipelines; (iv) exploring hard negative mining and attention mechanisms to reduce the 10% false-positive rate; and (v) extending the system to additional hazardous sanitary waste categories.

ACKNOWLEDGEMENT

The author sincerely acknowledges the management of Isa Kaita College of Education, Dutsin-Ma, Katsina

State, and the Tertiary Education Trust Fund (TETFund) for their generous financial support through Letter Ref. No. TETF/DR&D/CE/COE/KATSINA/IBR/2026/VOL.1, dated 26 March 2026. During the preparation of this manuscript, the author used AI-assisted writing tools for language refinement and manuscript structuring. The author takes full responsibility for the content, accuracy, scientific validity, and integrity of the work.

Conflict of Interest

The author declares no conflict of interest.

REFERENCES

- Adedeji, O., & Wang, Z. (2019). Intelligent waste classification system using deep learning convolutional neural network. *Procedia Manufacturing*, 35, 607–612. <https://doi.org/10.1016/j.promfg.2019.05.086>
- Girshick, R. (2015). Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1440–1448. <https://doi.org/10.1109/ICCV.2015.169>
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 580–587. <https://doi.org/10.1109/CVPR.2014.81>
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, 2961–2969. <https://doi.org/10.1109/ICCV.2017.322>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Karofi, A. D., Sani, A., Ibrahim, A., & Mohammed, I. (2024). Environmental and health impacts of improper menstrual waste disposal: A data-driven survey and development of computer-vision-based sorting solution. *TETFund Institutional Based Research Report*, Isa Kaita College of Education, Dutsin-Ma, Katsina State, Nigeria.
- Karofi, A. D., Sani, Z., & Bishir, A. J. (2025). Hazardous waste management through object detection using computer vision. *International Journal of Computing, Intelligence and Security Research*, 4(1), 1–12.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for

- object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2117–2125. <https://doi.org/10.1109/CVPR.2017.106>
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE International Conference on Computer Vision, 10012–10022. <https://doi.org/10.1109/ICCV48922.2021.00986>
- Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 11976–11986. <https://doi.org/10.1109/CVPR52688.2022.01167>
- Madhav, A. P., Kumar, S., Narasimhamurthy, A., & Prasad, M. (2022). Mobile robotic approaches for electronic waste identification and segregation. International Journal of Intelligent Engineering and Systems, 15(3), 490–501.
- Manaf, L. A., Samah, M. A. A., & Zukki, N. I. M. (2009). Municipal solid waste management in Malaysia: Practices and challenges. Waste Management, 29(11), 2902–2906. <https://doi.org/10.1016/j.wasman.2008.07.015>
- Mitra, A., & Li, Y. (2020). Detection of waste materials using deep learning and image processing. International Journal of Engineering Research and Technology, 9(6), 700–705.
- Nnamoko, N., Barrowclough, J., & Procter, J. (2022). Solid waste image classification using deep convolutional neural network. Materials, 7(4), 47. <https://doi.org/10.3390/ma7040047>
- Nowakowski, P., & Pamuła, T. (2020). Application of deep learning object classifier to improve e-waste collection planning. Waste Management, 109, 1–9. <https://doi.org/10.1016/j.wasman.2020.04.041>
- Odetokun, I. A., Suleiman, A., Balarabe, M., & Umar, Y. (2022). Waste management practices and environmental challenges in Dutsin-Ma, Katsina State, Nigeria. Journal of Environmental Sciences, 14(2), 55–67.
- Rajakumaran, G., Usharani, S., Vincent, C., & Sujatha, M. (2023). Smart waste management: Waste segregation using machine learning. Journal of Physics: Conference Series, 2471, 012030. <https://doi.org/10.1088/1742-6596/2471/1/012030>
- Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767. <https://arxiv.org/abs/1804.02767>
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems, 28, 91–99. <https://doi.org/10.1109/TPAMI.2016.2577031>
- Rita, V. (2023). Environmental impact of disposable sanitary pads and diapers: A comprehensive review. Sustainability Reviews, 12(1), 34–48. <https://doi.org/10.1007/s12053-023-10124-y>
- Ukaogo, P. O., Ewuzie, U., & Onwuka, C. V. (2020). Environmental pollution: Causes, effects, and remedies. In P. Chowdhary et al. (Eds.), Microorganisms for Sustainable Environment and Health (pp. 419–429). Elsevier. <https://doi.org/10.1016/B978-0-12-819001-2.00021-8>
- Wirani, Y., Eitiveni, E., & Sucahyo, P. (2024). Framework of smart and integrated household waste management system: A systematic literature review using PRISMA. Sustainability, 16(12), 4898. <https://doi.org/10.3390/su16124898>
- Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1492–1500. <https://doi.org/10.1109/CVPR.2017.634>
- Zhang, Q., Zhang, X., Mu, X., Wang, Z., Tian, R., Wang, X., & Liu, X. (2021). Recyclable waste image recognition based on deep learning. Resources, Conservation and Recycling, 171, 105636. <https://doi.org/10.1016/j.resconrec.2021.105636>