

Analysis and Comparative Review of CNN and Vision Transformer Based Models for Intelligent Waste Management

¹Sapna Pandey,²Neelesh Shrivastava

¹Department of Computer Science & Engineering, Aditya College of Technology and Science, Satna, M.P., India
sapna.pandey0001999@gmail.com

²Head of Department, Department of Computer Science & Engineering, Aditya College of Technology and Science, Satna, M.P., India
shrivastava.neelesh@gmail.com

Abstract—

Rapid urbanisation and rising municipal solid waste generation have made intelligent waste segregation an essential requirement for modern smart cities. Conventional classification techniques depend on manual sorting and are limited by cost, inconsistency and poor adaptability to complex environmental conditions. Deep learning has produced substantial improvements in automated waste recognition, but existing approaches frequently rely on single-model architectures, narrow datasets, or computationally expensive frameworks that restrict real-time deployment. This paper presents a comprehensive review of hybrid deep learning approaches for intelligent waste classification and smart waste management. Thirty-one representative studies published between 2016 and 2026 are organised into a four-generation taxonomy spanning classical convolutional networks, transfer and ensemble learning, vision transformers and hybrid architectures, and detection-oriented and deployable systems. The benchmark datasets used by the field are compared, the principal backbone architectures are contrasted by depth, parameter count and characteristic behaviour, and reported performance is analysed across the surveyed systems. The review shows that reported accuracy on standard corpora has risen from the low nineties for early convolutional systems to approximately ninety-nine per cent for recent fused convolutional-transformer designs, but that this progress is measured almost entirely on curated laboratory imagery. Five persistent gaps are identified — the absence of global context in single-CNN designs, limited dataset diversity and class imbalance, the separate treatment of classification and detection, computational cost that precludes edge deployment, and the absence of explainability — and each is mapped to a corresponding component of a proposed five-layer hybrid framework that integrates convolutional and transformer feature extraction, attention-based fusion, unified recognition and detection, and an explainable decision layer for smart city waste management.

Keywords—waste classification, deep learning, hybrid CNN, vision transformer, transfer learning, YOLO, smart waste management, intelligent recycling, computer vision, sustainable smart cities.

I. INTRODUCTION

Rapid growth in urban population, industrialisation and changing consumption patterns have produced an unprecedented increase in municipal solid waste. According to World Bank estimates, approximately 1.33 of every 2.01 billion tonnes of municipal solid waste generated worldwide is not managed sustainably, and global waste production is projected to reach 3.40 billion tonnes by 2050, with daily output in low- and middle-income countries expected to rise by forty per cent or more [27]. Improper segregation and inefficient recycling contribute directly to pollution, greenhouse gas emission and depletion of natural resources, which places effective waste management among the primary objectives of sustainable smart city development.

Traditional segregation depends on manual sorting, which is labour-intensive, slow, costly and error-prone. These limitations reduce recycling efficiency and raise operational cost, and they scale badly: the volume and heterogeneity of the waste stream in a large city exceed what manual inspection can process reliably. Researchers have therefore turned to artificial intelligence and computer vision to automate classification and support intelligent recycling.

Deep learning has become the dominant approach to image-based waste recognition. Convolutional neural networks including AlexNet, ResNet, EfficientNet and DenseNet extract discriminative visual features from waste images with high accuracy [5], [6], [13]. More recently, vision transformers, attention mechanisms, ensemble learning and transfer learning have improved recognition

by capturing local and global context jointly [10], [12], [29], while real-time detectors such as YOLOv8 and YOLOv12 have enabled automatic localisation and smart garbage monitoring through surveillance cameras and IoT-enabled bins [14], [15].

Despite these advances, practical challenges persist. Many models are trained on small or imbalanced datasets, which limits generalisation to diverse real environments. Cluttered backgrounds, varying illumination, occlusion and inter-class visual similarity continue to degrade accuracy. Several high-performing architectures require substantial computation, which makes deployment on edge devices and embedded smart waste systems difficult [33]. Recent work has responded with hybrid architectures that combine convolutional feature extraction, transfer learning, attention and detection networks into a single design [12], [29], [31].

This paper contributes the following. First, it organises the literature into a four-generation taxonomy and situates each surveyed study within it. Second, it compares the benchmark datasets and the backbone architectures on which the field depends, which is the information a practitioner needs to reproduce or extend any of the surveyed results. Third, it analyses reported performance across the surveyed systems and shows where the accuracy trend has reached. Fourth, it identifies five persistent research gaps and maps each to a corresponding component of a proposed hybrid framework, so that the proposal is traceable to the evidence rather than asserted.

Section II presents the taxonomy. Section III reviews the literature by generation. Section IV compares datasets and backbones. Section V analyses reported performance.

Section VI states the research gaps and the proposed framework, and Section VII concludes with directions for future work.

II. TAXONOMY OF WASTE CLASSIFICATION RESEARCH

Fig. 1 organises the field into four generations. The first applies a single convolutional network to a curated corpus. The second adds ImageNet pre-training, fine-tuning and ensembling of several backbones, together with explicit handling of class imbalance. The third introduces transformer architectures and convolutional-transformer hybrids that model global context. The fourth shifts the objective from classification to detection and deployment, covering YOLO-based localisation, IoT smart bins, edge inference, federated training and explainable output.

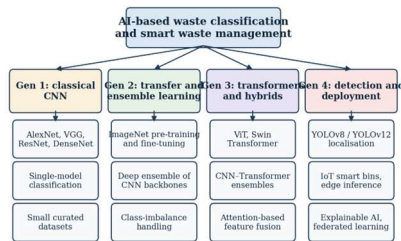


Fig. 1. Four-generation taxonomy of AI-based waste classification and smart waste management.

The generations are cumulative rather than exclusive. A fourth-generation detector still selects its backbone from the first three generations, and the most recent hybrid systems combine a convolutional stem, a transformer stage and a detection head in one model. What distinguishes a contemporary system is therefore the combination of components it selects, which is the view adopted throughout Section III.

III. LITERATURE REVIEW

A. Deep Learning Based Waste Classification

Sayem et al. introduced a robust deep learning framework for waste classification and object detection that improved recycling efficiency through comprehensive feature extraction and multi-class recognition [1]. Kaya proposed a smart garbage classification model for sustainable development, demonstrating the practical application of convolutional networks to intelligent segregation [2]. Navyasree et al. developed an AI-based garbage classification and disposal framework that integrates image classification with an automated disposal mechanism for smart city applications [3]. Kunwar introduced the Garbage Dataset, a large-scale multi-class benchmark designed to improve generalisation and support future research [4]. Bhasha et al. applied the AlexNet architecture with transfer learning and reported competitive recognition accuracy [5], and Kaur and Jaswal used ResNet50-based transfer learning to address class imbalance, improving performance on imbalanced garbage corpora [6]. Alkilinc et al. proposed a deep ensemble model combining several convolutional architectures to enhance robustness and reduce classification error [7], and Ijaz et al. presented GlobalWasteData, a large-scale integrated benchmark supporting robust classification and environmental

monitoring across diverse geographical conditions [8]. Collectively these studies show that convolutional networks, transfer learning, ensembling and large annotated corpora have substantially advanced automated classification, while complex backgrounds, inter-class similarity and real-time deployment remain unsolved.

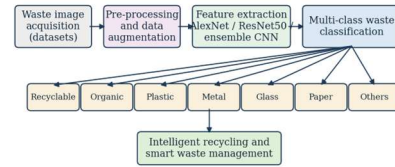


Fig. 2. General workflow of convolutional waste classification systems [1]–[9].

Fig. 2 summarises the workflow these systems share. Acquisition draws waste images from public corpora; pre-processing and augmentation standardise and expand them; feature extraction is performed by AlexNet, ResNet50 or an ensemble of convolutional backbones; and the resulting representation drives multi-class classification into recyclable, organic, plastic, metal, glass, paper and residual categories, whose output feeds intelligent recycling and smart waste management.

B. Vision Transformers, Object Detection and Smart Monitoring

Tanwer et al. evaluated vision transformer models for garbage bin fullness detection and reported superior performance in monitoring container capacity for efficient collection scheduling [10]. Koroğlu et al. investigated garbage container detection using deep object detection and presented a practical monitoring framework for urban environments [11]. Kumar et al. proposed a hybrid convolutional and transformer ensemble that combines local spatial features with global context, improving classification accuracy [12]. Muangnak et al. performed a comparative study of several deep models, highlighting the strengths and limitations of different convolutional architectures [13]. Chen et al. developed an enhanced YOLOv12 framework for construction waste detection, achieving accurate localisation at reduced computational cost [14], and Alsaihati et al. applied YOLOv8 to municipal waste detection and classification in real time [15]. Dixit et al. extended monitoring to aquatic environments with AI-based detection of floating garbage [16].

Recent work has pushed the hybrid direction further. Madhavi et al. proposed SwinConvNeXt, which fuses a Swin Transformer with a ConvNeXt backbone and a spatial attention mechanism for real-time garbage image classification over twelve categories, reporting 98.97% accuracy, 98.42% precision and 98.61% recall — currently among the strongest published figures for this task [29]. A dual-attention detector combining the convolutional block attention module with contextual transformer networks, coupled with DeepSORT tracking, has been proposed for real-time identification and tracking of solid waste, and evaluated on a purpose-built multi-label mixed-waste corpus alongside the standard benchmark [30]. Federated learning has been applied to trash classification in smart cities, comparing ten deep architectures on a large

benchmark and identifying ResNeXt-101 as the strongest [31], and reinforcement learning with generative augmentation and hyper-parameter optimisation has been used to address overfitting on small corpora [32]. An automated classification system reported by Alsubaei et al. surveys and extends this line, noting compact architectures that reach 96.10% on the standard benchmark despite very small parameter counts [33].

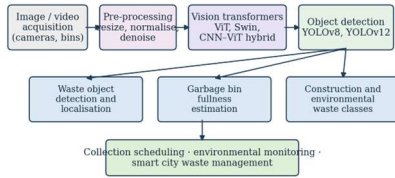


Fig. 3. Transformer and detection based smart waste monitoring architecture [10]–[16].

Fig. 3 presents the corresponding architecture. Acquisition draws image or video from surveillance cameras or IoT-enabled bins; pre-processing normalises the input; transformer and detection models then jointly detect waste objects, estimate bin fullness, localise materials and classify construction or environmental waste; and the processed information is transmitted for collection scheduling, environmental monitoring and smart city management.

C. Deep Learning Applications and Research Motivation

Deep learning has succeeded across computer vision applications beyond waste management, and that experience informs the design of a robust classification system. Nagar and Singh developed a secure AI-powered penalty detection and recommendation system for sports analytics [17]. Shakyawar and Singh proposed a facial expression recognition framework capable of analysing crowd emotion [18]. Prajapati et al. introduced a deep learning gender recognition model [19], and Gupta et al. applied ResNet152 to face mask recognition during pandemic monitoring [20]. Singh et al. applied machine learning and deep learning to intelligent job recommendation [21]. Vishwakarma et al. presented an EfficientNetB3 framework for brain tumour segmentation and detection [22], and Ray and Singh employed a hybrid task cascade region-based convolutional network for sports analytics [23]. Rajput et al. analysed cloud service providers for performance and availability, which bears on the scalable infrastructure such systems require [24]. Patel and Singh proposed age and gender recognition [25], and Singh and Songare applied GoogLeNet to monkeypox detection [26].

Further work by the same group reinforces the point across architectures and modalities: transfer-learning-based object detection with convolutional networks [34], convolutional glaucoma detection from retinal images [35], cancer prediction combining random forest with deep learning [36], face mask identification using OpenCV [37], and activity detection and people counting with Mask-RCNN and bidirectional ConvLSTM [38]. Taken together these studies confirm that transfer learning, EfficientNet, ResNet, GoogLeNet and hybrid architectures deliver

consistently strong performance across domains, which motivates their adoption for waste classification.

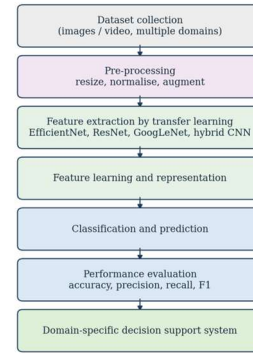


Fig. 4. Generalised deep learning pipeline for intelligent computer vision applications [17]–[26].

TABLE I. Systematic Literature Review on AI-Based Waste Classification and Smart Waste Management

Ref.	Year	Method / model	Dataset	Major contribution	Limitation / gap
[1]	2025	Deep CNN detection and classification	Multi-class waste	Improved classification and detection accuracy	High computational cost; limited real-time use
[2]	2023	CNN smart garbage classification	Smart waste images	Intelligent segregation framework	Degrades under complex conditions
[3]	2026	AI garbage classification and disposal	Garbage images	Classification with automated disposal	Limited dataset diversity
[4]	2026	Garbage Dataset (GD) benchmark	Large-scale garbage	Standardised multi-class benchmark	Requires expansion to real scenarios
[5]	2026	AlexNet transfer learning	Garbage images	Effective multi-class classification	Weaker features than modern backbones
[6]	2026	ResNet50 transfer learning	Imbalanced garbage	Improved imbalanced-class performance	Higher training cost
[7]	2025	Deep ensemble CNN	Waste classification	Robustness through ensembling	Larger model, higher latency
[8]	2026	GlobalWasteData	Global waste images	Integrated cross-region dataset	Validation across more regions needed
[10]	2025	Vision Transformer	Bin fullness images	Accurate fullness estimation	High computational requirement
[11]	2025	Deep object detection	Container images	Real-time container detection	Weak under severe occlusion
[12]	2025	CNN–Transformer ensemble	Waste images	Local and global feature fusion	Model complexity and latency
[13]	2021	Comparative CNN study	Garbage dataset	Benchmarked several architectures	Classical CNNs generalise poorly
[14]	2026	Enhanced YOLOv12	Construction waste	Fast, accurate localisation	Construction categories only
[15]	2025	YOLOv8	Municipal waste	Real-time detection and localisation	Affected by cluttered backgrounds
[16]	2026	AI water garbage detection	Water body images	Floating garbage detection	Aquatic scenes only
[29]	2025	SwinConvNeXt fusion	Garbage, 12 classes	98.97% accuracy, real-time	Single-corpus validation
[30]	2023	Dual-attention YOLO + DeepSORT	Mixed solid waste	Detection with real-time tracking	Purpose-built corpus
[31]	2024	Federated deep learning	TrashBox	Privacy-preserving smart-city training	Communication overhead

Ref.	Year	Method / model	Dataset	Major contribution	Limitation / gap
[32]	2026	Reinforcement learning + GAN augmentation	Benchmark corpus	Addresses small-data overfitting	Training instability risk
[33]	2025	Automated classification survey and model	Multiple corpora	Compact models at 96.10% accuracy	Laboratory imagery only

Table I summarises the surveyed work. Early studies employed conventional convolutional networks, while recent research has moved to transfer learning, ensembling, vision transformers and advanced detectors such as YOLOv8 and YOLOv12. Large-scale corpora including the Garbage Dataset and GlobalWasteData have improved generalisation and benchmarking. Nevertheless, computational complexity, deployment on resource-constrained devices, class imbalance, environmental variability and robustness under complex backgrounds remain unresolved, and most studies address classification or detection independently rather than integrating recognition, localisation, monitoring and decision support into one framework.

IV. BENCHMARK DATASETS AND BACKBONE ARCHITECTURES

A. Datasets

Reproducibility in this field depends on a small number of public corpora, compared in Table II.

TABLE II. Benchmark Datasets Used in the Surveyed Literature

Dataset	Images	Classes	Characteristics
TrashNet [28]	2,527	6	Uniform background, laboratory capture; the standard baseline
TACO [39]	Not stated	Multiple	Litter in context; fewer annotations than TrashNet
TrashBox [31]	Not stated	Multiple	Used for transfer-learning and federated benchmarking
Garbage Dataset (GD) [4]	Large scale	Multi-class	Standardised benchmark for segregation research
GlobalWasteData [8]	Large scale	Multiple	Cross-region integration for environmental monitoring
Garbage Classification [29]	Not stated	12	Paper, cardboard, biological, metal, plastic, glass variants, clothes, shoes, batteries, trash

Two observations follow. First, the standard baseline is small: TrashNet contains 2,527 images across six classes, which is roughly two orders of magnitude below the corpora on which the backbones themselves were pre-trained, and this is precisely why transfer learning rather than training from scratch dominates the literature [5], [6]. Second, the most widely used corpora are captured under controlled conditions with uniform backgrounds, whereas operational imagery contains clutter, occlusion, variable illumination and weather effects. The recent dataset contributions [4], [8] are a direct response to this, and the extent to which they close the gap is the most consequential open empirical question in the field.

B. Backbone Architectures

Table III contrasts the architectures used across the four generations.

TABLE III. Comparison of Backbone Architectures

Backbone	Depth	Parameters	Characteristic behaviour
AlexNet	8 layers	≈60 M	Early deep CNN; heavy fully connected block
GoogLeNet	22 layers	≈4 M	Inception modules capture multiple scales

Backbone	Depth	Parameters	Characteristic behaviour
ResNet50	50 layers	≈25 M	Residual shortcuts enable depth
DenseNet	121–201	≈8–20 M	Dense connectivity, feature reuse
EfficientNet	scaled	≈5–66 M	Compound scaling; best CNN efficiency
ViT / Swin	12–24 blocks	≈86 M+	Global self-attention; data hungry
CNN–Transformer hybrid	variable	variable	Convolutional bias plus global context
YOLOv8 / YOLOv12	variable	compact	Single-stage detection and localisation

The table explains why hybrids have come to dominate. A convolutional backbone carries a strong locality prior and therefore learns efficiently from a small corpus, but its receptive field grows only with depth, so global relationships between distant regions of a cluttered scene are captured late if at all. A transformer attends globally from the first layer but requires substantially more data or stronger augmentation. Placing a convolutional stem in front of an attention stack retains the inductive bias while adding global context, which is the design that SwinConvNeXt [29] and the CNN–Transformer ensemble of Kumar et al. [12] both adopt, and which the framework of Section VI generalises.

V. COMPARATIVE PERFORMANCE ANALYSIS

Fig. 5 plots reported performance across representative surveyed systems.

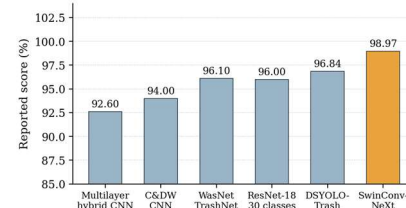


Fig. 5. Reported performance of representative waste classification systems. Corpora and metrics differ, so the comparison is indicative.

The trend is clear. A multilayer hybrid convolutional network reaches 92.60% on the standard benchmark [31], a convolutional classifier for construction and demolition waste reaches 94% [33], the compact WasNet architecture reaches 96.10% [33], ResNet-18 reaches 96% over thirty waste object classes, a metaheuristic transfer-learning detector reaches 96.84% precision [33], and the fused SwinConvNeXt architecture reaches 98.97% accuracy [29]. Across roughly a decade of architectural development the reported figure has risen by about six points and is now approaching the ceiling of the benchmark.

Two cautions apply. First, these figures come from different corpora, different class counts and in one case a precision rather than an accuracy figure, so the ordering is indicative rather than strict; controlled benchmarking under a fixed partition and preprocessing protocol would be required to establish it. Second, and more importantly, every figure plotted is obtained on curated imagery. The scoping literature on real-world deployment records markedly lower and more variable performance once background clutter, occlusion and illumination variation enter the input, and compact models such as WasNet score 64.5% on a general-purpose corpus against 96.10% on the waste benchmark [33] — a gap that measures dataset specificity rather than model quality.

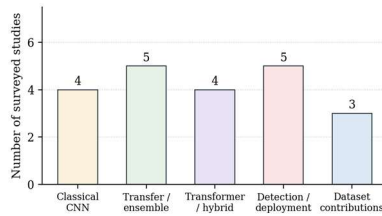


Fig. 6. Distribution of the surveyed studies across technique families.

Fig. 6 shows the distribution of surveyed studies. No family has displaced the others: classical convolutional work remains present alongside transformer and detection research, and dataset contributions form a distinct and growing category of their own. That pattern is consistent with the accuracy trend. Where architectural gains on the benchmark have become small, effort migrates toward the factors that actually limit deployment — data diversity, inference cost and integration — which is exactly what the fourth generation of Fig. 1 represents.

VI. RESEARCH GAPS AND PROPOSED FRAMEWORK

A. Identified Gaps

- Absence of global context. Most studies employ a single convolutional architecture such as AlexNet, ResNet or EfficientNet, which captures local spatial features while overlooking global context. Vision transformers address this but require larger corpora and greater computation, and few studies integrate the two into a unified lightweight design.
- Limited dataset diversity and class imbalance. Many public corpora contain controlled laboratory images with uniform backgrounds, whereas real waste imagery varies in illumination, occlusion, background clutter, scale and weather, so trained models generalise poorly across operational environments.
- Separate treatment of classification and detection. Current research addresses recognition or localisation independently; few studies simultaneously handle classification, localisation, bin monitoring, recycling decision support and collection scheduling within one framework.
- Computational cost. Several high-performing models demand resources that preclude deployment on embedded systems, IoT devices and edge platforms, and lightweight yet accurate hybrid architectures remain largely unexplored.
- Absence of explainability. Municipal authorities and environmental agencies cannot interpret the decision process of current models, and the integration of explainable artificial intelligence with waste classification remains an open direction.

B. Proposed Hybrid Framework

Fig. 7 presents a five-layer hybrid framework in which each gap identified above is addressed by a specific component.

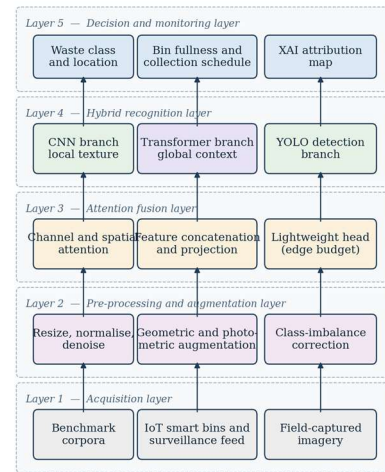


Fig. 7. Proposed five-layer hybrid deep learning framework for intelligent waste classification and smart waste management.

Layer 1 acquires input from benchmark corpora, IoT-enabled smart bins and field-captured imagery, so that the model is exposed to operational as well as curated data. Layer 2 performs pre-processing, geometric and photometric augmentation and explicit class-imbalance correction. Layer 3 fuses features through channel and spatial attention, concatenation and projection into a lightweight head sized to an edge inference budget. Layer 4 performs recognition through three parallel branches — a convolutional branch for local texture, a transformer branch for global context and a detection branch for localisation — so that classification and detection are produced by one model rather than two. Layer 5 emits the waste class and location, the bin fullness estimate and collection schedule, and an attribution map that identifies the image regions responsible for the decision.

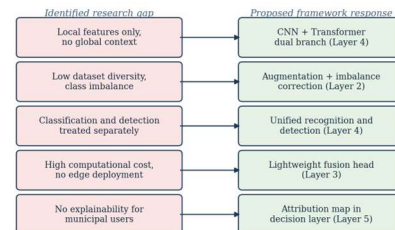


Fig. 8. Mapping of identified research gaps to components of the proposed framework.

Fig. 8 makes the correspondence explicit. Each of the five gaps of Section VI-A is answered by a named layer, so the framework is traceable to the review evidence rather than assembled from components chosen for novelty. This traceability is also the basis on which the framework should be evaluated: a validation study must show not merely that the combined model scores well, but that each component measurably closes the gap it was introduced to address.

VII. CONCLUSION AND FUTURE WORK

Intelligent waste classification has become an essential component of sustainable waste management and smart city development. This paper reviewed thirty-one representative studies published between 2016 and 2026, organised into a four-generation taxonomy spanning classical convolutional networks, transfer and ensemble

learning, vision transformers and hybrids, and detection-oriented and deployable systems. The benchmark datasets were compared and shown to be small and predominantly laboratory-captured, the principal backbones were contrasted by depth, parameter count and characteristic behaviour, and reported performance was analysed and shown to have risen from the low nineties to approximately ninety-nine per cent on curated imagery. Five persistent gaps were identified and each was mapped to a component of a proposed five-layer hybrid framework integrating convolutional and transformer feature extraction, attention-based fusion, unified recognition and detection, and an explainable decision layer.

Future research should develop lightweight hybrid models that retain classification accuracy while meeting the memory and power budget of edge devices and IoT-enabled smart bins. Vision transformers, attention mechanisms and multimodal sensor data can improve recognition under challenging environmental conditions. Large-scale, globally representative corpora captured in operational rather than laboratory conditions are needed to establish real generalisation, and controlled benchmarking protocols reporting repeated-run variance would allow genuine methodological progress to be distinguished from partitioning noise. Explainable artificial intelligence should be incorporated to improve transparency and user trust, and federated formulations [31] offer a route to training across municipalities without centralising imagery. Finally, combining classification with robotic sorting, autonomous collection, digital twins and reinforcement-learning-based route optimisation would extend the contribution from recognition to fully intelligent, real-time and sustainable waste management for future smart cities.

REFERENCES

- [1] F. R. Sayem, M. S. B. Islam, M. Naznine, M. Nashbat, M. Hasan-Zia, A. K. A. Kunju, and M. E. Chowdhury, "Enhancing waste sorting and recycling efficiency: Robust deep learning-based approach for classification and detection," *Neural Computing and Applications*, vol. 37, no. 6, pp. 4567–4583, 2025.
- [2] V. Kaya, "Classification of waste materials with a smart garbage system for sustainable development: A novel model," *Frontiers in Environmental Science*, vol. 11, Art. no. 1228732, 2023.
- [3] N. Navyasree, U. Navadeep, V. Haritha, C. Suresh, and N. A. Kumar, "AI-based intelligent garbage classification and disposal system," in *Proc. 6th Int. Conf. Trends in Material Science and Inventive Materials (ICTMIM)*, 2026, pp. 1174–1180.
- [4] S. Kunwar, "The Garbage Dataset (GD): A multi-class image benchmark for automated waste segregation," *arXiv:2602.10500*, 2026.
- [5] P. Bhasha, S. K. Sameera, P. C. Raj, P. S. K. Reddy, T. Ujwal, and M. Y. Noor, "Garbage classifier through AlexNet architecture," in *AIP Conf. Proc.*, vol. 3371, no. 1, Art. no. 030005, 2026.
- [6] N. Kaur and N. Jaswal, "Toward efficient smart waste management: ResNet50 based transfer learning for imbalanced garbage classification," in *Proc. Int. Conf. Smart Futuristic Technology*, 2026, pp. 1–6.
- [7] A. Alkılınc, F. Y. Okay, İ. Kök, and S. Özdemir, "Deep ensemble learning model for waste classification systems," *Sustainability*, vol. 18, no. 1, Art. no. 24, 2025.
- [8] M. Ijaz, S. U. R. Khan, A. U. Rehman, T. Asif, S. Vollmer, A. Dengel, and M. N. Asim, "GlobalWasteData: A large-scale, integrated dataset for robust waste classification and environmental monitoring," *arXiv:2602.07463*, 2026.
- [9] M. Yang and G. Thung, "Classification of trash for recyclability status," CS229 Project Report, Stanford University, 2016.
- [10] P. S. Tanwer, S. Maheshwari, S. Behera, A. Chauhan, and T. Sunil Kumar, "Evaluation of vision transformers for the detection of fullness of garbage bins for efficient waste management," *Frontiers in Artificial Intelligence*, vol. 8, Art. no. 1612080, 2025.
- [11] F. E. Köroğlu, S. Çakmak, M. E. Yurtsever, and S. Eken, "Smart waste management: A case study on garbage container detection," in *Pattern Recognition and Artificial Intelligence (MedPRAI 2024)*, Cham, Switzerland: Springer, 2025, pp. 881–890.
- [12] T. M. Kumar, V. R. Rajeswari, J. N. Kumar, P. S. Kumar, and P. Sai, "Optimizing garbage classification through CNN and transformer-based model ensembling," 2025.
- [13] N. Muangnak, C. Sirawattananon, M. G. Oza, and R. Sukthanapirat, "Comparative study considering garbage classification using in-depth learning techniques," in *Proc. Int. Conf. Computing and Information Technology*, Cham, Switzerland: Springer, 2021, pp. 169–180.
- [14] H. Chen, C. Liu, S. Qiao, J. Lei, T. Yu, and W. Lin, "Construction garbage classification detection using an enhanced YOLOv12 framework," in *Proc. 2nd Int. Conf. Electrical Automation and Artificial Intelligence (ICEAAI)*, 2026, pp. 1–4.
- [15] F. Alsaihati, H. Aldossary, R. Alzamil, R. Almadan, Z. Al Mousa, and A. Alahmadi, "Waste classification and detection model using YOLOv8 for waste management," in *Integrating Big Data and IoT for Enhanced Decision-Making Systems in Business*, vol. 2, Cham, Switzerland: Springer, 2025, pp. 473–484.
- [16] A. Dixit, V. Jadhav, P. Sadare, S. G. Patil, S. Kampli, and T. R. Goudar, "Detection of garbage in water bodies using AI based techniques," *Int. J. Sci. R. Tech.*, vol. 3, no. 5, 2026.
- [17] K. Nagar and U. Singh, "ProSoccerTrainer: A secure AI-powered penalty detection and performance recommendation system," in *Proc. 5th Intelligent Cybersecurity Conf. (ICSC)*, 2025, pp. 248–255.
- [18] P. Shakyawar and U. Singh, "Facial expression recognition using deep learning model on crowd," in *Proc. Int. Conf. Advances in Computing Research on Science Engineering and Technology (ACROSET)*, 2024, pp. 1–6.
- [19] A. Prajapati, N. Patel, and U. Singh, "Gender recognition using deep learning methodology," in *Proc. 2nd Int. Conf. Automation, Computing and Renewable Systems (ICACRS)*, 2023, pp. 798–804.
- [20] A. Gupta, P. Tripathi, B. P. Singh, and U. Singh, "Human face mask recognition using ResNet 152 model," in *Proc. 2nd Int. Conf. Automation, Computing and Renewable Systems (ICACRS)*, 2023, pp. 1503–1510.
- [21] D. Singh, N. Patel, and U. Singh, "Method for job recommendation based on machine learning and deep learning model," in *Proc. 2nd Int. Conf. Automation, Computing and Renewable Systems (ICACRS)*, 2023, pp. 875–883.
- [22] M. Vishwakarma, P. Tripathi, B. P. Singh, and U. Singh, "Segmentation and detection of brain tumors using EfficientNetB3 model," in *Proc. 2nd Int. Conf. Automation, Computing and Renewable Systems (ICACRS)*, 2023, pp. 1463–1470.
- [23] S. Ray and U. Singh, "Hybrid task cascade region-based convolutional neural network is used to analyse cricket scores," in *Proc. Int. Conf. Self Sustainable Artificial Intelligence Systems (ICSSAS)*, 2023, pp. 19–26.
- [24] A. Rajput, P. Gupta, P. Ghodeswar, S. Varma, K. K. Sharma, and U. Singh, "Study of cloud providers (Azure, Amazon, and Oracle) according to service availability and price," in *Proc. 3rd Int. Conf. Pervasive Computing and Social Networking (ICPCSN)*, 2023, pp. 1177–1188.
- [25] M. Patel and U. Singh, "Age and gender recognition using deep learning technique," in *Proc. 3rd Int. Conf. Smart Data Intelligence (ICSDMI)*, 2023, pp. 238–245.
- [26] U. Singh and L. S. Songare, "Analysis and detection of monkeypox using the GoogLeNet model," in *Proc. Int. Conf. Automation, Computing and Renewable Systems (ICACRS)*, 2022, pp. 1000–1008.
- [27] World Bank, *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*. Washington, DC, USA: World Bank, 2023.
- [28] M. Yang and G. Thung, "TrashNet dataset," Stanford University, 2016. [Online]. Available: <https://github.com/garythung/trashnet>
- [29] B. Madhavi, M. Mahanty, C.-C. Lin et al., "SwinConvNeXt: A fused deep learning architecture for real-time garbage image classification," *Scientific Reports*, vol. 15, Art. no. 7995, 2025, doi: 10.1038/s41598-025-91302-7.

- [30] "Garbage detection and classification using a new deep learning-based machine vision system as a tool for sustainable waste recycling," *Waste Management*, 2023, doi: 10.1016/j.wasman.2023.04.006.
- [31] "Enhancing trash classification in smart cities using federated deep learning," *Scientific Reports*, 2024, doi: 10.1038/s41598-024-62003-4.
- [32] "A solid waste classification using reinforcement learning boosted by data augmentation and hyperparameter optimization," *Scientific Reports*, vol. 16, Art. no. 18132, 2026, doi: 10.1038/s41598-026-46347-7.
- [33] "An automated waste classification system using deep learning techniques: Toward efficient waste recycling and environmental sustainability," *Knowledge-Based Systems*, 2025, doi: 10.1016/j.knosys.2025.113063.
- [34] B. Bamne, N. Shrivastava, L. Parashar, and U. Singh, "Transfer learning-based object detection by using convolutional neural networks," in *Proc. Int. Conf. Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, 2020, pp. 328–332, doi: 10.1109/ICESC48915.2020.9156060.
- [35] A. Saxena, A. Vyas, L. Parashar, and U. Singh, "A glaucoma detection using convolutional neural network," in *Proc. Int. Conf. Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, 2020, pp. 815–820, doi: 10.1109/ICESC48915.2020.9155930.
- [36] M. Ranjan, A. Shukla, K. Soni, S. Varma, M. Kuliha, and U. Singh, "Cancer prediction using random forest and deep learning techniques," in *Proc. IEEE 11th Int. Conf. Communication Systems and Network Technologies (CSNT)*, Indore, India, 2022, pp. 227–231, doi: 10.1109/CSNT54456.2022.9787608.
- [37] R. Baghel, P. Pahadiya, and U. Singh, "Human face mask identification using deep learning with OpenCV techniques," in *Proc. 7th Int. Conf. Communication and Electronics Systems (ICES)*, Coimbatore, India, 2022, pp. 1051–1057, doi: 10.1109/ICES54183.2022.9835884.
- [38] U. Singh, P. Gupta, and M. Shukla, "Activity detection and counting people using Mask-RCNN with bidirectional ConvLSTM," *Journal of Intelligent and Fuzzy Systems*, 2022, pp. 6505–6520.
- [39] P. F. Proença and P. Simões, "TACO: Trash Annotations in Context for litter detection," *arXiv:2003.06975*, 2020.