sustainability

MDPI

*Article*

# Sustainable Smart Waste Sorting Through Reject-Aware IoT and Calibrated Vision: A Measurement-Calibrated Selective Routing Prototype

**Abdalilah Alhalangy**

Department of Computer Engineering, College of Computer, Qassim University, Buraydah 52361, Saudi Arabia; a.alhalangy@qu.edu.sa

**Abstract**

Smart waste systems can support sustainable recycling only when sensing, classification, and routing decisions are linked to transparent uncertainty management. This study aims to develop and evaluate a measurement-calibrated, risk-aware selective routing prototype for sustainable smart waste sorting under bench-scale conditions. The proposed prototype integrates an HC-SR04 ultrasonic sensing layer, Arduino Nano control, ESP8266 wireless synchronization, calibration-based fill level conversion, median filtering, exponential smoothing, hysteresis, and persistence-driven alert logic with a conveyor-based single-item vision-routing layer. The sorting layer routes items into glass, metal, plastic, trash, or a reject stream using a reject-aware VGG19 classifier with true logit temperature scaling and calibrated confidence gating. A quantitative prototype-level experimental evaluation was conducted using a five class benchmark of 2527 TrashNet images under a stratified 75/10/15 hold-out design, with cardboard and paper serving as outlier-exposure samples for the reject class. The final model achieved 88.13% test accuracy, 88.15% weighted F1, and 96.60% reject class F1. With a final policy tau = 0.70, the system routed 76.09% of routing class test items with 92.66% routed precision, while reject recall reached 98.66% and reject false acceptance was limited to 1.34%. Structured perturbation testing showed that severe visual degradation reduced automatic routing coverage but mostly shifted uncertain cases toward rejection. The contribution of this work is a transparent prototype-level mechanism that links calibrated sensing, calibrated visual confidence, selective rejection, and actuation-aware routing to support recycling stream purity under bounded uncertainty. Future work should validate the approach under multi-bin, multi-object, and field deployment conditions.

**Keywords:** sustainable waste management; smart waste sorting; recycling stream purity; Internet of Things; calibrated ultrasonic sensing; reject-aware routing; outlier exposure; temperature scaling; selective classification; risk-aware actuation



Academic Editor: Abdelhadi Makan

Received: 16 June 2026
Revised: 15 July 2026
Accepted: 20 July 2026
Published: 22 July 2026



## 1. Introduction

Municipal solid waste management (MSWM) remains a persistent sustainability challenge because delayed collection, weak source separation, and inefficient material recovery affect environmental quality, operational efficiency, and public health at the same time [1–4]. Recent reviews further show that AI- and IoT-enabled waste systems are increasingly used to support collection visibility, sorting automation, route planning, and circular economy-oriented resource recovery [5,6]. The engineering concern for smart city trash systems is not whether a bin is able to report its state or an image model is able to classify an item but whether sensing and recognition can support defensible routing decisions that safeguard recyclable streams from contamination. Within this broader framework of sustainability, the present study addresses a narrower but essential bottleneck: how a low-cost cyber-physical prototype may proceed from sensing and recognition to selective physical routing while clearly disclosing measurement error and predictive uncertainty [5–11].

In sustainability-oriented waste management, the performance of automated sorting is not only evaluated through the number of items sorted but also by the degree to which ambiguous routing decisions are minimized to avoid contamination of recyclable materials. A low confidence routing error may reduce stream purity, increase downstream handling requirements, and lessen the environmental benefits of automated recycling. This concern is consistent with sensor-based and automated sorting literature, which emphasizes that material heterogeneity, surface contamination, and stream impurity remain major barriers to high-quality recovery [7,8]. We view uncertainty-aware selective routing as a sustainability-relevant decision mechanism, not only as a post-processing stage in computer vision. Manual waste sorting remains labor-intensive, difficult to scale, and vulnerable to inconsistent material identification, while purely cloud-centered vision workflows may introduce latency, connectivity, and actuation-timing constraints in conveyor-based sorting settings. These limitations motivate edge-supported smart waste architectures in which local sensing, calibrated measurement, and confidence-aware visual recognition are linked before a physical routing action is executed [5–9].

Recent studies have placed the emergence of the smart trash paradigm in the context of Industry 4.0, the circular economy, and data-driven urban infrastructure. Structured assessments show that waste management research has moved from traditional collection and disposal to digital integrated workflows, which include IoT monitoring, route optimization, predictive analytics, automated sorting, and resource recovery aid [12–14]. These advances are relevant to sustainability, as poor measurement, lack of separation, and forced misclassification can reduce material recovery and increase stream contamination. The research repeatedly points to problems of subsystem fragmentation, limited interoperability, insufficient diversity in real world data, and the persistent gap between laboratory-level visual performance and real world sensing to action dependability [12–18].

A key research area concerns bin-level monitoring and collection coordination. IoT-enabled bins, cloud dashboards, RFID-assisted sensors, wireless communication modules, and route-aware service logic [15–17,19,20] have greatly improved the visibility of the service and the efficiency of collection. However, many systems tend to treat bins as alerting devices rather than calibrated measurement nodes. Although these systems can provide information regarding when a bin is “full”, they often do not provide enough quantitative information on measurement error, repeatability, persistence of alerts under changing readings, robustness to temporary communication failures, and the relationship between the collection state estimate and the subsequent sorting decision. Thus, the sensing layer may affect downstream operations without clearly defined uncertainty boundaries [15–17,19,20].

A second major research stream concerns waste identification using computer vision and deep learning algorithms. Transfer learning is widely used because it provides strong classification performance for data-scarce scenarios and reduces training costs [21–26]. Recent efforts have extended this topic to include resilience under poor illumination, prototype-based sorting procedures, item identification frameworks, robotic manipulation, and realistic garbage category cases [18,25–32]. Collectively, this body of work indicates that deep models can distinguish recyclable materials under controlled conditions. However, these advances also reveal a fundamental limitation: classification accuracy does not necessarily ensure reliable routing. In practical sorting systems, uncertain predictions should often be rejected or redirected rather than automatically assigned to a recyclable output stream [7–11].

This research is situated at the convergence of sustainable smart waste management, measurement modeling, predictive uncertainty, and selective actuation. It does not claim to introduce a novel neural network architecture, outperform industrial multi-object detection systems, or demonstrate city-scale deployment. It recontextualizes intelligent waste sorting as a sustainability-focused selective decision problem characterized by dual uncertainty sources: measurement uncertainty at the sensing level and predictive uncertainty at the recognition level of the model. This study connects AI-driven waste recognition to the operational goal of maintaining recyclable stream purity by considering both as explicit inputs to the routing logic under defined prototype conditions [7,10,11].

In this regard, this study develops a two-stage selective routing prototype. The monitoring stage implements calibrated ultrasonic fill level estimation with median filtering, exponential smoothing, hysteresis-based alerting, buffered communication, and persistence-driven state transitions. The sorting stage assumes single-item inspection and combines transfer learning classifiers with validation set threshold selection, mixed/reject diversion, actuator mapping, and event-level recording. The essential design concern is therefore not only whether an image can be categorized correctly but whether the system can justify autonomous routing given measured uncertainty and restricted operational risk.

Unlike performance-centric smart waste studies, this study evaluates the decision boundary between sensing, prediction, and actuation as a controllable engineering interface under uncertainty.

The main objective of this study is to develop and evaluate a measurement-calibrated, reject-aware smart waste sorting prototype that links stabilized fill level sensing, calibrated visual confidence, selective rejection, and actuation verification within a bounded bench-scale workflow. The guiding research question is: How can calibrated sensing and confidence-gated visual recognition be integrated to support selective routing decisions that preserve accepted-stream purity under prototype-level uncertainty?

This work makes three main contributions. First, it formalizes sustainable smart waste sorting as a sensing-to-actuation decision interface under bounded uncertainty, rather than as a simple combination of IoT monitoring and image classification. This contribution links calibrated fill level sensing, persistence-based alerting, calibrated visual confidence, selective rejection, and actuation verification within one auditable prototype workflow.

Second, it empirically characterizes this decision interface using a transparent prototype-level evaluation protocol. The study reports ultrasonic calibration indicators, repeatability, hysteresis-based alert stability, reject-aware transfer learning, true logit temperature scaling, validation-selected confidence gating, threshold-policy sensitivity, perturbation response, and bootstrap-supported reporting under a unified stratified hold-out design.

Third, it demonstrates a low-cost bench-scale realization that connects smart bin monitoring, vision-guided selective sorting, and routing verification while explicitly limiting the interpretation to controlled prototype conditions. The contribution is therefore not a claim of municipal-scale deployment or complete open-world waste recognition but a reproducible mechanism for converting sensing and classification uncertainty into cautious recycling-oriented routing actions.

This framework defines the contribution not just in terms of algorithmic novelty but rather in terms of a disciplined conversion of sensing and classification uncertainty to auditable and selective routing decisions that can support safer recycling-oriented smart waste prototypes.

This study clarifies the design logic through four prototype-level objectives to avoid imprecise positioning, empirical scope, or overextended sustainability claims:

O1: Quantify the stability of the calibrated fill level sensing before any service state alert is issued.

O2: Evaluate whether confidence-gated vision decisions preserve accepted-stream precision while diverting uncertain cases to a mixed/reject pathway.

O3: Test whether perturbation and reject class inputs increase conservative rejection rather than silent misrouting.

O4: Interpret inference latency, actuator response, and cycle time as an end-to-end operational boundary rather than as an isolated model speed.

We evaluated these objectives with a direct chain of evidence: O1 was supported by fill level calibration and alert-stability results; O2 was supported by validation-threshold sensitivity and reject-aware routing analysis; O3 was supported by calibration, structured perturbation, and reject class diagnostics; and O4 was supported by latency, actuation, and analytical contamination-exposure modeling. Such mapping makes it clear that the manuscript is built on engineering aims and evaluation criteria, not statistical hypotheses, which is acceptable for a study of a prototype-level cyber-physical system.

These objectives distinguish what is measured in the present study from what continues to be a future field impact claim. This study therefore reports prototype-level feasibility, selective routing dependability, and stream purity controls but does not claim municipal-scale contamination reduction, economic savings, or lifecycle impact without pilot data.

## 2. Related Work, Sustainability Context, and Research Gap

Recent research on AI-augmented municipal waste management can be classified into four interconnected domains: (1) intelligent bin monitoring and collection coordination, (2) vision-based material identification, (3) object recognition and robotic sorting, and (4) integrated IoT-AI operational frameworks for sustainable urban infrastructure. Instead of just stating the lack of a particular technology, this analysis explores a subtle restriction, the absence of a clearly defined evaluation boundary between sensing, prediction, and physical action for recycling-oriented, contamination-aware garbage sorting. It questions whether present methods provide enough clarity on measurement uncertainty, rejection logic and route verification to enable a reasonable sustainability interpretation [5–14,18].

### *2.1. Smart Bin Monitoring and Collection Orchestration*

Recent smart bin studies have improved collection-side visibility by integrating Internet of Things (IoT) devices, low-power communication protocols, route-planning algorithms, cloud-based dashboards, and real-time alerting services. Recent AI-IoT reviews similarly emphasize that smart waste systems increasingly combine sensing, connectivity, prediction, and service planning but often remain more focused on operational visibility than on calibrated measurement uncertainty [5,6]. From an instrumentation perspective, a smart bin node normally consists of a physical sensing unit, an embedded control unit, a communication module, and a service-facing interaction layer. The sensing unit may include fill level, weight, proximity, inductive, or optical sensors, depending on the target waste stream. The embedded controller stabilizes readings, applies local decision rules, and prepares status records, while wireless modules transmit alerts or buffered records to cloud dashboards. In the present prototype, this instrumentation logic is implemented using an HC-SR04 ultrasonic sensor (ElecFreaks Co., Shenzhen, China) for fill level estimation, an Arduino Nano (Arduino S.r.l., Ivrea, Italy) for local control and filtering, and an ESP8266 module (Espressif Systems, Shanghai, China) for wireless synchronization. A QR code interface can also be used as a lightweight interaction mechanism between users, collection workers, and the smart bin service layer. For example, a QR code attached to the bin can link to a web or mobile page showing the bin status, sorting instructions, service history, fault reports, or feedback forms. This type of interface is useful for public-facing smart waste systems because it does not require a dedicated display on the bin. The current prototype does not implement QR-code interaction as a functional module; however, it is compatible with such an interface because the cloud layer already stores status records and alerts that could be exposed through a QR-linked service page in future deployments. Henaien et al. proposed a sustainable IoT-based solid waste management system using LPWAN connection and intelligent traffic principles [33]. Ishaq et al. [34] studied smart bin monitoring in biological waste environments. Abo-Zahhad and Abo-Zahhad proposed YOLO-based surveillance with intelligent support for waste collection and environmental sustainability [35]. These studies collectively enhance logistical efficiency and monitoring transparency.

In most implementations, however, the bin node is implicitly seen as a notification end-point rather than a calibrated measuring device. While the fill level status is commonly supplied, exact quantification of sensing error, repeatability, hysteresis stability, alert persistence under changing readings, and recovery behavior of communication are generally limited. Furthermore, the relationship between bin state estimates and downstream sorting decisions is rarely described in sufficient detail. As a result, collection triggers may depend on sensor outputs whose uncertainty is not operationally transparent.

### 2.2. Vision-Based Recognition and Transfer Learning

Another important subject is research on computer vision and deep learning waste recognition. Benchmark-oriented studies have demonstrated that convolutional neural networks can classify recyclable materials under controlled imaging settings [21–24]. Following these, low-illumination recognition, prototype-integrated sorting mechanisms, AI-IoT integration, robotic manipulation systems, and systematic waste classification surveys were introduced [18,25–32]. This body of work demonstrates that transfer learning and vision-based recognition can enhance waste classification performance for moderate-sized image datasets, particularly under controlled imaging or constrained sorting conditions [8,9].

Evaluation is often based on accuracy, precision, recall, or mAP under closed-set assumptions. Such indicators are important but do not capture the operational risk of uncertain projections. Forcing recycling of ambiguous items may taint the purity of the output stream and violate sustainability goals in real sorting systems. Rejection is therefore a design choice rather than a post hoc adjustment, and classification performance should be interpreted within a selective decision framework [10,11].

### 2.3. Object Detection and Robotic Sorting

A third line of work moves from image-level classification to object detection and segmentation, mainly using YOLO-based and Mask R-CNN-based architectures [36–40]. These solutions are well suited to multi-object cases, busy conveyors, and robotic picking scenarios that require accurate localization. Detection-based systems are more suitable for industrial contexts in which multiple waste items may appear within a single frame.

The current prototype is not designed to replace detection-oriented systems. Instead, it assumes single-item inspection, where items are physically separated before image capture. This constraint is practical for low-cost bench-scale prototypes and controlled experimental environments but is narrower than the operational needs of high-speed industrial conveyors. This study explains the scope of this assumption and avoids overgeneralization by making it explicit.

### 2.4. Uncertainty Handling and Operational Realism

However, explicit treatment of uncertainty and practical deployment constraints remains limited across these research directions. Confidence scores are sometimes treated as if they are a direct measure of reliability, despite evidence that softmax probabilities can be miscalibrated. This motivates explicit calibration diagnostics and post-hoc calibration methods such as temperature scaling when model confidence is used as an operational decision signal [10]. System-level evaluation does not regularly involve threshold sensitivity analysis, reporting of rejection rates, calibration diagnostics, and open-set handling. Selective prediction and reject-option learning provide a methodological basis for trading coverage against risk when uncertain predictions should be withheld from automatic action [11]. Likewise, latency, actuator delay, recovery from communication, maintenance load, and risk of contamination are mostly peripheral considerations and not key parts of deployment preparedness [12–18,31,32].

The significance of this omission is that a classifier could have excellent numerical accuracy but be operationally harmful, for example, if it misroutes uncertain items or its latency exceeds actuation tolerances. A sensing module may report plausible fill percentages yet emit unstable alerts near threshold boundaries. This transition from sensing and prediction to action is not well defined without a thorough understanding of these interfaces.

### 2.5. Research Gap and Positioning

Accordingly, the research gap addressed in this study is not the absence of smart bins, convolutional classifiers, or YOLO detectors. Instead, it lies in the limited methodological treatment of the decision boundary between sensing, recognition, and physical actuation. Specifically:

- How are fill level measurements stabilized, calibrated, and bounded before service state decisions are issued?
- How are uncertain visual predictions converted into accept/reject routing actions from a cost-sensitive perspective?
- How are accepted predictions verified as physical routing events within the latency and throughput constraints?

The present work addresses this gap by contributing a compact measurement-to-routing workflow that integrates explicit sensing calibration, confidence-gated selective routing, threshold sensitivity analysis, calibration diagnostics, perturbation-aware robustness testing, and bench-level functional verification. Table 1 summarizes representative studies and positions the proposed prototype relative to previous integrated IoT-AI waste-management systems.

**Table 1.** Methodological distinction from the closest integrated IoT-AI waste system study [18].

| Evaluation Dimension | Closest Integrated IoT-AI Study [18] | Present Manuscript | Substantive Distinction |
|---|---|---|---|
| System framing | Integrated smart waste management combining IoT bins and AI segregation | Measurement-calibrated and reject-aware selective routing prototype | Focus on decision interface between sensing, prediction, rejection, and actuation |
| Bin-side evidence | Collection notification and workflow logic emphasized | Calibration metrics (MAE/RMSE), repeatability, hysteresis, persistence, buffered recovery | Bin treated as calibrated measurement instrument |

**Table 1.** *Cont.*

| Evaluation Dimension | Closest Integrated IoT-AI Study [18] | Present Manuscript | Substantive Distinction |
|---|---|---|---|
| Vision model scope | VGG19-based segregation as central model | Reject-aware VGG19 with validation-fitted calibration and threshold policy | Evaluation aligned with routing and conservative rejection |
| Routing classes | Plastic, glass, metal | Glass, metal, plastic, trash, reject | Explicit reject pathway for unsupported or uncertain inputs |
| Uncertainty handling | Confidence not central methodological focus | Temperature scaling, threshold-policy sensitivity, reject recall, false acceptance, perturbation response | Sorting framed as selective decision under uncertainty |
| Operational metrics | System functionality and model accuracy | Latency, actuator timing, cloud delay, cycle time, verification logging | End-to-end operational boundary quantified |

### *2.6. Distinction from the Closest Integrated IoT-AI Waste System Study*

The closest prior study in terms of architectural integration was that of Alourani et al. [18], who introduced an intelligent solid waste management system combining smart bin collection with VGG-19-based segregation of plastic, glass, and metal. Because that work is conceptually close to combining IoT-based monitoring and deep learning-based classification, the present manuscript explicitly delineates its narrower sustainability-oriented contribution: the conversion of measurement and predictive uncertainty into auditable selective routing actions intended to protect stream purity under prototype-scale conditions.

The present study does not position itself as an alternative smart waste management platform or as a broader integration of IoT and AI components. Instead, it isolates and formalizes a narrower but underspecified problem: how measurement and predictive uncertainties are translated into physical routing actions within a prototype-scale cyber-physical system. Therefore, the contribution lies not in architectural novelty but in the explicit modeling of the decision interface between sensing, recognition, and actuation. The distinction from [18] can be summarized along the methodological and operational axes.

- Collection-side modeling:

While previous integrated systems are designed based on notification logic and connectivity, the proposed prototype treats the bin as a calibrated measurement node. The ultrasonic sensing is described with calibration points, MAE/RMSE reporting, repeatability assessment, thresholding with hysteresis, persistence rules, and buffered communication recovery. This exposes sensing uncertainty before service state decisions are issued, rather than presuming measurement reliability.

- Model evaluation scope:

The present study uses a VGG19 reject-aware classifier trained under a unified protocol aligned with routing categories and a conservative reject class. The evaluation is embedded within a measurement-to-actuation workflow rather than being treated as a stand-alone image benchmark.

- Routing semantics:

Apart from the recyclable classes (plastic, glass, metal), the current workflow clearly retains a trash class and a reject path. Cardboard and paper photos were used as outlier-exposure samples for the reject class to teach the algorithm when to stop autonomous material routing.

- Uncertainty handling:

Confidence is not a cosmetic score. Threshold sensitivity analysis, cost-sensitive accept/reject risk modeling, calibration diagnostics, and rejection coverage were precisely quantified. Thus, the system accounts for when not to sort automatically, which is important for contamination management.

- Operational coupling:

The evaluation boundary comprises latency, actuator response time, cloud update delay, cycle duration, and verification logging. This relates the recognition performance to the physical throughput and routing practicality, rather than evaluating image classification in isolation.

Table 1 shows a direct methodological and quantitative comparison between the closest combined IoT-AI work [18] and the present study. This comparison is not meant to detract from previous efforts but to define the contribution boundary of the present study more clearly. Rather than competing on the breadth of the system, the paper differentiates itself by making measurement assumptions, confidence thresholds, rejection behavior, and actuation verification transparent, auditable, and reproducible.

## 3. Sustainability-Oriented Measurement-Calibrated Selective Routing Architecture

The proposed architecture is a two-stage selective routing cyber-physical system that directly connects calibrated bin monitoring with item-level vision-based routing. In this architecture, sensing and classification are not treated as separate modules; every stage is seen as a decision module under bounded uncertainty. The collection stage should not generate unstable service alerts due to measurement fluctuation and the sorting stage should not propagate low-confidence predictions into recyclable output streams.

Conceptually, the system can be interpreted as a sequential decision pipeline.

$$\text{Measurement} \rightarrow \text{State Estimation} \rightarrow \text{Alert Decision} \rightarrow \text{Item Recognition} \rightarrow \text{Selective Routing} \rightarrow \text{Actuation Verification.}$$

Each transition in this pipeline represents a decision interface in which the uncertainty is either stabilized, bounded, or selectively absorbed through rejection logic.

### *3.1. Collection Side: Measurement-Aware Monitoring Layer*

Figure 1 demonstrates the collection side workflow. The smart bin node includes an ultrasonic sensor, embedded control, filtering, calibration-based conversion, and wireless communication modules. HC-SR04 distance readings are filtered for noise and calibrated for the bin geometry before being converted into estimated fill levels. These estimations are not immediately acted upon but submitted to persistence and hysteresis logic before initiating a transition in the service state.

Formally, the monitoring stage performs the following:

1. Signal acquisition (raw ultrasonic measurements).
2. Stabilization (median filtering and exponential smoothing).
3. Calibration-based state estimation (distance-to-fill mapping).
4. Decision gating (threshold + persistence rules).
5. State publication (cloud synchronization with buffering).

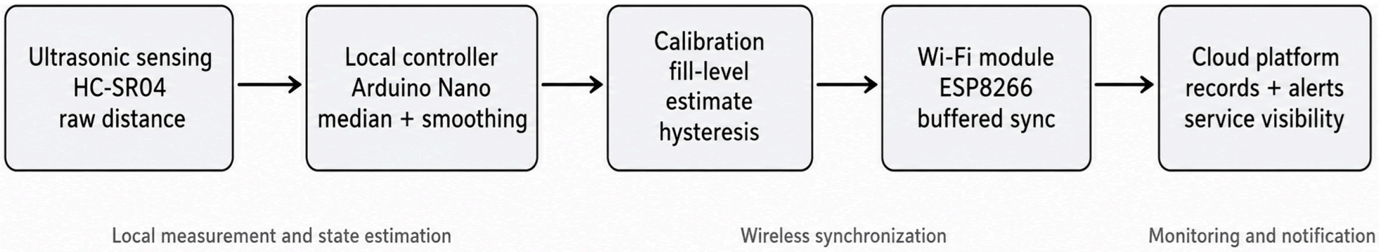


**Figure 1.** Smart bin monitoring workflow for fill level estimation, cloud synchronization, and alert generation.

The cloud layer records time-stamped data and publishes alerts only when the filtered estimate satisfies a persistence condition. The architecture introduces hysteresis and threshold persistence that decreases oscillatory state transitions around the decision boundaries. Thus, the bin is modeled as a calibrated measurement node rather than a binary notification device.

Uncertainty at this stage arises mainly from sensor noise, irregular waste surfaces, and environmental fluctuations. The architectural response is stabilization and bounded state-transition logic rather than immediate alert transmission.

### *3.2. Sorting Side: Confidence-Gated Selective Routing Layer*

Figure 2 presents the overall selective routing architecture. Waste items are sequentially introduced into a restricted inspection zone. The prototype assumes single-item presentation at the capture location to separate classification uncertainty from scene clutter effects.

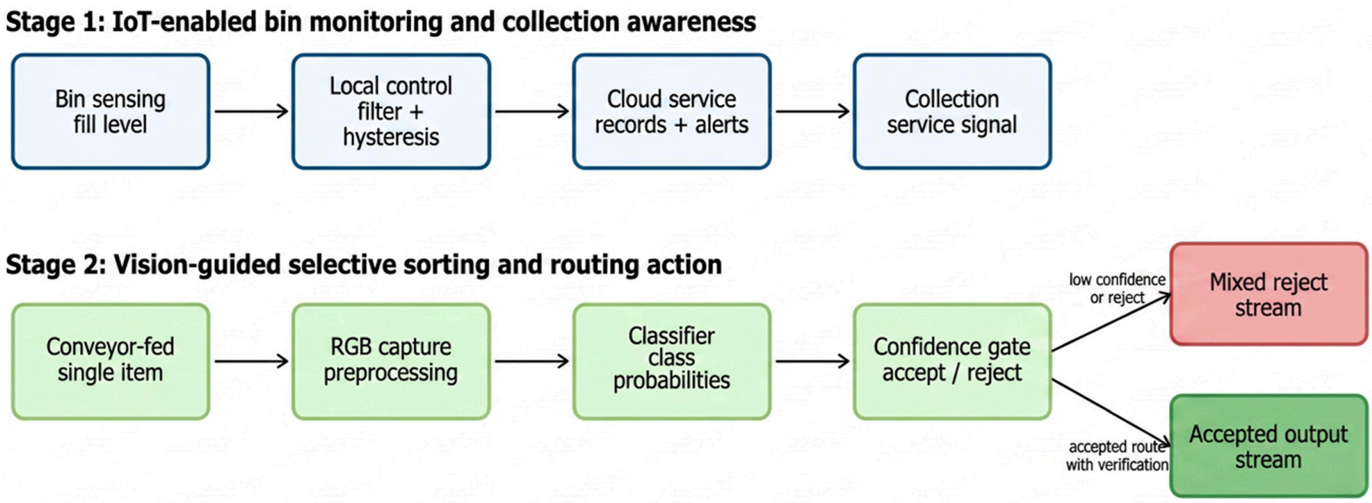


**Figure 2.** Two-stage architecture for integrated calibrated smart bin monitoring and confidence-gated vision-based routing.

The sorting stage transforms an image x into a probability vector $p = M(x)$, where M denotes the trained classifier and $p_c$ denotes the estimated probability for class c. Rather than enforcing the forced assignment $c^* = argmax_c\ p_c$, the system applies a selective acceptance policy.

If the predicted class is not rejected and $max_c\ p_c \geq \tau$, the item is routed to the corresponding output stream.

If the predicted class is rejected or if $max_c\ p_c < \tau$, after calibration, the item is diverted to the reject stream.

Here, $\tau$ denotes the calibrated confidence threshold selected on the validation split. This turns categorization into a selective decision problem under prediction uncertainty. The reject path acts as a protective buffer that absorbs unclear predictions rather than corrupting recyclable streams. Functionally, the sorting stage does:

1. Image capture under controlled geometry.
2. Preprocessing aligned with training assumptions.
3. Forward inference producing class probabilities.
4. Confidence evaluation against threshold $\tau$.
5. Actuator activation for accepted predictions.
6. Routing verification and event logging.
7. Line interruption if actuator or jam faults are detected.

Here, the uncertainty comes from the visual ambiguity, illumination variation, and model calibration error. The architectural response is selective gating and explicit rejection rather than blind actuation.

### *3.3. Uncertainty Propagation and Decision Boundary*

A significant conceptual characteristic of the design is that uncertainty is not eliminated but managed at two levels:

Filtering, calibration, and hysteresis are used to reduce measurement uncertainty prior to state publication.

Predictive uncertainty is constrained via confidence gating, calibration diagnostics and selective rejection before actuation.

The two-layer handling avoids early physical judgments based on unsteady sensor readings or overconfident misclassifications. Hence the design specifies a clear operational decision boundary: only sensing states and predictions that satisfy the stability and confidence constraints are allowed to trigger physical actions.

By formalizing these constraints, the prototype moves beyond simple IoT-AI integration and towards a risk-aware measurement-to-routing control workflow. The subsequent sections quantify how these architectural principles translate into calibration performance, threshold sensitivity, perturbation response, and throughput constraints.

### *3.4. End-to-End Operational Workflow and Decision Flow*

The operating workflow links the monitoring and sorting stages within a sequential measurement–actuation chain. The process begins with periodic sensing of the fill level of the bin being monitored. The cloud-linked layer updates the bin status and triggers a service alert when the smoothed and calibrated state estimation meets the persistence rule. The sorting station follows a confidence-gated routing approach and items are processed sequentially.

Therefore, each item passes through a structured selective decision sequence:

1. Capture,
2. Quality screening,
3. Preprocessing,
4. Probabilistic classification,
5. Confidence evaluation,
6. Routing actuation,
7. Physical verification and logging.

This layered decision flow prevents the direct influence of unstable sensor readings or poor confidence visual predictions on irreversible actions. Instead, the uncertainty is successively constrained before actuation.

The decision flow abstraction for stabilizing the measurement uncertainty before alert generation and filtering the predicted uncertainty by confidence gating before routing is shown in Figure 3. The figure highlights the architecture's dual uncertainty management principle: stabilization at the sensing layer and selective rejection at the recognition layer.

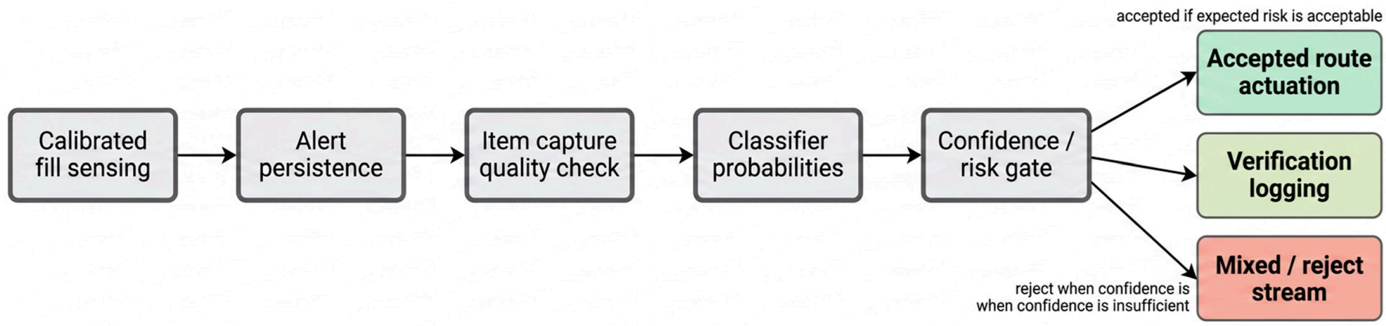


**Figure 3.** Decision flow representation showing how measurement uncertainty and predictive confidence are converted into alerting, rejection, routing, actuation, and verification actions within a two-stage architecture.

## 4. Prototype Realization, Measurement Characterization, and Selective Routing

The proposed architecture was built as a cyber-physical bench scale prototype that integrates a calibrated smart bin node with a confidence-gated selective routing station. The solution was designed to show, rather than hide, the two primary sources of uncertainty in the workflow:

1. Measurement uncertainty in ultrasonic fill level estimation.
2. Predictive uncertainty in image-based material recognition.

Thus, the collection-side subsystem was characterized in terms of calibration fidelity, repeatability, alert state stability, and communication recovery behavior. The sorting subsystem was tested for consistency in selective routing, inference latency, actuator response time and verification integrity. The dual assessment explicitly measures the sensing and actuation limits, instead of implicitly assuming them.

### *4.1. Smart Bin Node Operation, Calibration, and Uncertainty Model*

The monitoring node is realized as a small measurement-control unit with an Arduino Nano. It handles signal acquisition, filtering, state estimation from calibration, threshold evaluation and publication logic. The HC-SR04 ultrasonic sensor measures distance, while the ESP8266 module provides wireless connectivity.

Importantly, the node is not seen as a passive transmitter. It provides local stabilization and state transition control prior to surfacing a service alert. This design choice is at the heart of the contribution; measurement uncertainty is bounded at the edge, not delayed to the cloud.

Let H denote the calibrated internal measurement height of the bin, dt denote the filtered ultrasonic distance at sampling instant t, $f_t$ denote the instantaneous fill level estimate, and $s_t$ denote the smoothed fill level state. The instantaneous fill estimate is computed as:

$$f_t = clip(100 \times (H - d_t)/H, 0, 100). \tag{1}$$

In Equation (1), $clip(\cdot, 0, 100)$ bounds the estimated fill level within the operational range from 0% to 100%.

Because ultrasonic readings may fluctuate due to irregular waste geometry, reflection angle variation, and transient surface disturbance, stabilization is performed in two stages:

short-window median filtering, which is robust to local outliers, followed by exponential smoothing:

$$s_t = \alpha f_t + (1 - \alpha)\, s_{\{t-1\}}. \tag{2}$$

where $\alpha$ is the smoothing factor. This formulation transforms raw ultrasonic readings into a bounded and stabilized fill level state estimate.

A full bin alert is triggered only when $s_t \geq \theta_F$ for k consecutive sampling cycles, where $\theta_F$ is the full bin threshold and k is the persistence count. Reset occurs only when $s_t \leq \theta_R$ after servicing, where $\theta_R$ is the reset threshold. The reset threshold is lower than the full threshold ($\theta_R < \theta_F$), implementing hysteresis to prevent oscillation around the service boundary.

The thresholds $\theta_F$ and $\theta_R$ were selected as operational heuristic thresholds informed by the observed measurement-error margin and the need to avoid oscillation near the full bin boundary. They were not optimized through a deployment-scale search; rather, they were used to demonstrate stable service state publication under bounded bench-scale sensing uncertainty.

Calibration and error characterization

Calibration was conducted at five reference states: empty, 25%, 50%, 75%, and near full. For each level, repeated measurements were collected after settling intervals to reduce transient bias.

Performance was quantified using:

- Mean absolute error (MAE),
- Root mean square error (RMSE),
- Maximum absolute deviation,
- Repeatability standard deviation.

The purpose of this characterization is not metrological certification but operational transparency. By reporting bounded error indicators (Table 2), the sensing subsystem avoids black box presentation and enables inspection of decision stability margins.

**Table 2.** Bench-scale fill level measurement characterization of the smart bin node.

| Measurement Aspect | Procedure | Reported Indicator | Bench-Scale Result | Interpretation |
|---|---|---|---|---|
| Calibration points | Empty, 25%, 50%, 75%, and near full reference states | Reference-to-estimate deviation | MAE = 2.1%; RMSE = 2.8% | Suitable for service state estimation rather than precision metrology |
| Repeatability | Repeated readings at fixed reference levels after stabilization | Standard deviation of estimated fill level | ≤1.6 percentage points | Short window filtering reduced unstable readings |
| Worst-case deviation | Maximum absolute difference across reference states | Maximum absolute error | 5.4 percentage points | Supports use of hysteresis around the full threshold |
| Alert stability | Repeated threshold-crossing trials with persistence rule | False trigger tendency under transient readings | No alert before persistence condition | Reduces oscillation around service threshold |
| Communication recovery | Temporary disconnection followed by reconnection | Buffered publication behavior | Buffered records flushed after reconnection | Prevents loss of status records during short outages |

**Measurement interpretation**

The reported MAE (2.1%) and RMSE (2.8%) show that the fill level estimation is precise enough for service state transitions and not intended for high-precision volumetric metrology. That measurement error is thus clearly related to decision design, with the largest variance (5.4 percentage points) justifying a hysteresis between $\theta_F$ and $\theta_R$.

Making sensing error boundaries explicit and connecting them to alert state logic makes the monitoring layer more operationally valid.

The bin state update technique is implemented as a compact edge procedure. Ultrasonic readings are first screened and median filtered, converted to fill level, smoothed, tested against persistence and hysteresis thresholds and then published or buffered until connectivity is restored.

This buffering allows minor pauses in connectivity to not result in loss of state information that enables temporal continuity in cloud-side monitoring.

The reported calibration values are specific to the bin geometry, internal measurement height, sensor mounting position, and surface conditions used in this prototype. They should not be interpreted as directly transferable to bins with different height, shape, material, or sensor placement. For a new bin configuration, the calibration procedure should be repeated and the hysteresis margins should be rechecked before deployment.

### 4.2. Sorting Station and Routing Mechanism

The sorting station translates probabilistic predictions into physical routing actions under a selective decision policy. The station consists of:

- A fixed RGB camera positioned above a constrained inspection zone.
- A short conveyor transporting single items.
- Servo-actuated gates or arms implementing diversion.

Image preprocessing strictly follows the normalization assumptions used during training to avoid train-deployment mismatch. Given a captured image $x$, the trained classifier produces: $p = M(x)$, where $p$ is a probability vector over routing classes. The selective decision rule is defined as:

Accept and route if the calibrated prediction is not rejected and $max_c\ p_c \geq \tau$; otherwise, divert the item to the reject stream.

Thus, classification becomes a selective risk-minimization problem [11]. This routing logic differs from the classical forced classification in that it gives priority to the stream purity rather than the maximal coverage. The reject pathway acts as a contamination barrier for ambiguous or low-confidence predictions. Actuator verification and jam detection are related to the verification that logical decisions are followed by physical execution. If actuation failure is detected, the line is stopped and inspected manually to prevent silent misrouting.

The vision guided sorting routine is a one item inspection routine. The image quality is inspected, the frame is normalized, class probabilities are calculated, low-confidence predictions are routed to the mixed/reject stream, accepted predictions activate the mapped actuator, and the actuation event is logged for auditability [7,8,11].

### 4.3. Confidence-Threshold Selection and Risk-Aware Rejection Policy

The rejection threshold was selected exclusively on the validation split to prevent leakage into the final test report. This division guaranteed methodological integrity.

The adopted threshold reflects a validation-selected trade-off between routing coverage, routed precision, reject recall, and false acceptance into routing streams. To reduce redundancy, the threshold network was combined with policy interpretation in the subsequent threshold and policy analysis.

- Accepted coverage,
- Accepted item precision,
- Mixed/reject handling rates.

Decision-theoretic threshold formulation
Selective routing is framed as a cost-sensitive decision-making process [11].
For $c^* = argmax_c\ p_c$, let $p^* = p_{\{c^*\}} = max_c\ p_c$. The expected accept risk is approximated as:

$$R_{accept} = C_{mis}\ (1 - p^*) + C_{act}. \tag{3}$$

The reject risk is:

$$R_{reject} = C_{rej}. \tag{4}$$

Routing is justified when $R_{accept} < R_{reject}$; otherwise, rejection is preferred. Here, $C_{mis}$ is the contamination or misrouting cost, $C_{\mathbf{act}}$ is the actuation cost, and $C_{\mathbf{rej}}$ is the cost of diverting an item to the reject stream. Under this formulation, a higher contamination cost $C_{mis}$ shifts the operating point toward higher thresholds, lower coverage, and higher accepted stream purity.

Under stable class priors and fixed actuation costs, this rule reduces to a threshold condition on maximum calibrated probability $p^*$. Increasing the contamination-to-rejection cost ratio systematically shifts the operating point toward a lower coverage and higher purity.

This reframes the threshold selection as a policy variable rather than an arbitrary hyperparameter.

**Formal cost-sensitive policy interpretation**

The selective threshold can be interpreted as a cost-weighted probabilistic decision boundary. Using the previously defined $p^*$, automatic routing is justified when:

$$C_{mis}\ (1 - p^*) + C_{act} < C_{rej}. \tag{5}$$

Otherwise, rejection is preferred. Rearranging this inequality indicates that the acceptance threshold increases with the relative increase in the cost of contamination over the cost of rejecting an item. Thus, the strategy is not an arbitrary confidence threshold but a visible mechanism to balance coverage, accepted-stream purity, and mixed/reject handling under sustainability-oriented contamination risk [10,11].

### *4.4. Main Prototype Components*

The prototype was purposefully designed to be low cost and functionally integrated. An Arduino Nano (Arduino S.r.l., Ivrea, Italy) is responsible for local sensing and update scheduling, an HC-SR04 sensor (ElecFreaks Co., Shenzhen, China) provides a calibrated distance-to-fill estimation, an ESP8266 module (Espressif Systems, Shanghai, China) allows for online publication and buffered recovery, and the sorting unit is made of a fixed RGB camera, conveyor, and servo-driven diversion mechanism with a defined mixed/reject pathway.

### *4.5. Computing Environment and Reproducibility Context*

Model training and evaluation were conducted in a reproducible Python/TensorFlow (Python 3.10 and TensorFlow 2.13 (TensorFlow, Google LLC, Mountain View, CA, USA)) environment using Adam optimization, class-weighted sparse categorical cross-entropy, 224 × 224 RGB inputs, batch size 32, early stopping, ImageNet-initialized VGG19 transfer learning, and validation-logit-fitted temperature scaling. This description replaces a separate environment table to keep the manuscript concise while retaining the reproducibility information.

To enhance methodological transparency, the final model was trained as follows:

- A fixed five class-stratified split,
- A frozen backbone stage followed by last block fine-tuning,
- Early stopping based on validation loss,
- Class-weighted loss to mitigate imbalance, including a smaller trash class.

This ensured that the reported model reflected a controlled reject-aware training configuration rather than a post hoc threshold tuning.

## 5. Materials and Methods: Experimental Protocol and Evaluation Design

### *5.1. Benchmark Selection and Class Alignment*

The experimental approach was meant to mimic the physical logic of the selective routing station while providing a controlled and reproducible benchmark environment. It kept a five class baseline from the public TrashNet dataset [41], which includes glass, metal, plastic, trash, and reject.

The operational motivation to include a reject class is discussed. Images of cardboard and paper are used as outlier-exposure samples to educate the algorithm to not perform automatic routing when an item is outside the available material streams. This re-frames the problem as reject-aware routing instead than enforced closed-set classification.

The last benchmark consists of 2527 photos with a mild imbalance. Stratified partitioning is used to preserve class proportions in training, validation and test splits so that sample artifacts do not influence threshold selection and the final reporting. Table 3 gives an overview of the composition of the dataset.

**Table 3.** Prototype-aligned benchmark composition and stratified partitioning.

| Class | Total Images | Training Set | Validation Set | Test Set | Practical Relevance to Sorting Station |
|---|---|---|---|---|---|
| Glass | 501 | 376 | 50 | 75 | Directly mapped to the recyclable glass stream |
| Metal | 410 | 307 | 41 | 62 | Directly mapped to the recyclable metal stream |
| Plastic | 482 | 362 | 48 | 72 | Directly mapped to the recyclable plastic stream |
| Trash | 137 | 102 | 14 | 21 | Assigned to the non-recyclable routing stream |
| Reject (cardboard + paper) | 997 | 749 | 99 | 149 | Unsupported material categories used for outlier-exposure-assisted rejection |
| Total | 2527 | 1896 | 252 | 379 | Five class reject-aware benchmark aligned with prototype outputs |

Dataset origin and evaluation boundary: All images utilized for model assessment were drawn from the public TrashNet benchmark and were split before to calibration, threshold selection, and final reporting. The manuscript thus disjoins training, validation, calibration, threshold selection, and final test reporting to reduce the danger of data leakage and maintain the presented evidence in line with prototype-level evaluation. The reject class should be viewed as outlier-exposure-assisted rejection, rather than full open-world recognition [11]. Therefore, the reported reject performance should not be interpreted as zero-shot detection of arbitrary unseen waste categories such as textiles, ceramics, organic waste, or mixed overlapping items; it reflects conservative rejection behavior for the outlier-exposure categories represented by cardboard and paper under the controlled benchmark setting.

This dataset is treated as a prototype-aligned reject-aware benchmark, and its limitations are acknowledged explicitly. The goal is not to claim open-world representativeness, but to evaluate selective routing under controlled class alignment and conservative rejection behavior.

### 5.2. Image Preprocessing and Augmentation Strategy

All images were resized to $224 \times 224$ RGB format and converted to float tensors. The final model used VGG19-compatible preprocessing inside the network pipeline rather than external [0, 1] normalization, thereby preserving consistency between training, calibration, perturbation testing, and held-out inference.

Augmentation was applied exclusively to the training split to avoid information leakage. The augmentation policy included:

- ○ Small random rotation (0.05),
- ○ Horizontal flipping,
- ○ Random zoom (0.10),
- ○ Random contrast (0.10).

Augmentation intensity was intentionally bounded. The objective was to maintain controlled training behavior rather than maximize benchmark scores through aggressive synthetic expansion.

### 5.3. Stratified Hold-Out Validation Protocol

The evaluation employed a stratified 75%/10%/15% hold-out protocol for training, validation, and testing, respectively. No k-fold cross-validation was conducted.

The separation of roles was strict.

- Training set: model fitting.
- Validation set: checkpoint selection, early stopping, temperature scaling, and threshold selection.
- Test set: Final performance reporting only.

No hyperparameter tuning, threshold adjustment, or calibration fitting was performed on the test set. This separation ensures that the selective routing policies are reproducible and not tuned to the reporting data.

The validation split was used only for checkpoint selection, temperature-scaling fitting, and threshold policy selection. Although bootstrap resampling provides uncertainty bounds for the reported held-out metrics, it does not replace a full multi-seed threshold-stability study. Therefore, the selected threshold should be interpreted as validation-selected under the present stratified split, and multi-seed or repeated-split validation is identified as future work.

To reduce overinterpretation, the model performance was interpreted conservatively based on the size of the held-out test set. The main manuscript focuses on uncertainty reporting in calibration, threshold policy sensitivity, perturbation response, and reject-aware routing behavior, where the operational implications of uncertainty are most directly visible.

### 5.4. Selected Transfer-Learning Backbone

The final visual model uses VGG19 with ImageNet initialization. VGG19 was retained as a reference backbone because it aligns with the closest integrated IoT-AI waste-system study and provided stable transfer-learning behavior under the adopted reject-aware protocol. The study does not claim architectural superiority; comparative lightweight backbone benchmarking is treated as future work because the present contribution focuses on calibrated selective routing behavior rather than architecture ranking.

The reject-aware model was trained in two stages:

- Stage 1: the convolutional VGG19 base was frozen while the task-specific classification head learned the five class mapping.
- Stage 2: the final VGG19 convolutional block was fine-tuned with a reduced learning rate to improve class separation without destabilizing validation loss.
- Both stages used the same train/validation/test split.
- Checkpoint selection used validation loss only.
- Temperature scaling was fitted on true pre-softmax validation logits; threshold selection was fitted on the resulting calibrated validation outputs.
- Final performance was reported only on the held-out test split.

This protocol aligns model selection with the reject-aware selective routing task rather than with isolated benchmark accuracy. The evaluation therefore examines calibrated routing behavior within a measurement-to-actuation workflow, not only image classification scores.

### *5.5. Scope Relative to Object Detection and Segmentation*

The use of image-level classification reflects a deliberate choice of scope. The prototype assumes single-item isolation at the inspection points. Under this constraint, classification is sufficient to test confidence-gated selective routing.

For multi-object scenes, overlapping waste, or high-speed conveyors, object detection and segmentation models such as YOLO or Mask R-CNN are more appropriate [36–40]. The present evaluation does not claim to be equivalent to detection-based industrial systems. Instead, it isolates the decision-theoretic contribution in a controlled single-item routing environment.

This explicit scope prevents overgeneralization.

### *5.6. Evaluation Metrics and Statistical Scope*

The performance was evaluated using the following metrics:

- Test accuracy,
- Macro-precision, macro-recall, and macro F1,
- Weighted F1,
- Reject class precision, recall, and F1,
- Routed precision, routing coverage, reject recall, and reject false acceptance under the final policy,
- Calibration diagnostics (ECE and Brier scores).

Macro metrics were prioritized because they penalized class imbalance and reflected routing fairness across streams.

To enhance statistical caution, accuracy and macro-level metrics were interpreted as point estimates from a moderately sized held-out test set rather than as definitive population-level performance claims. This framing guards against deterministic interpretations beyond the controlled prototype-aligned benchmark.

The experimental design was a controlled prototype-aligned evaluation. It does not attempt to replicate the full heterogeneity of municipal waste streams.

### 5.7. Calibration, Perturbation, and Throughput Analysis Design

**Calibration Protocol**

Because softmax confidence does not guarantee probabilistic correctness, calibration diagnostics were applied to validation outputs. The expected calibration error (ECE) was computed using 15 equal-width confidence bins over the maximum calibrated softmax probability, alongside the Brier score, reliability diagrams, and temperature scaling.

- Expected calibration error (ECE),
- Brier score,
- Reliability diagrams,
- Temperature scaling.

Temperature scaling was fitted exclusively on the true pre-softmax logits of the final reject-aware VGG19 model and applied unchanged to test logits. To obtain logits without retraining, the weights and bias of the final softmax dense layer were reused in an equivalent dense layer with linear activation. A consistency check confirmed that softmax applied to the reconstructed logits exactly reproduced the original model probabilities, with maximum absolute difference = 0.0. No class ranking was modified by calibration.

Test set ECE and Brier score are reported to verify calibration transferability beyond validation fitting.

**Structured Domain-Shift and Perturbation Protocol**

Beyond accuracy, a structured domain-shift layer was introduced. After partitioning, held-out test images were subjected to controlled perturbations:

- Dark and bright lighting shifts,
- Average blur simulation,
- Central occlusion,
- Horizontal flip.

Perturbation severity was deterministic and fixed for reproducibility: dark lighting used a 0.65 intensity multiplier, bright lighting used a 1.35 multiplier with clipping, blur used a $5 \times 5$ average filter, center occlusion used a $64 \times 64$ pixel mask at $224 \times 224$ resolution, and horizontal flip used deterministic left-right reversal. These conditions are reported as structured stress diagnostics rather than field-level robustness claims.

For each condition, the following were reported:

- Argmax accuracy and policy accuracy,
- Routing coverage and routed precision,
- Reject recall,
- Reject false acceptance rate.

The objective is not to simulate full field variability but to test whether the reject mechanism scales conservatively under visual degradation rather than forcing unstable routing decisions.

Reject-aware outlier-exposure evaluation design

To assess rejection behavior beyond the four routing classes, the benchmark includes a reject class constructed from TrashNet cardboard and paper categories. These categories are not mapped to physical routing streams in the prototype; instead, they function as outlier-exposure samples that train the model to withhold automatic routing. This design differs from zero-shot OOD detection because the reject categories are seen during training, and the results are therefore reported as reject-aware routing performance rather than as pure open set recognition.

- Reject class precision and recall,
- False acceptance of reject items into routing streams,

- Routing false reject rate,
- Threshold-sensitive routing coverage and routed precision.

This analysis tests whether the reject pathway absorbs non-routing categories instead of misrouting them into recyclable streams. Because cardboard and paper are used as outlier-exposure categories, the results are interpreted as a conservative reject-routing diagnostic, not as proof of complete open-world reliability.

**Cost-Sensitive Sensitivity Analysis**

Selective routing is examined across a contamination-to-rejection cost-ratio grid $\left(C_{mis} : C_{rej}\right)$. Each ratio is mapped to the nearest empirical threshold obtained from the validation sensitivity grid. This approach transforms the routing policy into a transparent, parameterized decision boundary rather than relying on a fixed heuristic.

**Throughput Constraint Decomposition**

The routing cycle was decomposed into:

- Image capture and quality check,
- Inference time,
- Command transmission,
- Actuator response,
- Gate reset,
- Verification logging.

The throughput boundary and safe belt speed were interpreted from the module-level timing decomposition and actuator cycle constraints rather than from classifier speed alone.

### *5.8. Selective Risk-Coverage Analysis*

To formalize the selective routing behavior beyond a single threshold, the evaluation adopts a risk-coverage perspective from selective classification. For a threshold tau, coverage is the proportion of items automatically routed, whereas selective risk is approximated as one minus accepted item precision. This separates the classification quality from the decision selectivity and allows the routing policy to be interpreted as a controllable risk-coverage tradeoff.

Risk coverage behavior was evaluated directly from the validation-selected threshold grid. Because the final model includes both explicit reject prediction and a calibrated confidence gate, the principal reporting emphasis is on routing coverage, routed precision, reject recall, and reject false acceptance, rather than on a single aggregate AURC or pAURC value. Full AURC reporting was therefore omitted to keep the analysis aligned with policy-level routing decisions rather than benchmark-level selective classification ranking.

### *5.9. Statistical Reliability and Reporting Discipline*

To avoid overinterpreting point estimates, this study treats accuracy, macro F1, and weighted F1 as prototype-level results from a controlled held-out test set. Final model selection was based on balanced routing class behavior, strong reject class performance, calibration suitability, and routing-oriented interpretability rather than on a claim of universal architectural superiority.

For reproducibility, the supplementary workflow includes held-out prediction logs, true logit calibration outputs, bootstrap confidence interval summaries, threshold policy tables, perturbation outputs, and seed/protocol notes. The fixed split indices and bootstrap seed are provided in the Supplementary Materials to support the recomputation of the reported point estimates and uncertainty intervals.

# 6. Results and Discussion

## *6.1. Final Reject-Aware VGG19 Performance Under the Unified Benchmark Protocol*

Table 4 summarizes the final reject-aware VGG19 results under the stratified five class hold-out protocol. The test set contained 379 images, including 230 routing class items and 149 reject class items. The model achieved 88.13% test accuracy, 87.84% balanced accuracy, 85.20% macro F1, and 88.15% weighted F1.

**Table 4.** Five class reject-aware VGG19 performance on the held-out test set.

| Class/Summary | Precision (%) | Recall (%) | F1-Score (%) | Support | Result Type | Operational Interpretation |
|---|---|---|---|---|---|---|
| Glass | 86.57 | 77.33 | 81.69 | 75 | Routing class | Usable routing; residual cross-confusion |
| Metal | 80.56 | 93.55 | 86.57 | 62 | Routing class | High recall; moderate precision |
| Plastic | 84.85 | 77.78 | 81.16 | 72 | Routing class | Usable; visually confusable |
| Trash | 68.97 | 95.24 | 80.00 | 21 | Routing class | High recall; small heterogeneous class |
| Reject | 97.93 | 95.30 | 96.60 | 149 | Reject class | Strong reject absorption |
| Accuracy | 88.13 | 88.13 | 88.13 | 379 | Overall | Overall held-out accuracy |
| Macro average | 83.77 | 87.84 | 85.20 | 379 | Aggregate | Class-balanced behavior |
| Weighted average | 88.75 | 88.13 | 88.15 | 379 | Aggregate | Support-weighted behavior |
| Balanced accuracy | 87.84 | 87.84 | 87.84 | 379 | Aggregate | Class-balanced summary for the imbalanced five class split |

To prevent overinterpretation of point estimates, the final model is interpreted as a prototype-level operating point rather than as universal evidence of architectural superiority. Because the five class split is moderately imbalanced, balanced accuracy, macro F1, and reject/routing policy metrics are treated as primary decision evidence, while overall accuracy is reported as a secondary descriptive metric.

A closer inspection reveals distinct decision profiles:

- Reject behavior was the strongest component, with precision = 97.93%, recall = 95.30%, and F1 = 96.60%.
- Routing classes remained usable but heterogeneous: metal and trash showed high recall, whereas glass and plastic exhibited more cross-confusion.
- The final model should therefore be interpreted as a reject-aware routing policy rather than as a purely accuracy-optimized classifier.

Within the selective routing framework, reject recall and routed precision are operationally more informative than aggregate accuracy because they reflect stream purity protection. Accordingly, the VGG19 reject-aware model was selected for calibration and threshold-policy analysis.

However, these findings should not be generalized beyond the controlled benchmark. The five class TrashNet-derived split contains limited inter-class ambiguity and constrained visual diversity relative to real municipal waste streams. The results therefore characterize reject-aware routing behavior within a selective routing prototype, not universal model superiority [25–32,36–41].

### *6.2. Class-Wise Behavior of the Selected Backbone*

Class-wise behavior confirms that the reject class is reliably identified while routing class performance remains differentiated by material. On the held-out test set, reject F1 reached 96.60%, metal F1 reached 86.57%, glass F1 reached 81.69%, plastic F1 reached 81.16%, and trash F1 reached 80.00%. The lower trash precision reflects the semantic heterogeneity of the trash category and the small support of that class.

The confusion matrix showed that 142 of the 149 reject class items were correctly assigned to the reject class, while only 7 reject class items were falsely accepted into routing categories. Conversely, only 3 routing class items were falsely assigned to reject under direct argmax prediction. This asymmetry is desirable for a stream purity-oriented prototype because uncertain or unsupported categories should be absorbed by the reject pathway rather than silently routed.

These results should be interpreted as controlled-benchmark evidence of reject-aware separability, not as proof of field robustness or universal open-world recognition.

### *6.3. Uncertainty Calibration, Risk-Aware Thresholding, and Selective Routing Behavior*

Inference Latency and Operational Feasibility

The throughput interpretation was deliberately conservative. Because no additional standalone batch-1 latency benchmark is reported in this submission, neural inference is interpreted as part of a module-level software-and-actuation boundary rather than a field-validated throughput claim. The nominal 0.7–1.0 s routing cycle is therefore discussed primarily as an actuator-limited prototype boundary.

This observation shifts the engineering bottleneck from classification speed to actuator response and mechanical timing, reinforcing the need to evaluate routing decisions within physical constraints rather than accuracy alone.

Calibration Transferability

Table 5 reports the calibration diagnostics for the reject-aware VGG19 model. Temperature scaling was fitted to the true pre-softmax validation logits, selecting T = 1.24. After scaling, the validation set produced an ECE of 0.0486 and a Brier score of 0.1625, whereas the held-out test set produced an ECE of 0.0349 and a Brier score of 0.1697.

**Table 5.** Calibration diagnostics for the reject-aware VGG19 model after validation-fitted temperature scaling.

| Scenario | Top-1/Policy Result | ECE | Brier Score | Temperature T | Decision-Level Interpretation |
|---|---|---|---|---|---|
| Validation after temperature scaling | Accuracy = 89.68% | 0.0486 | 0.1625 | 1.24 | Validation-fitted calibration was used for threshold selection. |
| Held-out test after temperature scaling | Accuracy = 88.13% | 0.0349 | 0.1697 | 1.24 | Confidence alignment is transferred to the test without refitting. |
| Final policy at tau = 0.70 on test | Coverage = 76.09%; routed precision = 92.66% | Policy-level | Policy-level | 1.24 | Reject if the predicted rejection or calibrated confidence < 0.70. |
| Reject behavior under final policy | Reject recall = 98.66%; false acceptance = 1.34% | Policy-level | Policy-level | 1.24 | Unsupported categories are rarely routed into the material streams. |

Importantly, the calibration was fitted exclusively on validation logits and applied unchanged to the held-out test logits, preserving methodological separation. The test mean confidence (0.8790) closely matched the test accuracy (0.8813), indicating an acceptable probability alignment for selective routing decisions.

The final decision policy rejects an item if the predicted class is rejected or if the calibrated maximum confidence is below tau. The final threshold tau = 0.70 was selected on the validation set as the lowest threshold satisfying reject false acceptance ≤ 2.5% and routed precision ≥ 90%, while maximizing routing coverage among eligible thresholds. On the test set, this policy routed 76.09% of routing class items with 92.66% routed precision, while the reject recall reached 98.66% and reject false acceptance was limited to 1.34%.

**Bootstrap uncertainty analysis**

To quantify the sampling uncertainty in the held-out test split, non-parametric bootstrap resampling was applied to the test prediction log using 5000 resamples. Table 6 reports the 95% confidence intervals for the primary classification and final policy metrics. These intervals are descriptive uncertainty bounds for the controlled benchmark and should not be interpreted as field-level performance guarantees for the model.

**Table 6.** Bootstrap 95% confidence intervals for primary held-out test metrics under the final true logit-calibrated policy.

| Metric | Point Estimate (%) | 95% Bootstrap CI (%) | Operational Reading |
|---|---|---|---|
| Top-1 accuracy | 88.13 | 84.70–91.29 | Overall top-1 performance |
| Balanced accuracy | 87.84 | 84.13–91.39 | Class-balanced performance |
| Macro F1 | 85.20 | 80.63–89.23 | Class-wise balance |
| Weighted F1 | 88.15 | 84.83–91.45 | Support-weighted performance |
| Reject F1 | 96.60 | 94.33–98.50 | Reject-stream separability |
| Routing coverage at tau = 0.70 | 76.09 | 70.42–81.39 | Automatic routing retained |
| Routed precision at tau = 0.70 | 92.66 | 88.71–96.28 | Precision of routed items |
| Reject recall at tau = 0.70 | 98.66 | 96.53–100.00 | Reject items safely diverted |
| Reject false acceptance at tau = 0.70 | 1.34 | 0.00–3.47 | Reject items wrongly routed |

Cost-Sensitive Threshold Sensitivity

Table 7 reframes the threshold selection as a policy parameter rather than a static hyperparameter. As tau increases, the operating point shifts toward

**Table 7.** Reject-aware threshold policy sensitivity after temperature scaling.

| Split | Threshold Tau | Routing Coverage (%) | Routed Precision (%) | Reject Recall (%) | Reject False Acceptance (%) | Operational Reading |
|---|---|---|---|---|---|---|
| Validation | 0.70 | 76.47 | 92.44 | 97.98 | 2.02 | Validation-selected operating point |
| Test | 0.50 | 89.57 | 85.31 | 96.64 | 3.36 | Coverage-oriented setting |
| Test | 0.70 | 76.09 | 92.66 | 98.66 | 1.34 | Adopted prototype policy |

**Table 7.** *Cont.*

| Split | Threshold Tau | Routing Coverage (%) | Routed Precision (%) | Reject Recall (%) | Reject False Acceptance (%) | Operational Reading |
|---|---|---|---|---|---|---|
| Test | 0.85 | 59.57 | 95.62 | 100.00 | 0.00 | High-purity, lower-coverage setting |
| Test | 0.95 | 44.78 | 97.09 | 100.00 | 0.00 | Very conservative setting |

- Lower accepted coverage,
- Higher accepted-item precision,
- Increased mixed/rejected handling.

This monotonic pattern confirms the theoretical consistency with the risk formulation in Section 4.3. The adopted threshold tau = 0.70 provides a balanced prototype setting that prioritizes stream purity protection while avoiding excessive loss of automatic routing coverage.

This explicit tradeoff transparency strengthens reproducibility, as alternative municipalities can select different operating points by adjusting the cost ratio rather than modifying the architecture.

### *6.4. Structured Perturbation Response and Operational Boundary*

Perturbation analysis was designed to test whether visual degradation was absorbed through increased rejection rather than misclassification.

Perturbation analysis showed that the final policy conservatively responded to visual degradation. Under the clean test condition, the routing coverage was 76.09%, and the routed precision was 92.66%. Under dark lighting and blur, coverage decreased to 59.13% and 57.83%, respectively, indicating that uncertain cases were increasingly diverted to rejection. Under center occlusion, coverage fell to 30.87%, but routed precision remained 90.28%, and reject false acceptance remained below 1%. The compact robustness and operational boundary results are presented in Table 8.

**Table 8.** Compact reject-aware robustness and operational boundary summary.

| Evidence Layer | Key Measured Result | Operational/Sustainability Reading |
|---|---|---|
| Final clean policy | $C_{mis}(1 - p^*) = 0.70$: coverage 76.09%, routed precision 92.66%, reject recall 98.66%, and false acceptance 1.34%. | Adopted selective routing operating point. |
| Perturbation response | Dark lighting and blur reduced coverage to 59.13% and 57.83%, respectively; reject recall remained $\geq$98.66%. | Degraded inputs were mostly shifted to rejection. |
| Severe occlusion | Center occlusion reduced the coverage to 30.87%, routed precision remained at 90.28%, and false acceptance was 0.67%. | Coverage is sacrificed in the event of severe evidence loss. |
| Calibration transferability | T = 1.24; test ECE = 0.0349; Brier = 0.1697. | Calibrated confidence supports the thresholding. |
| Throughput boundary | No new standalone batch-1 latency benchmark is claimed; the safe cycle remains approximately 0.7–1.0 s under bench assumptions. | Throughput is treated as actuator- and spacing-limited. |
| Belt speed boundary | A 20 cm spacing and 1.0 s cycle imply v $\leq$ 0.20 m/s. | The spacing should follow the slowest safe actuation. |

Crucially, the perturbation response shows a safety-oriented behavior: degraded visual evidence reduces automatic routing coverage rather than increasing silent contamination. The reject recall remained high across perturbations, ranging from 98.66% to 99.33%.

The center occlusion condition revealed the main boundary of the prototype. The system maintains conservative rejection but cannot preserve high coverage when the visual evidence is severely degraded. This is a practical limitation rather than a failure of the rejection policy.

These results should be interpreted as structured stress diagnostics rather than as field validation substitutes.

### 6.5. Reject-Aware Outlier-Exposure Evaluation

The previous zero-shot OOD interpretation was replaced with a reject-aware outlier exposure design. Cardboard and paper images were incorporated as a reject class because they were not mapped to the implemented glass, metal, plastic, or trash routing streams. This design allows the model to learn conservative non-routing behavior for unsupported material categories while avoiding unsupported claims regarding pure open-set recognition.

The resulting reject behavior was strong: on the held-out test set, the reject precision reached 97.93%, reject recall reached 95.30%, and reject F1 reached 96.60%. Under the final calibrated policy, tau = 0.70, reject recall increased to 98.66%, and reject false acceptance into routing streams fell to 1.34%.

A source-category diagnostic under the final tau = 0.70 policy further showed that all 60 held-out cardboard items and 87 of the 89 held-out paper items were rejected. This diagnostic indicates consistent behavior across the two outlier-exposure sources, but it remains within the seen reject categories and should not be interpreted as leave-one-category-out or fully unseen open-set generalization.

These results confirm that outlier-exposure-assisted rejection is more reliable for the present prototype than fixed-threshold OOD screening based solely on the maximum softmax confidence. Therefore, the claim is bounded; the system learns a practical reject route for unsupported TrashNet categories, but it remains a bench-scale model that requires broader field validation before open-world deployment.

### 6.6. Throughput Boundary Interpretation

Throughput interpretation situates the recognition performance within mechanical constraints. Because a fresh standalone latency benchmark could not be added to the final submission package, this study does not claim a new measured GPU/edge latency value. Instead, the safe cycle and belt speed discussions are framed as bench-scale operational boundaries dominated by mechanical actuation and item spacing. The timing boundaries are summarized in Table 8.

The derived belt speed constraint ($\leq$0.20 m/s for 20 cm spacing and 1.0 s cycle) clarifies that routing safety is constrained by mechanical timing rather than by neural computation.

This integrated timing analysis supports a conservative module-level interpretation of the measurement-to-actuation chain, rather than a claim of continuous high-throughput deployment.

### 6.7. Training Dynamics and Convergence Behavior

Figures 4 and 5 confirm the stable convergence of the selected reject-aware model. The frozen-backbone stage improved the validation accuracy from 52.78% in the first epoch to the best validation loss at epoch 26, and fine-tuning further reduced the validation loss to 0.339 at the selected checkpoint. However, smooth convergence on a controlled dataset does not imply robustness to open-world or field-level variations. The convergence curves validate internal consistency but not external generalization.

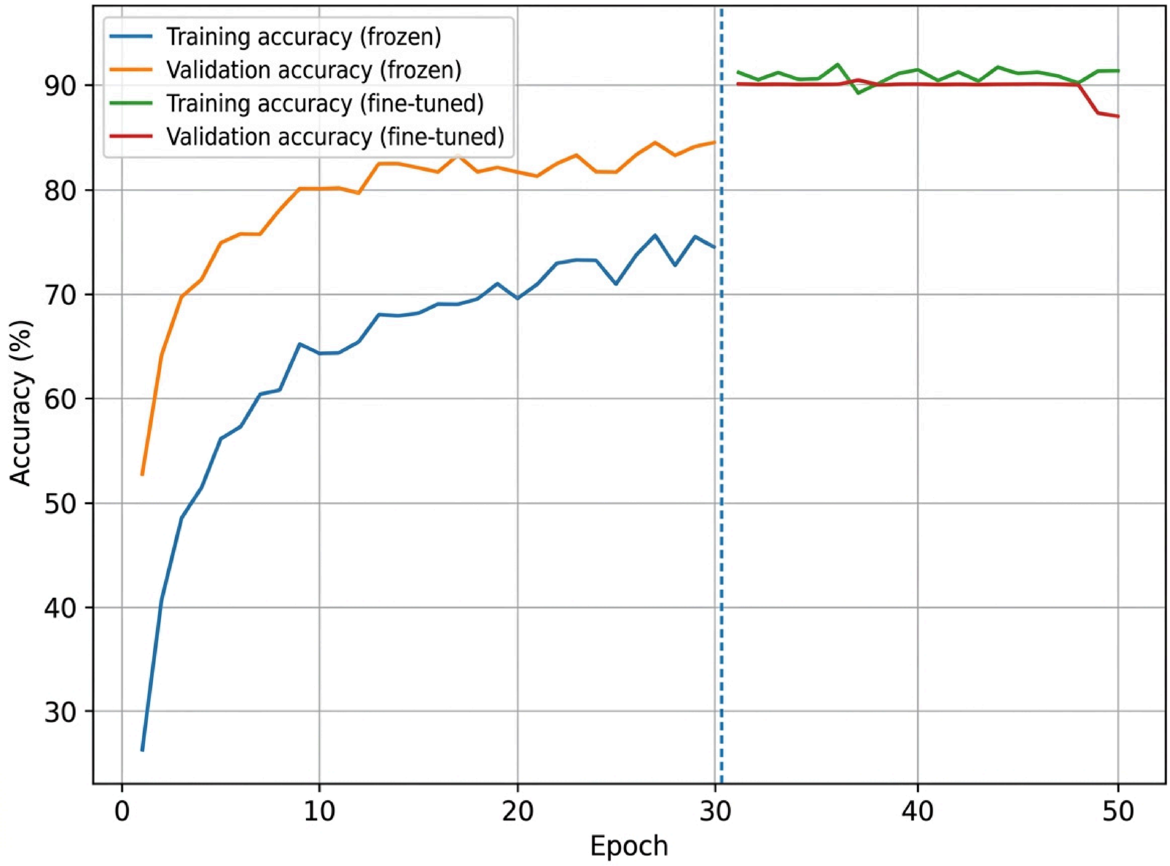


**Figure 4.** Training and validation accuracy trajectories for the selected best-performing models. The dotted vertical line indicates the transition point from frozen-backbone training to fine-tuning stage.

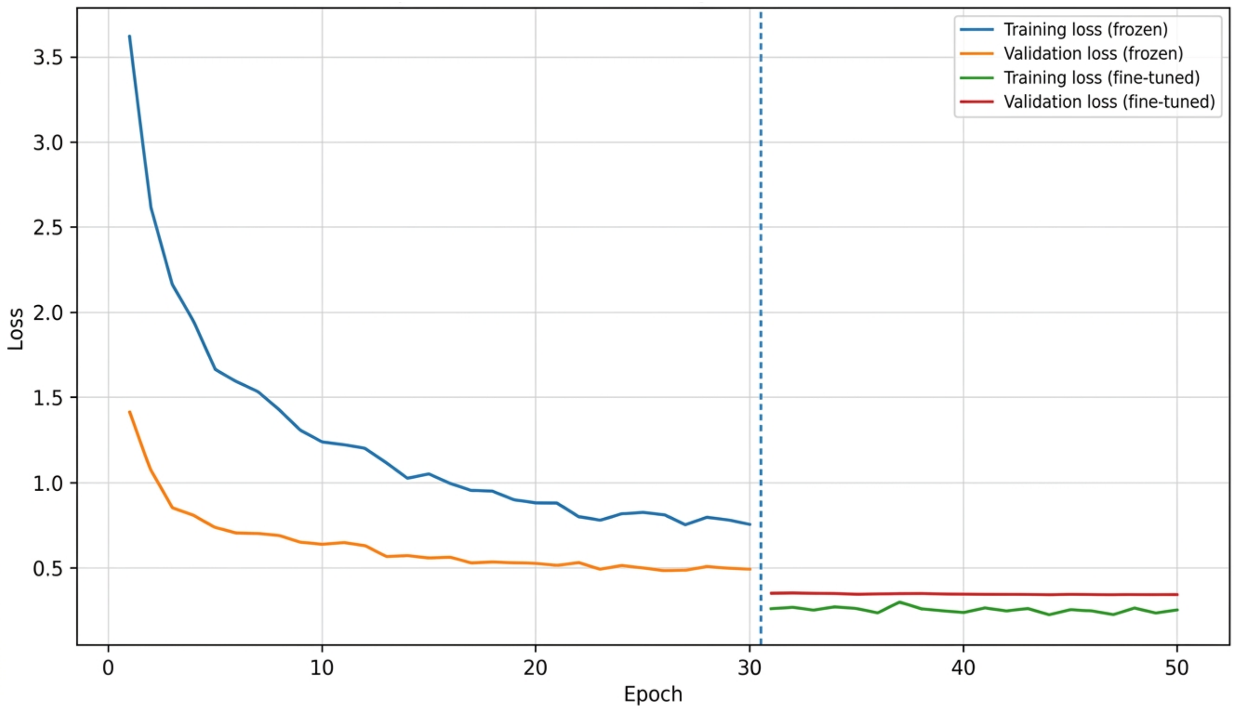


**Figure 5.** Training and validation loss trajectories of the selected best-performing model. The dotted vertical line indicates the transition point from frozen-backbone training to fine-tuning stage.

### 6.8. Quantitative Prototype-Level Functional Verification

The prototype-level functional verification is summarized in a narrative form to reduce table density. Hysteresis is justified by the measurement error limitations, persistence logic avoids false warnings, buffering ensures state continuity, confidence gating prevents the forced routing of ambiguous items, and actuator verification connects the recognition output to the physical action. Taken together, the evidence demonstrates technological coherence at prototype level, but does not promise industrial throughput or field-scale sustainability implications.

The benchmark evaluation and actuator timing verification were tested as connected prototype modules, rather than as a continuous long-duration sorting deployment. Thus, the evidence supports module-level coherence of the measurement-to-actuation workflow, but it does not ensure continuous field operation or industrial throughput.

The key outcome is coherence.

- The measurement error bounds justify the hysteresis.
- The persistence logic prevents false alerts.
- Cloud buffering preserves the state continuity.
- Confidence gating prevents the forced misrouting.
- Actuator verification ensures the consistency of physical execution.

This narrative synthesis transforms the component-level evaluation into system-level validation, reinforcing the measurement-to-actuation integrity claim.

### 6.9. Integrated Discussion and Sustainability Implications

Overall, the results show that the proposed measurement-calibrated and reject-aware workflow is technically coherent at the bench size and applicable for sustainability-oriented smart waste systems. The selective routing approach not only improves classification metrics but also transforms decision behavior under uncertainty by defining when automated routing is warranted and when diversion to a reject stream is safer for recyclable stream purity.

Three key insights emerge from this study.

1. Uncertainty stabilization at the sensing layer reduces alert volatility and prevents premature service state transitions.
2. Confidence-gated routing transforms classification into a risk-aware selective decision that preserves stream purity under ambiguous evidence.
3. Mechanical timing constraints dominate throughput feasibility, emphasizing that routing evaluation must extend beyond the neural inference.

The contribution boundary is therefore explicit. The benchmark was controlled and aligned with the prototype. Perturbation analysis can approximate the visual domain shift, but it cannot substitute for field experiments. The reject mechanism mitigates risk for the included unsupported categories but does not address unrestricted open-world recognition.

Accordingly, the results support prototype-level viability within a restricted operational envelope. The main contribution is methodological: exposing measurement error, calibration behavior, rejection policy, and actuation verification as inspectable elements of a compact selective routing system.

From a sustainability point of view, the prototype has three practical consequences. First, calibrated fill level detection eliminates wasteful collection through stable service notifications. Second, confidence-gated routing treats stream contamination as an explicit decision cost, which is closer to recycling quality than a forced high-coverage classification. Third, perturbation and rejection-aware analyses reveal when the system should limit the automated routing coverage, instead of silently misrouting uncertain items. The

https://doi.org/10.3390/su18147498

implications are still at the prototype stage and do not substitute for field measurements of contamination reduction, recovery improvement, or lifecycle cost.

### *6.10. Analytical Sustainability Impact Modeling Under Selective Routing*

To connect the selective routing behavior to sustainability-oriented outcomes without adding further table density, this section uses a compact analytical contamination exposure model. The model does not claim field-measured recovery gains; instead, it estimates the extent to which automatic misrouting exposure is reduced when low-confidence items are rejected rather than forced into recyclable streams.

The expected automatic misrouting exposure can be approximated by the routing coverage times the residual routed error rate using the stated final policy. At tau = 0.70 on the held-out test set, routing coverage among routing class items was 76.09% and routed precision was 92.66%, giving an estimated exposure of $0.7609 \times (1 - 0.9266) = 0.0559$, or approximately 5.6%. This value is analytical rather than field-measured, but it shows how the selective policy trades lower automatic throughput for reduced unsafe routing exposure under the controlled benchmark. The rejection false acceptance was 1.34%, indicating that unsupported cardboard/paper items were rarely routed into the implemented material streams.

Recent publications in Sustainability have discussed automated trash classification, robotic sorting, and IoT-enabled smart waste management for improved recycling efficiency and sustainable urban infrastructure [36,42–45]. The present study follows this journal line but narrows the contribution to decision boundary formalization. Instead of claiming a full-scale operational impact, it evaluates how calibrated sensing, calibrated confidence, rejection behavior, and actuation verification can reduce the estimated misrouting exposure under a transparent prototype-level boundary.

### *6.11. Theoretical Implications*

This study contributes theoretically by reframing sustainable smart waste sorting as a sensing-to-actuation decision problem under bounded uncertainty rather than as a conventional combination of smart bin monitoring and closed-set image classification. This framing extends the literature on AI-enabled waste management by emphasizing the decision interface between measurement, prediction, rejection, and physical routing. In doing so, it links three research streams that are often treated separately: calibrated IoT sensing, confidence-aware visual recognition, and selective actuation.

The findings also support the theoretical value of reject-aware and calibration-aware evaluation in recycling-oriented automation. Instead of interpreting classification accuracy as the primary indicator of sustainability relevance, the study shows that routed precision, reject recall, false acceptance, calibration error, and perturbation response provide a more operationally meaningful basis for evaluating whether an automated sorting prototype can protect accepted streams from uncertain or unsupported inputs. This perspective positions uncertainty handling as a central design principle for sustainable cyber-physical waste systems.

### *6.12. Practical Implications*

From a practical perspective, the proposed prototype provides a transparent design logic for low-cost smart waste systems that need to avoid unsafe automatic routing under uncertainty. The calibrated fill level monitoring layer can help service operators avoid unstable alert publication caused by noisy ultrasonic readings, while the confidence-gated sorting layer provides a mechanism for diverting uncertain items to a mixed/reject stream instead of forcing them into recyclable outputs.

For prototype developers and municipal technology planners, the results suggest that practical smart waste systems should not be assessed only by classification accuracy or bin connectivity functionality. They should also document sensing calibration, alert persistence, confidence calibration, rejection behavior, routing verification, and actuator-cycle boundaries. These elements are important for moving from demonstration-level AI-IoT integration toward auditable and sustainability-oriented waste-management automation. However, the practical interpretation remains limited to bench-scale conditions until the approach is validated in multi-object, multi-bin, and field-deployment settings.

## 7. Limitations and Future Work

Despite the structured evaluation and coherent bench-scale integration demonstrated in this study, several limitations define the operational boundary of the proposed prototype. First, the image-based evaluation relied on a controlled five class split derived from TrashNet. Although this design supports reproducibility and routing-specific benchmarking, it does not fully represent the heterogeneity of municipal waste streams, where overlapping objects, deformation, contamination, background clutter, severe illumination variation, occlusion, and dynamic camera geometry are common [25–32,36–40]. Accordingly, the reported classification and selective routing results should be interpreted within a controlled-imaging context rather than as field-generalized sustainability performance. The reported reject performance should also not be interpreted as complete open-world recognition because cardboard and paper were used as outlier-exposure samples for the reject class. A stronger test of unseen category rejection would require leave-one-reject-category-out or broader open-set evaluation protocols. Similarly, the reported routed precision should be interpreted under the single-item inspection assumption; it does not directly transfer to cluttered, overlapping, or multi-object conveyor scenes without additional detection, segmentation, or singulation modules.

Second, the sensing, threshold, and operational-timing components remain specific to the bench-scale configuration. The ultrasonic calibration values are tied to the bin geometry, internal measurement height, sensor mounting position, and surface conditions used in this prototype; therefore, recalibration and hysteresis margin verification would be required for different bin designs. The confidence threshold was selected using the present validation split, while bootstrap resampling was used to quantify uncertainty in reported held-out metrics. However, bootstrap analysis does not replace a full multi-seed or repeated split threshold stability study. Continuous actuator event logs and a fresh standalone batch-1 latency benchmark were also not available for this manuscript. Therefore, the reported evidence should be interpreted as module-level and prototype-level rather than as long-duration operational benchmarking. Long-term reliability, communication resilience under unstable networks, hardware wear, power efficiency, and scalability across distributed bins and sorting stations remain open engineering questions for future deployment-oriented studies [12–18,31,32].

Third, practical sustainability impact was not directly measured in field conditions. The manuscript formalizes contamination risk through selective rejection and models stream purity exposure under threshold variation, but it does not report field-measured impurity rates, recovery gains, collection route efficiency improvements, emissions effects, lifecycle cost, or economic savings. Such indicators require pilot-scale implementation under operational conditions. Future work should therefore proceed along three complementary axes: expanding visual realism through multi-object, cluttered, and broader open-set datasets; scaling the operational workflow through detection-based routing, edge device inference benchmarking, and actuator robustness testing; and conducting pilot deployments to evaluate contamination reduction, recovery improvement, route optimization, and cost–benefit performance under real operating conditions. The present work defines these boundaries to position the contribution as a transparent prototype-level method rather than a claim of municipal-scale readiness.

## 8. Conclusions

This study developed and evaluated a measurement-calibrated, reject-aware selective routing prototype for sustainable smart waste sorting under controlled bench-scale conditions. The prototype links ultrasonic fill level sensing, edge stabilization, cloud-supported alert logic, image-based material recognition, calibrated rejection, and actuator-level verification within a single sensing-to-actuation workflow. Rather than treating IoT monitoring and image classification as parallel subsystems, the study reframes smart waste sorting as a bounded risk-aware control problem in which sensing uncertainty and predictive uncertainty must be made visible before physical routing is executed.

The experimental results support the technical coherence of this prototype-level approach. The sensing layer reported explicit calibration, repeatability, hysteresis, and buffered communication evidence, while the vision layer used true logit temperature scaling, reject-aware classification, validation-selected thresholding, cost-sensitive policy interpretation, and perturbation testing. Under the adopted five class benchmark, the final VGG19 model achieved 88.13% test accuracy, 87.84% balanced accuracy, 88.15% weighted F1, 96.60% reject class F1, and test expected calibration error (ECE) = 0.0349 after temperature scaling. With the final tau = 0.70 policy, the system routed 76.09% of routing class test items with 92.66% routed precision, while reject recall reached 98.66%. These results indicate that selective routing can reduce silent misrouting exposure by shifting uncertain or unsupported cases toward the mixed/reject stream.

The contribution of this work is therefore not a new neural architecture or a municipal-scale smart waste platform, but an uncertainty-explicit prototype-evaluation methodology for deciding when automated routing is justified within a bounded cyber-physical waste-sorting system. The findings should be interpreted as analytical and prototype-level evidence, not as field-measured proof of municipal contamination reduction, lifecycle improvement, economic savings, or full deployment readiness. Future work should extend the framework through multi-object and cluttered datasets, detection-based routing, broader open-set uncertainty evaluation, edge device benchmarking, long-duration actuator logging, and pilot deployments that directly measure contamination reduction, recovery improvement, route efficiency, and cost–benefit performance.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su18147498/s1, Supplementary File S1, which includes split index documentation, final reject-aware prediction outputs, raw true logit outputs, true logit temperature-scaling diagnostics, threshold-policy results, perturbation summaries, training histories, and the final policy summary used to support the reproducibility of the reported prototype-level experiments.

**Funding:** This research received no external funding for the study design, data collection, analysis, interpretation, or manuscript preparation. The article processing charge (APC) is supported by the Deanship of Graduate Studies and Scientific Research at Qassim University (www.qu.edu.sa) under Support Code QU-APC-2026.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The image benchmark used in this study was derived from the publicly available TrashNet Dataset. The five class partition counts, evaluation metrics, threshold policy results, calibration diagnostics, true logit outputs, perturbation summaries, and prototype-level verification results are reported in this manuscript. Supplementary File S1 provides supporting material related to split documentation, prediction-output summaries, raw logit exports, true logit temperature scaling, perturbation outputs, threshold policy analysis, final policy selection, and training histories. Additional prototype logs and trained model files can be made available from the corresponding author upon reasonable request, subject to institutional and repository constraints.

**Acknowledgments:** The author would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University (www.qu.edu.sa) for financial support (QU-APC-2026).

**Conflicts of Interest:** The author declares no conflicts of interest.

## References

1. Nanda, S.; Berruti, F. Municipal solid waste management and landfilling technologies: A review. *Environ. Chem. Lett.* **2021**, *19*, 1433–1456. [CrossRef]
2. Vinti, G.; Bauza, V.; Clasen, T.; Medlicott, K.; Tudor, T.; Zurbrugg, C.; Vaccari, M. Municipal solid waste management and adverse health outcomes: A systematic review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 4331. [CrossRef] [PubMed]
3. Abubakar, I.R.; Maniruzzaman, K.M.; Dano, U.L.; AlShihri, F.S.; AlShammari, M.S.; Ahmed, S.M.S.; Al-Gehlani, W.A.G.; Alrawaf, T.I. Environmental sustainability impacts of solid waste management practices in the global South. *Int. J. Environ. Res. Public Health* **2022**, *19*, 12717. [CrossRef] [PubMed]
4. Javed, A.R.; Shahzad, F.; Rehman, S.U.; Zikria, Y.B.; Razzak, I.; Jalil, Z.; Xu, G. Future smart cities: Requirements, emerging technologies, applications, challenges, and future aspects. *Cities* **2022**, *129*, 103794. [CrossRef]
5. Lakhouit, A. Revolutionizing urban solid waste management with AI and IoT: A review of smart solutions for waste collection, sorting, and recycling. *Results Eng.* **2025**, *25*, 104018. [CrossRef]
6. El Jaouhari, A.; Samadhiya, A.; Kumar, A.; Mulat-Weldemeskel, E.; Luthra, S.; Kumar, R. Turning trash into treasure: Exploring the potential of AI in municipal waste management—An in-depth review and future prospects. *J. Environ. Manag.* **2025**, *373*, 123658. [CrossRef] [PubMed]
7. Zhao, Y.; Li, X. Sensor-Based Technologies in Effective Solid Waste Sorting: Successful Applications, Sensor Combination, and Future Directions. *Environ. Sci. Technol.* **2022**, *56*, 17524–17538. [CrossRef] [PubMed]
8. Giel, R.; Fiedeń, M.; Dąbrowska, A. Real-Time Automatic Identification of Plastic Waste Streams for Advanced Waste Sorting Systems. *Sustainability* **2025**, *17*, 2157. [CrossRef]
9. Alexakis, G.; Pellegrino, M.; Rodriguez-Turienzo, L.; Maniadakis, M. Enhanced Waste Sorting Technology by Integrating Hyperspectral and RGB Imaging. *Recycling* **2025**, *10*, 179. [CrossRef]
10. Guo, C.; Pleiss, G.; Sun, Y.; Weinberger, K.Q. On Calibration of Modern Neural Networks. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; Proceedings of Machine Learning Research; Volume 70, pp. 1321–1330. Available online: https://proceedings.mlr.press/v70/guo17a.html (accessed on 16 May 2026).
11. Geifman, Y.; El-Yaniv, R. SelectiveNet: A Deep Neural Network with an Integrated Reject Option. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; Proceedings of Machine Learning Research; Volume 97, pp. 2151–2159. Available online: https://proceedings.mlr.press/v97/geifman19a.html (accessed on 16 May 2026).
12. Kannan, D.; Khademolqorani, S.; Janatyan, N.; Alavi, S. Smart waste management 4.0: The transition from a systematic review to an integrated framework. *Waste Manag.* **2024**, *174*, 1–14. [CrossRef] [PubMed]
13. Fang, B.; Yu, J.; Chen, Z.; Osman, A.I.; Farghali, M.; Ihara, I.; Hamza, E.H.; Rooney, D.W.; Yap, P.-S. Artificial intelligence for waste management in smart cities: A review. *Environ. Chem. Lett.* **2023**, *21*, 1959–1989. [CrossRef] [PubMed]
14. Dawar, I.; Srivastava, A.; Singal, M.; Dhyani, N.; Rastogi, S. A systematic literature review on municipal solid waste management using machine learning and deep learning. *Artif. Intell. Rev.* **2025**, *58*, 183. [CrossRef]
15. Abuga, D.; Raghava, N.S. Real-time smart garbage bin mechanism for solid waste management in smart cities. *Sustain. Cities Soc.* **2021**, *75*, 103347. [CrossRef]
16. Hannan, M.A.; Begum, R.A.; Al-Shetwi, A.Q.; Ker, P.J.; Al Mamun, M.A.; Hussain, A.; Basri, H.; Mahlia, T.M.I. Waste collection route optimisation model for linking cost saving and emission reduction to achieve sustainable development goals. *Sustain. Cities Soc.* **2020**, *62*, 102393. [CrossRef]
17. Ijemaru, G.K.; Ang, L.M.; Seng, K.P. Transformation from IoT to IoV for waste management in smart cities. *J. Netw. Comput. Appl.* **2022**, *204*, 103393. [CrossRef]
18. Alourani, A.; Ashraf, M.U.; Aloraini, M. Smart waste management and classification system using advanced IoT and AI technologies. *PeerJ Comput. Sci.* **2025**, *11*, e2777. [CrossRef] [PubMed]
19. Ali, T.; Irfan, M.; Alwadie, A.S.; Glowacz, A. IoT-based smart waste bin monitoring and municipal solid waste management system for smart cities. *Arab. J. Sci. Eng.* **2020**, *45*, 10185–10198. [CrossRef]

20. Catarinucci, L.; Colella, R.; Consalvo, S.I.; Patrono, L.; Rollo, C.; Sergi, I. IoT-aware waste management system based on cloud services and ultra-low-power RFID sensor-tags. *IEEE Sens. J.* **2020**, *20*, 14873–14881. [CrossRef]
21. Zhang, Q.; Zhang, X.; Mu, X.; Wang, Z.; Tian, R.; Wang, X.; Liu, X. Recyclable waste image recognition based on deep learning. *Resour. Conserv. Recycl.* **2021**, *171*, 105636. [CrossRef]
22. Zhang, Q.; Yang, Q.; Zhang, X.; Bao, Q.; Su, J.; Liu, X. Waste image classification based on transfer learning and convolutional neural network. *Waste Manag.* **2021**, *135*, 150–157. [CrossRef] [PubMed]
23. Mao, W.-L.; Chen, W.-C.; Wang, C.-T.; Lin, Y.-H. Recycling waste classification using optimized convolutional neural network. *Resour. Conserv. Recycl.* **2021**, *164*, 105132. [CrossRef]
24. Zhang, S.; Chen, Y.; Yang, Z.; Gong, H. Computer vision based two-stage waste recognition-retrieval algorithm for waste classification. *Resour. Conserv. Recycl.* **2021**, *169*, 105543. [CrossRef]
25. Qiao, Y.; Zhang, Q.; Qi, Y.; Wan, T.; Yang, L.; Yu, X. A waste classification model in low-illumination scenes based on ConvNeXt. *Resour. Conserv. Recycl.* **2023**, *199*, 107274. [CrossRef]
26. Lin, Y.-H.; Mao, W.-L.; Fathurrahman, H.I.K. Development of intelligent municipal solid waste sorter for recyclables. *Waste Manag.* **2024**, *174*, 597–604. [CrossRef] [PubMed]
27. Fotovvatikhah, F.; Ahmedy, I.; Noor, R.M.; Munir, M.U. A systematic review of AI-based techniques for automated waste classification. *Sensors* **2025**, *25*, 3181. [CrossRef] [PubMed]
28. Sayem, F.R.; Islam, M.S.B.; Naznine, M.; Nashbat, M.; Hasan-Zia, M.; Kunju, A.K.A.; Khandakar, A.; Ashraf, A.; Majid, M.E.M.; Kashem, S.B.A.; et al. Enhancing waste sorting and recycling efficiency: Robust deep learning-based approach for classification and detection. *Neural Comput. Appl.* **2025**, *37*, 4567–4583. [CrossRef]
29. Le, H.T.N.; Ngo, H.Q.T. Application of the vision-based deep learning technique for waste classification using the robotic manipulation system. *Int. J. Cogn. Comput. Eng.* **2025**, *6*, 391–400. [CrossRef]
30. Gurcan, F.; Soylu, A. Automated waste classification for smart recycling: A multi-class CNN approach with transfer learning and pre-trained models. *Environ. Technol. Innov.* **2026**, *41*, 104673. [CrossRef]
31. Ghosh, D.; Goswami, A. Enhanced deep learning framework for efficient garbage classification in smart waste management systems. *Inf. Sci.* **2025**, *719*, 122462. [CrossRef]
32. Kumar, Y.; Bhardwaj, P.; Malhotra, S.; Prasad, A.; Kim, W.; Ijaz, M.F. Deep residual and hybrid CNN models for confidence-aware real-world waste classification for sustainable waste management. *Sci. Rep.* **2026**, *16*, 10424. [CrossRef] [PubMed]
33. Henaien, A.; Elhadj, H.B.; Fourati, L.C. A sustainable smart IoT-based solid waste management system. *Future Gener. Comput. Syst.* **2024**, *157*, 587–602. [CrossRef]
34. Ishaq, A.; Mohammad, S.J.; Bello, A.A.D.; Wada, S.A.; Adebayo, A.; Jagun, Z.T. Smart waste bin monitoring using IoT for sustainable biomedical waste management. *Environ. Sci. Pollut. Res.* **2025**, *32*, 19434–19449. [CrossRef] [PubMed]
35. Abo-Zahhad, M.M.; Abo-Zahhad, M. Real time intelligent garbage monitoring and efficient collection using Yolov8 and Yolov5 deep learning models for environmental sustainability. *Sci. Rep.* **2025**, *15*, 16024. [CrossRef] [PubMed]
36. Arishi, A. Real-Time Household Waste Detection and Classification for Sustainable Recycling: A Deep Learning Approach. *Sustainability* **2025**, *17*, 1902. [CrossRef]
37. Ren, Y.; Li, Y.; Gao, X. An MRS-YOLO Model for High-Precision Waste Detection and Classification. *Sensors* **2024**, *24*, 4339. [CrossRef] [PubMed]
38. Son, J.; Ahn, Y. AI-based plastic waste sorting method utilizing object detection models for enhanced classification. *Waste Manag.* **2025**, *193*, 273–282. [CrossRef] [PubMed]
39. Moktar, M.H.; Mohamed, H.; Hajjaj, S.S.H.; Baharuddin, M.Z. Medical waste sorting machine development with IoT and YOLO model utilization. *J. Eng. Appl. Sci.* **2025**, *72*, 86. [CrossRef]
40. Miski, A. Lightweight YOLO object detectors for PET and HDPE classification in recycling facilities. *Sci. Rep.* **2026**, *16*, 2745. [CrossRef] [PubMed]
41. Thung, G.; Yang, M. *Classification of Trash for Recyclability Status*; Stanford University CS229 Project Report; [Online]; Stanford University: Stanford, CA, USA, 2016. Available online: https://github.com/garythung/trashnet (accessed on 16 May 2026).
42. Cheema, S.M.; Hannan, A.; Pires, I.M. Smart waste management and classification systems using cutting edge approach. *Sustainability* **2022**, *14*, 10226. [CrossRef]
43. Ahmed, M.I.B.; Alotaibi, R.B.; Al-Qahtani, R.A.; Al-Qahtani, R.S.; Al-Hetela, S.S.; Al-Matar, K.A.; Al-Saqer, N.K.; Rahman, A.; Saraireh, L.; Youldash, M.; et al. Deep learning approach to recyclable products classification: Towards sustainable waste management. *Sustainability* **2023**, *15*, 11138. [CrossRef]

44. Cheng, T.; Kojima, D.; Hu, H.; Onoda, H.; Pandyaswargo, A.H. Optimizing waste sorting for sustainability: An AI-powered robotic solution for beverage container recycling. *Sustainability* **2024**, *16*, 10155. [CrossRef]
45. Alnanih, R.; Elrefaei, L.; Al-Ahwal, A. Advancing sustainability through an IoT-driven smart waste management system with software engineering integration. *Sustainability* **2025**, *17*, 9803. [CrossRef]