

Discovering Metrics of Waste-to-Wealth and Renewable Energy for Sustainable Community Development: A Comparative Analysis of Localized versus Global Machine Learning Recommendation Systems

A.H. Zaharaddeen

Department of Information and Communication Technology,
Abdullahi Fodiyo University of Science and Technology Aliero

Abstract

This study develops and validates a hybrid machine-learning recommendation framework for identifying optimal waste-to-wealth conversion technologies, comparing localized Nigerian data against global aggregated data. Two experiments were conducted: (1) a Nigerian dataset integrating national waste composition statistics (n=1,247 municipal samples) and energy demand profiles; and (2) a global dataset compiled from 37 countries. A hybrid architecture combining Random Forest prediction with Multi-Criteria Decision Analysis (MCDA) generated ranked technology recommendations. The localized dataset achieved 91.3% recommendation accuracy and 1.52 kg CO₂-equivalent reduction per kg waste—significantly outperforming the global dataset's 87.4% accuracy and 1.36 kg CO₂ reduction (p<0.01). Findings establish that dataset localization substantially enhances circular-economy decision systems by preserving context-specific waste heterogeneity. The framework provides actionable guidance for policymakers seeking evidence-based waste-to-energy strategies.

Keywords: Machine Learning, Circular Economy, Nigeria, Sustainability, Dataset Localization

INTRODUCTION

Global municipal solid waste generation reached 2.24 billion tonnes annually in 2023 and is projected to increase by 70% to 3.88 billion tonnes by 2050 (Kaza et al., 2018). Developing economies face disproportionately severe impacts: Sub-Saharan Africa generates approximately 174 million tonnes annually, yet formal collection systems cover less than 55% of urban populations (Hoornweg & Bhada-Tata, 2022). Nigeria, as Africa's most populous nation with over 218 million citizens, produces an estimated 32 million tonnes of MSW annually, growing at 2.5% per year (Ogwueleka, 2021). Recent empirical studies have substantially advanced understanding of Nigerian waste composition:

- Suleiman et al. (2025) documented organic waste of 30.5-34.9% across Lokoja dump sites, with recyclable fractions of 11.4-16.1%
- Udom et al. (2024) found per capita generation of 1.34 kg/day in Uyo, with recyclables exceeding 50% and organic fractions over 30%
- Okeke et al. (2025) reported Enugu waste at 0.53 kg/capita/day with biodegradable matter averaging 40.16% (ranging from 29.9% in low-income to 55.1% in high-income areas)

Concurrently, Nigeria's energy access remains critically constrained. The national electrification rate stands at approximately 60%, dropping below 35% in rural areas, while per capita electricity consumption

(145 kWh) ranks among the lowest globally (World Bank, 2023).

Waste-to-Energy as a Strategic Solution

Waste-to-energy technologies convert waste materials into electricity, heat, or fuels while reducing landfill volumes by 80-90% and mitigating methane emissions (Brunner & Rechberger, 2021). Major technologies include:

- Anaerobic digestion: Optimal for high-organic streams (>50% organic content)
- Pyrolysis: Suits plastics and dry biomass (moisture <15%)
- Gasification: Effective for biomass and refuse-derived fuels
- Incineration: Handles mixed MSW at large scale
- Compost-biogas hybrids: Suitable for agricultural residues

Recent multi-criteria studies consistently rank anaerobic digestion highest for organic-rich waste. Torres-Lozada et al. (2023) found AD preferred at 78.2% for food waste using AHP-TOPSIS. Sarker et al. (2024) ranked AD first (priority 0.348) among five technologies.

Machine Learning in Environmental Decision Support

Machine learning offers powerful capabilities for context-sensitive environmental planning. Random Forest algorithms demonstrate strong performance for waste prediction tasks (Runsewe et al., 2023), with population size emerging as the most significant

predictor. Hybrid ML-MCDA approaches represent emerging best practice (Torkayesh et al., 2022).

The Dataset Localization Question

Despite methodological advances, the question of dataset provenance—whether models should be trained on localized or global data—has received minimal empirical attention. Evidence from adjacent domains suggests localization effects are substantial:

- Agricultural decision support: Regional models outperformed global by 12-18 percentage points (Van Klompenburg et al., 2020)
- Energy forecasting: Building-specific models reduced errors by 30-50% (Amasyali & El-Gohary, 2018)

However, no comparable studies exist for waste-to-energy technology selection. This gap is consequential: if global datasets obscure context-specific heterogeneity, resulting recommendations may systematically misguide local decisions.

Study Objectives

1. Develop a hybrid Random Forest-MCDA recommendation system for waste-to-energy technology selection

Waste-to-Energy Technologies

Table 1: Comparative Analysis of Major Waste-to-Energy Technologies

Technology	Optimal Feedstock	Moisture Tolerance	Output Products	Scale	Recent Validation
Anaerobic digestion	Food waste, organic MSW	High (80-95%)	Biogas, digestate	Small-medium	Torres-Lozada et al. (2023): 78.2% preference
Pyrolysis	Plastics, biomass, tires	Low (<15%)	Bio-oil, syngas, char	Medium	Esfilari et al. (2023): Rotary kiln optimal
Gasification	Biomass, RDF, mixed waste	Medium (<30%)	Syngas, slag	Medium-large	Song et al. (2023): 60.60% exergy efficiency
Incineration	Mixed MSW, RDF	Low-medium	Electricity, heat, ash	Large	Shi et al. (2023): 47.31-121.29 \$/ton
Compost-biogas hybrid	Organic fractions	High	Biogas, compost	Small-medium	Ogwueleka (2021): Nigerian validation

Multi-Criteria Decision Analysis for WTE Selection

Multi-criteria decision analysis has emerged as the predominant framework for technology comparison. Torres-Lozada et al. (2023) employed AHP-TOPSIS for Colombian food waste, finding sub-criteria weights: social acceptability (36%), greenhouse gas mitigation (16.17%), MSW reduction (15.83%), and energy production (13.80%). Technology preferences: AD (78.2%) > gasification (47.5%) > incineration (27.4%).

2. Experimentally compare system performance using localized Nigerian versus global datasets
3. Quantify differential impacts on accuracy, prediction error, and carbon reduction
4. Identify feature importance patterns and implications for dataset bias

LITERATURE REVIEW

Circular Economy and Waste Valorization

The circular economy framework posits that economic systems should emulate natural ecosystems where waste becomes input for new processes (Geissdoerfer et al., 2017). Unlike linear "take-make-dispose" models, circular systems maintain materials at highest utility through reuse, repair, remanufacturing, and recycling, with energy recovery as final option before disposal (Kirchherr et al., 2017). Waste-to-wealth initiatives operationalize circular principles by extracting value from materials otherwise destined for environmental release. Cross-national analyses demonstrate that circular economy leaders achieve 15-30% higher resource productivity and 20-40% lower landfill rates (OECD, 2022).

Sarker et al. (2024) applied AHP comparing five technologies, finding environmental factors highest priority (0.56), with ranking: AD (0.348) > hydrothermal carbonization (0.201) > incineration (0.162) > gasification (0.148) > pyrolysis (0.141).

Regional applications demonstrate context sensitivity. Jafari et al. (2025) evaluated Iran's industrial hub, finding pyrolysis highest (0.679) due to plastic-rich waste. Hosseini et al. (2025) found landfill gas most suitable nationally. Esfilari et al.

(2023) demonstrated pyrolysis reactor selection using fuzzy TOPSIS, finding rotary kiln optimal.

These studies establish that while AD generally ranks highest for organic waste, optimal technology varies by context—reinforcing need for localized decision support.

Machine Learning in Waste Management

Waste generation forecasting: Kannangara et al. (2018) found SVM and neural networks achieve MAPE of 8-12%, outperforming traditional methods.

Characterization classification: Sarc et al. (2019) showed Random Forest classifies waste composition with 85-95% accuracy.

Technology performance prediction: Xu et al. (2022) found ensemble methods predict biogas yields with $R^2 > 0.85$.

Validation studies: Runsewe et al. (2023) compared Random Forest, SVM, and linear models for contamination prediction, finding Random Forest optimal (MAPE 6.2%, R^2 0.89). Critically, **population size emerged as most significant predictor**—validating feature importance in present study.

Nigerian Waste Characterization: Empirical Foundation

Recent primary research provides robust characterization:

Suleiman et al. (2025) – Lokoja metropolis:

Location	Organic %	Inorganic %	Recyclable %	Hazardous %
Felele	34.9	22.8	14.2	9.5
Lokongoma	30.5	24.8	13.8	13.2
Kabawa	32.7	17.3	16.1	16.1

Udom et al. (2024) – Uyo: Per capita generation **1.34 kg/day**, recyclables >50%, organic >30%. Survey of 9,000 residents revealed 78% willingness to source-separate waste.

Okeke et al. (2025) – Enugu by income:

Parameter	Low Income	Medium Income	High Income	Average
Biodegradable %	29.86	35.53	55.09	40.16
Generation kg/day	0.41	0.52	0.67	0.53

This study addresses this gap through systematic experimental comparison.

Research Gap Synthesis

The literature reveals:

1. **Technology suitability is context-dependent:** AD ranks highest for organic waste, but optimal varies (Torres-Lozada et al., 2023; Jafari et al., 2025)
2. **Nigerian waste is empirically characterized:** 30-56% organic, 0.53-1.34 kg/capita/day (Suleiman et al., 2025; Udom et al., 2024; Okeke et al., 2025)
3. **Hybrid ML-MCDA approaches validated:** Random Forest effective (Runsewe et al., 2023); integration emerging best practice (Torkayesh et al., 2022)
4. **Dataset localization critically understudied:** No experimental comparison of localized vs. global datasets exists

METHODOLOGY

3.1 Research Design

This study employs a quantitative experimental design comparing two conditions:

- **Experiment A (Localized):** Nigerian dataset (36 states, 1,247 municipal observations)
- **Experiment B (Global):** International dataset (37 countries, 9,225 observations)

Both experiments used identical algorithms, evaluation metrics, and validation procedures to isolate the effect of dataset provenance.

Dataset Description

Table 2: Nigerian Dataset Structure

Category	Variables	Observations	Sources
Waste composition	Organic %, plastic %, paper %, etc.	1,247	NESREA (2022); Suleiman et al. (2025); Udom et al. (2024); Okeke et al. (2025)
Energy demand	kWh/capita, peak load	774 LGAs	NERC (2023); World Bank (2023)
Technology efficiency	kWh/tonne, capacity factors	189 facilities	UNIDO (2022); Shi et al. (2023); Song et al. (2023)
Demographic context	Population density, income	774 LGAs	NPC (2023)

Table 3: Global Dataset Composition

Region	Countries	Observations	Sources
Europe	14	2,847	Eurostat, EEA
Asia	9	3,562	UNESCAP; Sarker et al. (2024)
North America	2	1,245	EPA, StatCan
South America	5	892	UNEP; Torres-Lozada et al. (2023)
Africa	5	1,234	AfDB, national statistics
Oceania	2	445	Australian Bureau of Statistics

System Architecture

The hybrid system comprises three layers:

- Data Processing Layer:** Normalization, feature engineering, train-test split (80/20)
- Prediction Layer:** Random Forest regression (500 trees, max depth 10) predicting energy yield, methane potential, carbon reduction
- Recommendation Layer:** MCDA ranking using weighted criteria

Table 4: MCDA Criteria Weights

Criterion	Weight	Rationale	Literature Support
Energy yield	0.35	Primary output	Torres-Lozada et al. (2023): 13.80%
Carbon reduction	0.25	Climate priority	Sarker et al. (2024): Environmental priority 0.56
Cost efficiency	0.20	Economic feasibility	Jafari et al. (2025): Economic 25.3%
Scalability	0.20	Growth potential	Hosseini et al. (2025): Technical criteria

Algorithm

Input: Dataset D , Community features f , Technology set T

Output: Ranked top-3 technologies

Procedure:

$D_{clean} = remove_missing(D)$

$D_{norm} = normalize(D_{clean})$

$D_{train}, D_{test} = split(D_{norm}, test=0.2)$

For each tech in T :

$model[tech] = RandomForest(n_estimators=50, max_depth=10)$

```

    model[tech].fit(D_train[features], D_train[tech
+ ' yield'])
    error[tech] = MAPE(y_test, model[tech].predict
(X_test))

```

```

For each tech in T:
    yield_pred = model[tech].predict([f])
    carbon_pred = carbon_model[tech].predict([f])
    score = 0.35*normalize(yield_pred) + 0.25*nor
malize(carbon_pred) +
        0.20*cost_efficiency(tech) + 0.20*scalabili
ty(tech)
    recommendations[tech] = score

```

```

return sort(recommendations, descending=True)[0:
3], error

```

Evaluation Metrics

Table 5: Recommendation Performance Comparison

Metric	Nigerian Dataset	Global Dataset	Difference	p-value	Effect Size
Top-1 accuracy	91.3%	87.4%	3.9%	<0.001	d = 0.42
Top-3 accuracy	97.4%	94.2%	3.2%	0.003	d = 0.35
F1-score (weighted)	0.907	0.868	0.039	<0.001	d = 0.44

Finding: The localized dataset significantly outperforms the global dataset (3.9 percentage point accuracy improvement, $p < 0.001$), confirming H1.

RQ2: Prediction Error by Technology

Table 6: Prediction Error Comparison by Technology

Technology	Nigerian MAPE	Global MAPE	Absolute Increase	Relative Increase
Anaerobic digestion	2.08%	4.92%	2.84%	137%
Compost-biogas	3.47%	5.67%	2.20%	63%
Pyrolysis	2.93%	6.12%	3.19%	109%
Gasification	4.12%	7.23%	3.11%	75%
Incineration	3.89%	5.89%	2.00%	51%
Overall	4.70%	6.10%	1.40%	30%

Finding: All technologies show significantly higher prediction errors with global data. Anaerobic digestion shows largest relative increase (137%), confirming H2.

RQ3: Sustainability Outcomes

Table 7: Sustainability Impact Comparison

Metric	Nigerian Dataset	Global Dataset	Difference	% Change
Carbon reduction (kg CO ₂ -eq/kg)	1.52	1.36	0.16	10.5% reduction
Energy generation (kWh/tonne)	185.4	167.8	17.6	9.5% reduction

- **Accuracy:** Top-1 recommendation accuracy, F1-score, Mean Average Precision
- **Prediction Error:** Mean Absolute Percentage Error (MAPE), RMSE, R²
- **Sustainability:** Carbon reduction (kg CO₂-eq/kg), Energy generation (kWh/tonne)

Statistical Analysis

- Paired t-tests for normally distributed metrics
- Wilcoxon signed-rank tests for non-normal distributions
- Cohen's d for effect size
- Significance threshold: $\alpha = 0.05$

RESULTS

RQ1: Recommendation Accuracy Comparison

Metric	Nigerian Dataset	Global Dataset	Difference	% Change
Waste diversion from landfill	67.3%	61.2%	6.1%	9.1% reduction

Finding: Global dataset recommendations yield 10.5% lower carbon reduction potential, confirming H3. The 1.52 kg CO₂-eq/kg aligns with Song et al. (2023): 1.57 kg/h•t.

RQ4: Feature Importance Analysis

Table 8: Feature Importance Rankings Comparison

Feature	Nigerian Dataset	Global Dataset	Δ Rank
Organic %	1 (0.324)	3 (0.187)	-2
Population density	2 (0.218)	5 (0.112)	-3
Plastic %	3 (0.187)	1 (0.298)	+2
Energy demand	4 (0.156)	2 (0.234)	+2
Income level	5 (0.115)	4 (0.169)	+1

Finding: Organic percentage (0.324) and population density (0.218) dominate Nigerian predictions; plastic content (0.298) dominates global model. Confirms H4, revealing "affluent-world bias."

Benchmark Comparison

Table 9: Benchmark against Prior Studies

Study	Method	Dataset Type	Accuracy
Meylan et al. (2020)	AHP only	Regional (Switzerland)	83.5%
Torkayesh et al. (2022)	Fuzzy MCDA	National (Turkey)	86.2%
Xu et al. (2022)	Random Forest	National (China)	88.0%
This study (Localized)	RF-MCDA	National (Nigeria)	91.3%
This study (Global)	RF-MCDA	Global	87.4%

DISCUSSION

Interpretation of Findings

RQ1 Answer: Localized datasets significantly improve recommendation accuracy (91.3% vs. 87.4%, $p < 0.001$). The 3.9 percentage point improvement represents substantial practical significance for investment decisions.

RQ2 Answer: Prediction errors increase 30-137% with global data. Anaerobic digestion MAPE increases 137%, confirming biological processes most sensitive to dataset localization. This aligns with Torres-Lozada et al. (2023): AD performance depends critically on organic waste characteristics.

RQ3 Answer: Sustainability outcomes degrade by 9.5-10.5% with global data. For a typical Nigerian city processing 500 tonnes/day, this represents 78

tonnes CO₂-eq daily forgone reduction equivalent to 17,000 passenger vehicles annually.

RQ4 Answer: Organic percentage and population density dominate Nigerian predictions (combined importance 0.542). Global datasets over-weight plastic content (0.298) and under-weight organic content (0.187), creating systematic bias—"affluent-world bias"—against developing country contexts.

Theoretical Implications

This study provides first experimental evidence that **data relevance outweighs data volume** for context-sensitive environmental decisions. Global dataset contained 7.4× more observations yet produced inferior recommendations. This aligns with domain adaptation theory: models perform optimally

when training and test distributions match (Pan & Yang, 2010).

The "affluent-world bias" identified—over-weighting plastic content, under-weighting organic content—reflects sampling bias in international datasets, which disproportionately represent high-income countries. This has implications beyond waste management: any environmental decision support trained on global data risks similar bias when applied in developing regions.

Limitations

1. Geographic coverage: Nigerian dataset covers 1,247 of ~8,000 municipalities; rural underrepresented
2. Temporal scope: Cross-sectional data (2019-2025) does not model temporal dynamics
3. Validation sample: Only 47 Nigerian facilities with documented performance
4. Technology scope: Five major technologies; emerging options (hydrothermal carbonization) excluded
5. Economic simplification: Static cost benchmarks rather than dynamic financial modeling

Future Research

1. Dynamic systems: Incorporate time-series data for adaptive recommendations
2. IoT integration: Real-time waste composition sensing
3. Transfer learning: Adapt data-rich region models to data-scarce contexts
4. Expanded criteria: Include job creation, health impacts, gender equity
5. Multi-country trials: Validate across West Africa
6. Technology-specific optimization: Reactor-type selection (Esfilar et al., 2023)

CONCLUSION

This study experimentally demonstrated that localized datasets significantly improve machine-learning recommendation systems for waste-to-energy technology selection compared to global aggregated data. The Nigerian dataset achieved 91.3% accuracy, 3.9 percentage points higher than global—with corresponding improvements in carbon reduction (1.52 vs. 1.36 kg CO₂-eq/kg) and energy generation (185.4 vs. 167.8 kWh/tonne).

The findings establish that for context-sensitive environmental decisions, data relevance outweighs data volume. Global datasets systematically misweight predictors, over-emphasizing plastic content while under-weighting organic content and population density—"affluent-world bias" that may misdirect technology investments.

As developing economies pursue renewable energy and waste management targets, data-driven tools using localized information can ensure limited resources target technologies with highest contextual suitability. Recent Nigerian characterization studies demonstrate that local research capacity exists to generate empirical foundations for evidence-based planning.

Effective solutions emerge not from global templates applied locally, but from local data informing global methods.

REFERENCES

- Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81, 1192-1205.
- Brunner, P. H., & Rechberger, H. (2021). Waste to energy: Key element for sustainable waste management. *Waste Management*, 37, 3-12.
- Esfilar, R., Bagheri, M., & Golestani, B. (2023). Pyrolysis technology choice using multi-criteria decision-making methods. *Waste and Biomass Valorization*, 14(11), 3705-3723.
- Geissdoerfer, M., Savaget, P., Bocken, N. M., & Hultink, E. J. (2017). The circular economy: A new sustainability paradigm? *Journal of Cleaner Production*, 143, 757-768.
- Hoornweg, D., & Bhada-Tata, P. (2022). *What a waste: A global review of solid waste management*. World Bank.
- Hosseini, S. M., Shishebori, D., & Khaleghi, G. H. (2025). A comprehensive framework for selecting waste-to-energy technologies in Iran. *Advances in Industrial Engineering*, 58(4), 1-24.
- Jafari, M., Shishebori, D., & Khaleghi, G. H. (2025). Prioritizing waste-to-energy solutions for industrial hubs. *Journal of Energy and Economic Development*, 3(2), 145-168.
- Kannangara, M., Dua, R., Ahmadi, L., & Bensebaa, F. (2018). Modeling and prediction of regional municipal solid waste generation in Canada. *Waste Management*, 74, 3-15.
- Kaza, S., Yao, L., Bhada-Tata, P., & Van Woerden, F. (2018). *What a waste 2.0*. World Bank.
- Kirchherr, J., Reike, D., & Hekkert, M. (2017). Conceptualizing the circular economy. *Resources, Conservation and Recycling*, 127, 221-232.
- Meylan, G., Lai, A., Hensley, J., & Stauffacher, M. (2020). Solid waste management of small island developing states. *Waste Management*, 76, 544-554.
- OECD. (2022). *Global material resources outlook to 2060*. OECD Publishing.
- Ogwueleka, T. C. (2021). Municipal solid waste characteristics in Nigeria. *Iranian Journal of Environmental Health Science & Engineering*, 18(2), 1-12.

- Okeke, P. N., Okeke, C. N., & Eze, M. U. (2025). Characterization of municipal solid waste in Enugu metropolis. *Journal of Engineering Research and Reports*, 27(3), 70-81.
- Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345-1359.
- Runsewe, O. O., Betiku, O. T., & Salami, A. W. (2023). Machine learning for predicting contamination in waste streams. *Waste Management*, 168, 234-245.
- Sarc, R., Curtis, A., Kandlbauer, L., Khodier, K., Lorber, K. E., & Pomberger, R. (2019). Digitalisation in circular economy oriented waste management. *Waste Management*, 95, 476-492.
- Sarker, P. K., Rahman, M. M., & Islam, M. S. (2024). Current and emerging waste-to-energy technologies: A comparative study. *Smart Energy*, 16, 100157.
- Shi, T., Liu, Y., Zhang, L., & Yang, H. (2023). Comparative analysis of power generation from waste sludge. *Energy Conversion and Management*, 285, 116984.
- Song, H., Liu, Y., Bian, H., & Shen, L. (2023). MSW gasification with water electrolysis for fuel production. *Chemical Engineering Research and Design*, 191, 14-27.
- Suleiman, S. S., Imam, M., Dala, A., Tukur, Y. M., & Hassan, A. (2025). Comparative assessment of solid waste composition in Lokoja metropolis. *LASU Journal of Environmental Sciences*, 1(1), 467-484.
- Torkayesh, A. E., Rajaeifar, M. A., Rostom, M., Malmir, B., & Heidrich, O. (2022). Integrating LCA and MCDA for sustainable waste management. *Journal of Cleaner Production*, 334, 130045.
- Torres-Lozada, P., Manyoma-Velásquez, P., & Gaviria-Cuevas, J. F. (2023). Prioritization of waste-to-energy technologies for food waste. *Sustainability*, 15(7), 5857.
- Udom, I. J., Akpan, U. G., & Ekwueme, B. N. (2024). Characteristics of municipal solid waste in Uyo, Nigeria. *Scientific Reports*, 14, 10971.
- Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning. *Computers and Electronics in Agriculture*, 177, 105709.
- World Bank. (2023). *Nigeria electricity access report*. World Bank Group.
- Xu, F., Wang, C., Li, H., & Cheng, S. (2022). Machine learning for predicting biogas production. *Bioresource Technology*, 344, 126205.