Pixels to Policy: A Multi-Modal AI Framework for Proactive, Sustainable Waste Management Across Urban & Rural Regions

MSc Research Project
MSc Data Analytics (MSCDAD_A)

Devanshu Singh
Student ID: x23267143

School of Computing
National College of Ireland

Supervisor: David Hamill

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Devanshu Singh		
Student ID:	x23267143		
Programme:	MSc. Data Analytics	**Year:**	2024-2025
Module:	Research Practicum 2		
Supervisor:	David Hamill		
Submission Due Date:	11-08-2025		
Project Title:	Pixels to Policy: A Multi-Modal AI Framework for Proactive, Sustainable Waste Management Across Urban & Rural Regions		
Word Count:	**Page Count:** 21		

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Devanshu Singh

Date: 10-08-2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Pixels to Policy: A Multi-Modal AI Framework for Proactive, Sustainable Waste Management Across Urban & Rural Regions

Devanshu Singh
x23267143

Abstract
The rapid growth of global waste, driven by urbanization and population expansion, contributes to the increasing demands of smarter, faster, and more adaptable waste management solutions. This research proposes an AI-powered, multi-modal framework that unifies waste detection using YOLOv8s, trend/pattern forecasting using LSTM, and suggestive policy generation using Zephyr-7b-beta into a single, automated system. By analyzing images of waste, predicting future generation patterns for specific locations, and producing tailored, actionable strategies, the system empowers municipalities and communities to take proactive measures before waste issues escalate. This solution is designed to operate in both urban and rural environments, including regions where the data is scarce while shifting from reactive waste management responses to anticipatory interventions. The suggestive policies that are generated are context-specific, addressing regional needs, available resources, and sustainability goals, and are evaluated for relevance, feasibility, and innovation. Real-world validation of the solution demonstrates the framework's ability to bridge the gap between raw environmental data and practical decision-making, offering a scalable, location-agnostic solution that supports Sustainable Development Goals and enables more efficient, responsible, and inclusive waste management practices.

1. Introduction

1.1 Background and Overview
The use of Artificial Intelligence and Machine Learning technologies is changing the way various industries operate when it comes to Waste Reduction and Management. Handling and managing diverse types of waste using AI tools has already become a prominent aspect for research[1], which has brought about innovative solutions to classify and segregate different waste materials, enabling us to dispose of them efficiently. The rising level of waste generation calls for a growing demand for such applications[2], which utilizes advanced technologies to analyze the underlying problems of waste production and its management in urban and rural areas.
Existing waste management solutions rely on static rules, isolated operations for various tasks, and delayed response for policy establishment to the change in waste generation trends which can be time consuming and inefficient[3]. The role of AI and Machine learning in waste management is not just limited to industrial operations like classification of waste in factories, but it holds the potential to form complex solutions leveraging the basic operations for dedicated use case of waste management[4]. Thus, this research dwells on exploring one such niche of solving waste generation by providing a solution which controls litter management in urban and rural areas.
The solution is designed in such a way that will focus on providing an initiative-taking solution to use the existing infrastructures for classifying waste types, capturing waste generation trends, and formulating dynamic suggestive policies for handling & reducing waste. This not only provides a flexible solution but also addresses significant gaps concerning the practical implementation of accessible solutions in a single pipeline which offers high adaptability for deployment in any urban or rural region.

1.2 Problem Statement
Some of the major concerns that this research aims to address include how AI-driven methodologies & solutions can be effectively utilized for proactive waste management and reduction by overcoming the limitations of conventional systems. Traditional waste management practices often rely heavily on generalized static rules set for a country or state[3], reactive decision making, and poor adaptability of

solutions across data lacking regions for waste management. In many low-resource settings, particularly rural regions, over 2 billion people lack access to basic waste collection services[2], while municipalities often allocate 20-50% of their budgets to ineffective waste operations[2].

These challenges highlight the pressing need for scalable and accessible technological intervention which can help with tackling the litter generation in urban and rural regions. Therefore, this research focuses on developing a scalable, accessible, and adaptable automated system which will leverage classification and predictive analysis of litter generation for forming dynamic policy suggestions in real time to ensure applicability across different regions, including data-scarce environments.

1.3 Significance

The escalating volume of global waste is estimated to reach 3.4 billion tons annually by 2050[2], which poses a critical threat to sustainable development, environmental integrity as well as public health. This growth along with the rapid urbanization, expansion of the population, and increased consumerism, demands the urgent need for more intelligent, responsive, and equitable waste management systems[5].

This research is important as it focuses on developing an AI-driven, data-centric solution that utilizes the hybrid approach and integrates isolated operations like litter classification, trend analysis & prediction, and dynamic suggestive policy generation based on the forecasted litter/waste under a single pipeline. This automated system will be accessible, adaptable by any region, even where the data is scarce and will also be initiative-taking in analyzing the growing cause of litter/waste in real-time, while operating with minimum human intervention and requiring only modest infrastructure investment.

Furthermore, the research aligns directly with key United Nations Sustainable Development Goals (SDGs), particularly SDG 11 (Sustainable Cities and Communities) and SDG 12 (Responsible Consumption and Production)[6]. By contributing to the development of explainable, adaptable, and ethically sound AI applications, this study not only advances academic knowledge but also addresses a globally relevant environmental challenge.

1.4 Purpose

The purpose of this research is to formulate a solution which tackles waste/litter generation in a way that works in real time, classifying waste using images of the litter scattered and utilizing the image's metadata (latitude, longitude, time of the picture taken etc.) for trend analysis. This allows us to identify litter generation patterns and forecast it, providing solid evidence for our solutions to generate suggestive policies based on history and future data to enhance waste management outcomes across diverse environments.

This research also focuses on:

- Solving a real-world problem by addressing inefficiencies in traditional waste management systems, like generalized static rules, manual processes for policy establishment and optimization, which affect the sustainability efforts for rural and data-scarce regions[5].
- Extending current knowledge by integrating isolated operations (classification, prediction, and decision-making) into a single, adaptable pipeline that employs a hybrid approach, wherein the output of each stage directly serves as the input for the subsequent stage.
- Adding value to literature by proposing a scalable and generalizable approach that utilizes technologies like YOLOv8s, LSTM, and Zephyr-7b-beta models to automate and optimize waste strategies.
- Addressing a critical gap by targeting accessibility and sustainability challenges in both urban and rural areas, where data may be scarce and the existing conventional/traditional or overly complex AI models fail to generalize.

1.5 Research Question

The primary research question that is guiding this research is: How can a multi-modal AI approach be leveraged to proactively automate waste management policy generation and promote sustainability in both urban and rural areas by integrating isolated operations through data-driven suggestions?
The growing environmental impact, lack of data-based forecasting, and real-time optimization solutions which create an important gap in achieving sustainable waste handling methods. Existing studies highlight the potential of using AI for isolated waste management solutions, but multi-modal, scalable solution for analyzing litter generation in real-time and providing suggestive strategies for data scarce region is still barely explored. This research focuses on addressing that gap by introducing an automated proactive AI-powered framework that contributes and closely aligns to sustainability and SDG 11 (Sustainable Cities and Communities).
This research will develop a framework which will include multiple components including classifying waste types, predicting waste generation trends using temporal models for any region, generating and validating suggestive waste management strategies and decision-making. It will validate the system on real-world image datasets with urban-rural representation, measuring accuracy, adaptability, impact on sustainability, and evaluate its performance under data-scarce scenarios to make sure that it can be generalized.
The machine learning models which effectively align with our use case are: YOLOv8s (You Only Look Once - small), LSTM (Long Short-Term Memory), and Zephyr-7b-beta (Transformer-Based model). Model evaluation will use metrics such as Precision, Recall for classification, RMSE/MAE for prediction, and policy effectiveness scores (Relevance, Specificity, Actionability, Completeness, Innovation) for validation.

1.6 Aims and Objectives

The aim of this research is to explore how emerging AI technologies can be harnessed to create innovative, sustainable, and adaptable solutions for improving waste management systems, with the broader goal of supporting environmental sustainability (SDGs) and enhancing the quality of life in both urban and rural communities.

The objectives of this research are as follows:

- Develop a unified AI-driven system that integrates real-time waste classification, regional waste forecasting, and policy optimization to support proactive waste management.
- Enhance waste management strategies by enabling data-driven, dynamic decision-making that is scalable and adaptable across both urban and rural regions, including data-scarce environments.
- Analyze and evaluate the performance, feasibility, and sustainability of the proposed AI framework in terms of classification accuracy, forecasting reliability, policy impact, ethical compliance, and alignment with UN Sustainable Development Goals (SDGs).

2. Related Work

TACO: Trash Annotations in Context for Litter Detection: This research introduced the TACO dataset, a well-defined, open-access resource for waste detection with 1,500+ images annotated across 60 litter categories in diverse environments. While TACO advanced litter localization benchmarks, it lacks temporal metadata and geographical context which limits its utility for predictive analysis. Future work of this paper highlights the need to integrate spatio-temporal attributes to enable forecasting, a gap our research addresses by augmenting similar datasets (e.g., Kaggle's trash detection.v35) with timestamps and coordinates for LSTM-based trend analysis.[1]

Litter Detection with Deep Learning: A Comparative Study: This research validated YOLOv5 and Mask R-CNN for waste classification, demonstrating robust performance in a controlled environment suitable for it. However, their study focused solely on detection, without exploring downstream applications like policy generation and failing to address data-scarce rural deployments. Our framework extends this work by integrating YOLOv8s (chosen for speed-accuracy balance) with forecasting and suggestive policy components, ensuring adaptability across diverse environments, including low-resource regions.[4]

pLitterStreet: Street-Level Plastic Litter Detection and Mapping: This research mapped plastic waste hotspots using geotagged imagery which enabled them for targeted cleanups. Despite its innovation, the approach is restricted to plastic waste and reactive interventions approach, which lacks predictive capabilities or policy suggestions for waste management. Our system generalizes this concept by forecasting multi-category waste trends (via LSTM) and converting predictions into proactive policies, thus shifting from spatial mapping to anticipatory governance.[8]

Solid Waste Generation & Disposal Using ML/DL: It surveyed LSTM's efficiency in waste forecasting, noting its superiority over statistical models in capturing temporal patterns. Even then, their review highlighted a critical gap showing isolated forecasting models rarely inform real-time decision-making. Our work bridges this gap by feeding LSTM outputs (quantile-binned forecasts) into Zephyr-7b-beta for policy generation, creating a closed-loop system that transforms predictions into actionable strategies[9]

Direct Preference Optimization (DPO): This paper proposed DPO to align LLMs with human preferences efficiently, therefore replacing complex RLHF. While DPO improves response coherence, it does not support structured, constraint-based outputs for environmental policy. Zephyr-7b-beta (DPO-tuned) adapts to generate policies which are grounded in waste forecasts and metadata, augmented by rule-based filters for evaluation to ensure operational feasibility and addressing DPO's tendency toward generic suggestions.[11]

Dublin City Council Litter Management Plan: This policy document (2020–2022) outlines static interventions (e.g., bin placements) but relies on manual audits and lacks automate & dynamic adaptation to real-time waste trends. Our AI framework automates such strategies by generating context-aware policies (e.g., "increase collection frequency in parks during summer") scored for relevance, innovation, and actionability which transforms rigid plans into adaptive, data-driven governance tools aligned with SDG 11.[16]

These studies collectively highlight the three limitations that our research resolves which are as follows: (1) Siloed operations (detection/forecasting/policy), (2) Static or reactive interventions, and (3) Urban bias neglecting data-scarce regions. By combining and integrating YOLOv8s, LSTM, and DPO-tuned LLMs into a single pipeline that is validated for scalability, policy quality, and SDG alignment, our framework adapts proactive waste management that converts environmental data into executable sustainability strategies.

3. Research Methodology

This section covers the research methodology that follows a systematic approach which is followed for developing, training, and evaluating and integrated AI framework, for the automation of effective and sustainable future suggestive waste policy generation. It outlines the approach of data acquisition, preprocessing, model architecture, the process of training, and the metrics used for performance evaluation that ensure the framework's accuracy, adaptability, and relevance across both urban and rural waste management scenarios.

3.1 Data Collection and Understanding

The image dataset used in this research was sourced from Kaggle (https://www.kaggle.com/datasets/ahnaftahmeed/trash-detection-image-dataset/data?select=trash-detection.v35.yolov9), which provides a YOLO-formatted version like TACO (Trash Annotations in Context) dataset. This dataset is specifically designed for object detection operation/classification focused on environmental waste and includes 4000+ high-resolution images of distinct types of litter captured in diverse real-world contexts divided into 3 folders, train, valid and test. Every image contains its respective YOLO-standard annotation file containing normalized bounding box coordinates and class labels, which helps us with precise localization and categorization of waste items and the bounding box annotations are normalized according to the image size, ranging from 0 to 1. The wide range of litter types (29 waste types from common waste) and environmental settings in the dataset ensure that models trained on it can generalize well across urban and rural scenes[1]. This makes it particularly suitable for our study, which aims to build a robust AI framework for real-time waste classification and sustainable policy suggestion.

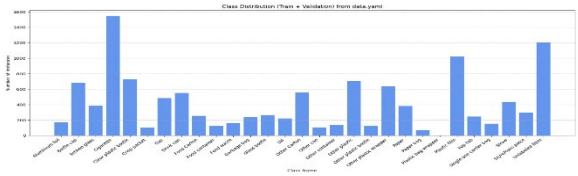

Fig. 1 - Waste Image Dataset Class Frequencies

The class frequency graph in the image shows that the dataset exhibits a noticeable class imbalance, with certain classes like Class 3 and Class 28 having over 1,000 instances, while others such as Class 22 and Class 21 have fewer than 100. This skew in class distribution calls for the need of augmentation or balancing techniques to ensure effective model training.

	image_id	date	day	time	season	lat	long	place_type	waste_types
0	000064_JPG.rf.7b1ed5445bd3072be6594ab6811fe21b...	2022-10-12	Wednesday	12:00	Autumn	53.338353	-6.258680	Park	['Paper']
1	000017_jpg.rf.000bf05f31c5cce9248fb0accd35ab21...	2022-10-12	Wednesday	14:00	Autumn	53.344062	-6.254571	College	['Unlabeled litter']
2	000035_jpg.rf.f98c8a9523c7fc068f061873550f1670...	2022-10-12	Wednesday	16:15	Autumn	53.348341	-6.246732	Hotel	['Plastic film']
3	000017_jpg.rf.0e95ec1c997eaec736d4e5e858f64e79...	2022-10-13	Thursday	13:00	Autumn	53.338353	-6.258680	Park	[]
4	000098_JPG.rf.3be4ef641dcd7f512682c0f80fbecf19...	2022-10-13	Thursday	10:30	Autumn	53.344062	-6.254571	College	['Food container']

Fig. 2 - Spatio-Temporal Dataset for LSTM

For the waste forecasting (LSTM) phase, a spatio-temporal dataset was curated by combining YOLO-based waste classifications after the trained model predicts the waste/litter type (image_id, waste_types) on randomly selected set of 3000 images, and integrated the result with corresponding metadata (date, time, lat, long) along with some known columns (day, seasons, place_type). This dataset represents daily waste patterns across three distinct locations Hotel, Park, and College, where each image corresponds to one day. This allows for seamless integration of the Classification and the forecasting of waste data using a singular dataset.

3.2 Data Preparation and Preprocessing

Some of the data preprocessing steps performed to make sure that the waste image dataset used is fair, diverse, and not overfitted, included:

1. **Image Resizing:** This step is done to make sure all input images across training, validation, and test split folders are resized to a uniform resolution of 640×640 pixels using OpenCV. This helps the dataset to align with the input requirements of YOLOv8s models and standardize the input dimensions, which in turn helps us in efficient batch processing and consistent learning across samples.

2. Tiered Oversampling for Class Balancing: Due to the class imbalance observed during EDA, a tiered oversampling strategy was applied:

- Classes with fewer than 100 instances were oversampled up to 600.
- Classes with 100–299 instances were increased to 500.
- Classes with 300–499 instances were raised to 400.
- Well-represented classes (>500) were left unchanged.

Oversampling was done by duplicating image-label pairs with new filenames to balance minority classes and reduce model bias.

To prepare the curated spatio-temporal dataset for sequential modeling using LSTM, the following preprocessing steps were applied:

1. **Time Indexing:** In this the raw date and time fields have been converted into a single datetime column, set it as the DataFrame index, and sorted chronologically. The series was then resampled to a daily frequency, to make sure a consistent day-by-day timeline for downstream modeling and sequence generation and filtered out days with zero total waste to reduce noise.
2. **Feature Engineering**: To add contextual signals, one-hot encoded was applied to the metadata fields place_type and season. Detected waste_types were transformed with a MultiLabelBinarizer, then aggregated by day (resample + sum) to obtain per-day counts for each waste category. Finally, metadata one-hot is concatenated with the per-type daily count columns into a single, aligned daily feature table that the model consumes.

3. **Sliding Window Creation:** The daily table is converted into supervised samples with the use of rolling window: with a lookback of 7 days, each input sequence comprises the previous seven daily feature vectors, and the target is the next day's vector (per-type counts). This preserves temporal order and allows the LSTM to learn dependencies across the past week to predict the following day.

3.3 Real-Time Waste Classification (YOLOv8s)

The first step of the operation involves classification of the waste type from the image dataset that utilizes YOLOv8s[7] which is a fast, single-stage object detection, pretrained model with a CSPDarknet backbone, PANet neck, and a unified detection head that helps in predicting bounding boxes and class probabilities in a single forward pass. The model takes in annotated images using labels for each image normalized in the range of 0-1 and outputs bounding boxes with class predictions for various types of waste.

After the preprocessing is completed, the model training is performed as follows:

Training and Validation: The model is trained for 100 epochs using 640x640 sized images, SDG optimizer (Stochastic gradient descent) to iteratively update the model's weights for minimizing loss function to balance learning speed and convergence ability, and cosine learning rate scheduling for decreasing the learning rate following the cosine curve. It uses classification, objectness, and bounding box regression loss during training and its validation is done using a specified image directory in data.yaml to monitor generalization and overfitting.

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \mathcal{L}_{obj} + \mathcal{L}_{box}$$

Where:

- $\mathcal{L}_{cls}$: Classification loss
- $\mathcal{L}_{obj}$: Objectness confidence loss
- $\mathcal{L}_{box}$: Bounding box regression loss

Testing and Inference: After the training, a dedicated testing phase is performed which loads the trained model from Google drive (where the best model was stored after training) and runs inference on the images present in the test/image directory.
The reason YOLOv8s has been chosen is because of its ability to balance performance and computational efficiency. Although, it can sometimes struggle detecting very small or overlapping items in visually cluttered environments, it offers high-speed inference which is extremely important for performance in real-time scenarios[8] allowing it to be more diverse with the operation of classification.

3.4 Regional Waste Generation Forecasting (LSTM)

The next step involves using the curated spatio-temporal dataset for forecasting daily waste generation patterns and understanding trends for different regions (Hotel, Park, and College) using a framework that incorporates LSTM (Long Short-Term Memory) model. Its networks are a specialized form of RNNs capable of learning not only long-term trends/dependences but also analyzing short term localized patterns, which makes it effective for modeling temporal trends in time series data[9].
The historical daily waste counts from the curated spatio-temporal dataset are used as sequential input features, along with other relevant attributes that is engineered, such as one-hot encoded location tags and day-of-week indicators which help the model predict the next day's waste count for a given region. LSTM cell operations are governed by the following equations:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * \tanh(C_t)$$

Where:

- f_t: Forget gate
- i_t: Input gate
- C_t: Cell state
- o_t: Output gate
- h_t: Hidden state (prediction vector)

The loss function minimized during training is:

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Through this, the model captures seasonality, fluctuations, and periodic behaviors in waste/litter generation of a region, offering more accurate and location-specific predictions more effectively than simple statistical methods[10].

After preprocessing, a sequential LSTM model was built using TensorFlow/Keras which contains a single LSTM layer with 64 units, followed by a dense output layer with linear activation for regression. The model is trained using the Mean Squared Error (MSE) loss function and the Adam optimizer which adaptively adjusts the learning rate for each parameter by combining the benefits of momentum and RMSProp improving handling of sparse or noisy gradients, with early stopping enabled to prevent overfitting. The input sequences are generated using a sliding window of the previous 7 days to predict the waste quantity of the next day. The training and validation datasets are split region-wise and is then evaluated using metrics like MSE and RMSE, plotted against actual vs. predicted values to assess temporal accuracy for each class of different waste types.

LSTM's ability to learn/understand sequential patterns using historical data, enables us to not only understand the waste generation trends but also enables us to forecast the types of waste that will be generated for suggestive policy generation.

3.5 Dynamic Policy Generation (Transformer-Based Language Model - zephyr-7b-beta)

In the final step, once the forecasted waste generated data for a week is obtained, Transformer-based Large Language Model (LLM), HuggingFaceH4/zephyr-7b-beta is used to generate dynamic and context-aware suggestive waste management strategies/policies. It is aligned fine-tune of Mistral-7B using Direct Preference Optimization (DPO), this 7-billion parameter model is designed to generate coherent, instruction-aligned responses based on complex, structured input prompts[11][12].
The model takes structured data derived from the result of LSTM forecasting phase and additional contextual parameters, including:

- Forecasted waste volumes by region and category.
- Regional metadata such as urban/rural designation, population density, or location-specific trends.
- Static constraints or policy rules (e.g., resource availability, compliance requirements).

Based on the input, the model generates 10 policy suggestions for each place such as increasing bin collection frequency in parks, or launching awareness campaigns in hotels etc. These generated suggestions help domain experts and municipal planners to interpret and understand more effectively, which can be further implemented or incorporated into their waste management decision making processes for that region.
The model's self-attention mechanism allows it to learn dependencies across different types of inputs and generate context-sensitive strategies which makes it a much better option than conventional rules-based systems in terms of adaptability and language fluency. It depends on the input prompt structure to provide responses, the lack of which will produce generic responses which can be resolved by introducing a lightweight rule-based post-filtering layer is implemented to refine the output for domain validity, ethical considerations, and operational feasibility. By utilizing Zephyr's generative capabilities, the research offers a scalable solution for policy generation that is not only explainable, but it makes the solution adaptable across different waste management scenarios, which in turn helps to close the gap between predictive analytics and actionable decision-making.

3.6 Evaluation

To understand and gain insights from the final evaluation of the waste classification performance, the YOLOv8s model was assessed using industry-standard object detection metrics that help in capturing both localization accuracy and class prediction reliability. These metrics were computed using the val() method from the Ultralytics YOLO framework after training completion.
The evaluation metrics included:

- **Mean Precision (mP):** It indicates how precise the model is in detecting waste objects which measure the proportion of correct positive amongst all positive predictions.
- **Mean Recall (mR):** It reflects the model's ability to detect all instances of waste that occurred which helps in reducing false negatives and measuring the proportion of correct positives identified.
- **Mean Average Precision at IoU 0.5 (mAP@0.5):** This metric helps quantify how accurately the bounding boxes overlap with ground truth. It computes the average precision for each class at an Intersection over Union (IoU) threshold of 0.5.
- **Mean Average Precision from IoU 0.5 to 0.95 (mAP@0.5:0.95):** This is a comprehensive evaluation that averages the AP over multiple IoU thresholds (from 0.5 to 0.95 in steps of 0.05), providing a more holistic view of model performance.
- **Per-Class AP@0.5**: It shows AP@0.5 values for each waste type class, helping identify which specific waste types are well-classified and which need further improvement.

For evaluating the LSTM-based forecasting model, regression based metrices were used to calculate the forecasting accuracy for each waste type. After generating the prediction, the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were calculated for each of the 29 waste type classes, which enables us to understand if the model's predicted daily waste volumes match the original values and provide information about the reliability of the prediction across different waste categories. Lower RMSE and MAE values indicate higher forecasting precision, which is crucial for suggestive policy generation in waste management.

To evaluate the quality of policies generated by the transformer-based large language model (HuggingFaceH4/zephyr-7b-beta), a custom scoring framework was implemented using Sentence-BERT (SBERT) embeddings and semantic similarity techniques. Five key attributes that were used to score the suggestive policies that were generated are as follows:

Metric	Purpose In Evaluation
Relevance	Evaluates how well the generated policy aligns with the contextual background of the waste scenario (e.g., season, location type, dominant waste items).
Specificity	Checks whether the policy addresses detailed spatio-temporal and situational factors such as day of the week, place type, and seasonal patterns.
Actionability	Assesses if the policy includes clear, practical steps such as assigned roles, required tools, and follow-up actions.
Completeness	Determines whether the policy covers both frequent and rare waste types for the given context.
Innovation	Measures how novel the policy is compared to baseline policies, encouraging creativity and unique solutions.

Each of these following dimensions in the above table is scored between 0 and 1, and the final policy score is computed as the mean of all five to ensure a balanced evaluation across relevance, depth, practicality, coverage, and originality. These scores are generated for all the policies for different places and only the best 3 overall policies from each place are considered.

4. Design Specification

This section outlines the core architectural framework specifications which comprises of various components in a hybrid AI-driven approach for waste management, details about the data processing methodologies which operate on the realistic and readily available dataset, and integration of automated, context -aware policy recommendation system. This research aims to utilize existing infrastructure like surveillance cameras to gather data and combining it with other isolated waste management operations. This approach not only addresses the gap of producing a localized, actionable waste management system for generating strategies specific to both rural/urban or data scarce regions but also enforces real-time waste monitoring capabilities. By generating policy suggestions, it shows a significant potential to support scalable, adaptive, and sustainable waste management solutions which address the urgent global challenges.

This section explores the idea and the reason behind the proposed framework, highlighting how its architectural design, methodological choices, and integration of AI techniques are not only just technically feasible but also strategically aligned with addressing pressing environmental and societal needs. The potential to close the gap between the theoretical advancements and their practical implementations serves as the justification for this research which ensures that system can produce actionable insights accessible data, which in turn contributes to the data-driven environmental policy making and the pursuit of sustainable development goals.

4.1 Data Specification

4.1.1 Significance of using image dataset for Waste Classification

The image dataset used for this research is a YOLO-formatted dataset sourced from Kaggle which contains over 4k high-resolution images of diverse litter types across different real-world environments. Each image is annotated using YOLO-standard annotation files which contain class labels and coordinates across 29 distinct waste types, enabling accurate object detection.

The dataset is quite suitable for our use case since it offers diversity in waste types, background and environmental conditions which helps the model generalize effectively without being constrained. This generalization ability helps the hybrid approach as it can be deployed in any place while still providing informed policy suggestions for sustainable waste management. From a technical perspective, the annotated data ensures efficient training for object detection frameworks like YOLO, while its broad range of waste scenarios enhances the model's adaptability to real-world variations.

4.1.2 Integration of Image Metadata for Spatio-Temporal Waste Forecasting

Before the waste forecasting (LSTM) phase, spatio-temporal dataset was curated by integrating YOLO-based waste classification results with the corresponding image metadata which includes latitude, longitude, date and time, where each image represents a single day mapped to 3 different locations i.e. Hotel, Park, College, capturing localized waste generation patterns.

This approach is beneficial since it utilizes the integration between real-time waste classification and forecasting which creates a dataset that reflects daily trends/variations of waste generation without relying on existing structured time series dataset. By leveraging the use of metadata out solution, we can dynamically generate time series data for forecasting based on captured images making it adaptive and deployable in any region. Unlike the existing system which depends greatly on historical waste generation data, our framework removes any such dependency, allowing us to reconstruct temporal data using real-time images and their metadata[13] making the forecasting accessible and deployable in any region.

While annotated waste images with metadata are not readily available publicly, the solution shows that it can still operate without any hinderance to its performance or adaptability. Such integration of visual classification and spatio-temporal metadata offers a robust and scalable solution, which addresses the pressing need for flexible, region-independent waste management systems in today's data-driven environmental strategies.

4.1.3 Scalability Across Data Scarcity Scenarios

A key advantage of this solution is that it can adapt to regions with minimal or no waste management records. Since it does not rely on external historical data, deployment of this approach can be accessible as it can capture images from existing public cameras, followed by automated metadata tagging, and scale to full waste forecasting and policy generation. This enables the framework to adopt AI-driven waste management making it scalable and relevant and cost effective across diverse socio-economic contexts.

4.2 Model Specification

This section helps in understanding and explaining the different models chosen, it efficiency, scalability, impact for forming a unified solution exploring the AI framework developed in this research which integrates three core models which are YOLOv8s for waste classification, LSTM for waste forecasting, and HuggingFaceH4/zephyr-7b-beta for policy generation. Each model contributes to a different stage of the automated waste management pipeline where the selection of these models was established by their technical strengths, alignment with our dataset structure, and their ability to work together to produce end-to-end, actionable waste management insights.

4.2.1 YOLOv8s for Waste Classification

YOLOv8s, the small variant of Ultralytics was chosen as it has exceptional balance of speed, accuracy and efficiency which makes it suitable for real-time waste classification in various contexts. It is a pretrained model which uses learned weights from large-scale object detection tasks, allowing it to generalize effectively with comparatively fewer training epochs[8]. This was further amplified using waste image dataset, which is natively annotated in YOLO format with normalized bounding box coordinates (0–1 range) and class labels which significantly reduces preprocessing overhead and ensures our model can generalize across varied scenes without being restricted to any single geographic region.

Classification with training ResNet18 from scratch using labeled categories of non-annotated waste images were used, which were combined into composite images through multiple layout strategies like 2x2 grid, horizontal splits, vertical splits, random paste, and three-layout compositions, to simulate complex real-world waste arrangements. The area coverage of each waste category in the combined synthetic images was calculated and used as label percentages, incorporating slight jittering for realism.

Even after all these steps, the model struggled to learn accurate multi-class proportions, which highlights the role of bounding box annotations of waste detection. It also validated the decision to use YOLOv8s with the annotated dataset, as it shows that pretrained detection models with high-quality annotated data are significantly more effective for this task than training from scratch on synthetic composites which simulate real world waste[7].

4.2.2 LSTM: Regional Waste Generation Forecasting

For capturing trends and forecasting daily waste generated across various regions, this research is utilizing LSTM because of the following reasons:

- **Capturing Temporal Dependencies:** LSTM's memory mechanism learns short- and long-term patterns (daily, weekly, seasonal), enabling accurate waste trend forecasting[9] in varied environments.
- **Resilience to Noisy Inputs:** Handles irregular data such as missing days, seasonal spikes, or deviations, maintaining robust predictions despite fluctuations.
- **Multivariate & Multi-Region Forecasting:** Processes waste counts alongside contextual features (location labels, day-of-week) to capture how multiple factors influence waste patterns.
- **Established Research Precedent:** Widely used in environmental forecasting (e.g., air pollution, energy, water), aligning with waste generation's nonlinear temporal behavior.
- **Low Data Requirement & Robustness:** Needs less training data than Transformers, making it suitable for urban and rural areas with limited annotated datasets.

Studies of waste generation forecasting have shown LSTM having better accuracy & adaptability compared to linear or statistical modes achieving lower Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) across diverse time scales and contexts[10]. It enables us to learn about the temporal relationships in waste generation which make it practical and reliable for predicting waste trends, and its ability to forecast across different regions supports proactive planning enabling responsive policy intervention, which is critical for efficient, data-driven waste management.

4.2.3 Transformer LLM for Policy Generation - HuggingFaceH4/zephyr-7b-beta

HuggingFaceH4/Zephyr-7b-beta is used as it is a pretrained 7B transformer model which is fine tuned Direct Preference Optimization (DPO) on instruction data, and it turns structured input from LSTM into actionable suggestive policy without needing any supervised dataset. Being DPO-aligned, it is more stable and efficient than RLHF (Reinforcement Learning from Human Feedback), which helps when we evaluate in terms of quantities (e.g., predicted volumes by type) and constraints (e.g., resources, region type). LSTM output is converted into prompt schema which is used by Zephyr to generate targeted actions which align with emerging evidence that LLMs are effective decision-support tools in environmental and regulatory contexts, where they augment expert workflows with context-aware synthesis and rationale generation.

Initially cGAN, Conditional Generative Adversarial Network where the data generation is guided by a specific condition or input combined with LLM and reinforced by RL was explored in generated suggestive waste generation policies. It would generate JSON output which would be converted to human-like fluent policies and then scored for its effectiveness using RL. However, cGANs for controllable text generation are notoriously unstable and depend on labeled/conditioned training data (good/bad policy pairs or rich conditioning) and training it from the scratch will introduce dependency on an external dataset, which will undermine the accessibility and reproducibility. By using Zephyr, the unified pipeline is still maintained while delivering practical, explainable policies from the same inputs that drive forecasting and will work in areas where the data is scarce.

4.3 Framework Specification

This section explores the core framework and the reason behind choosing this structure for implementation which serves as the core of the single-pipeline, multimodal system that turns raw images into actionable, location-aware waste policies by chaining three interoperable components: Real-time classification, Spatio-Temporal forecasting and Suggestive policy generation, which is effectively designed around one accessible waste image dataset and their metadata while preserving scientific rigor and reproducibility.

This research's solution pipeline is a one-stop solution for generating suggestive waste generation policies since it does not rely on multiple external datasets yet produces explainable text that is scored and filtered thus keeping the approach transparent, data-minimal, and controllable. By tying the quality of detection directly to forecast fidelity and then to policy utility, the framework creates a chain from perception to decision which can be measured: Better bounding boxes will lead to cleaner daily counts which will produce more reliable forecasts and will generate higher-quality policy drafts. Because each stage emits auditable artifacts (mAP curves, RMSE/MAE, policy scores on relevance, specificity, actionability, completeness, innovation), it can quantify how improvements at one layer propagate to societal outcomes at the end of the chain. This end-to-end visibility is rarely available in siloed studies and is crucial for evidence-based environmental policy[5].

Since the model is trained on diverse YOLO-formatted datasets, it can be ported to new cities or rural areas without any concerns for the historical data or tables. The reliance on one data modality (images) plus standard metadata reduces onboarding costs[13] for the solution making it sustainable, while the LLM's text output fits naturally into municipal workflows (work orders, public notices, staffing briefs) along with the scoring layer which supports human-in-the-loop governance so that the authorities can accept, edit, or reject policies with justifications. Moreover, many municipal actions follow certain overflow events or complaints, whereas this hybrid system, forecasts trigger policy drafts before any such kind of peaks occurs, which changes operations from reactive fixes to anticipatory interventions, and since the generate suggestive policies are automatically generated and ranked using five scoring dimensions, it functions as a continuous policy recommender that can scale across districts while preserving oversight and local nuance.

Overall, this framework covers a lot of gaps for defining a proper waste management system with aim of not just making operations better, but to provide a niche open-ended complex solution which can leverage AI in different ways for not just waste management but to other domains like Agricultural Resource Optimization, Traffic & Mobility Management etc. or allowing other external factors like climate, lifestyle of the people etc. to be implemented as parameters to making the policy generation even more effective.

5. Implementation

This section talks about the final implementation of the hybrid automated suggestive waste policy generation methodology as it focuses on what is developed, integrated, and tested based on the design structure discussed earlier in section 4. The implementation includes dataset preparations, classification, forecasting and policy generation & scoring, while paying attention to the accessibility, scalability, ethical considerations, dataset balancing, and technical optimization, ensuring that the system can be implemented and adapted to any region which is why it was conducted within a cloud-based environment. The outputs generated at this stage form the functional backbone of the research, demonstrating the feasibility and applicability of the proposed approach in real-world waste management contexts.

5.1 Overview of the System Implementation

The final stage of the proposed multi-modal waste management framework integrates the trained models, processed datasets, and evaluation pipeline into a unified system capable of generating proactive and context-aware waste management policies which begins with importing the waste classification dataset from Google Drive into the Google Colab environment, ensuring proper directory structure and accessibility for subsequent stages. At this stage, the implementation outputs include:

- **Transformed and Annotated Datasets**: Waste images were cleaned, structured, and annotated using YOLO-compatible formats, which incorporate both visual data and contextual metadata (e.g., location type, season, time). Class frequency distributions and dataset statistics were generated for training insights.
- **Object Detection Model**: A YOLO-based waste classification model was trained to detect 29 distinct waste categories from real-world images that produce bounding box coordinates, class predictions, and confidence scores, enabling accurate, real-time waste identification.
- **Forecasting Model**: A Long Short-Term Memory (LSTM) network was trained using spatio-temporal waste data to predict future waste trends for various location types which captures daily, weekly, and seasonal variations in waste generation.
- **Policy Generation Module**: A Zephyr-based large language model (LLM) text generation module, supported by contextual embeddings, produces waste management policy suggestions tailored to the detected waste composition and predicted trends.
- **Evaluation Metrics**: Semantic similarity-based metrics (Relevance, Specificity, Actionability, Completeness, and Innovation) were implemented using SBERT embeddings to score the generated policies against contextual and baseline references.
- **Final Output**: For any given image input and metadata, the system produces:
  - A set of detected waste items with locations and confidence levels.
  - Forecasted waste trends for the specific location type.
  - An AI-generated, context-specific waste management policy suggestion, evaluated and scored for quality and applicability.

This integrated pipeline ensures the system meets its objectives of being multi-modal, proactive, automated, and suggestive, while remaining adaptable for deployment in both urban and rural environments[1].

5.2 Setup of Dataset and Ethical Considerations

This research utilizes litter imagery dataset, which is publicly available, which makes sure that it is compliant with open-access licensing and proper citation of source. Ethically the project ensures that no personal or sensitive identifiable information was present in the dataset, thus avoiding any risk of privacy. In line with responsible AI development principles, the modelling process was designed to maintain transparency, reproducibility, and accountability along with the responsible use of AI for research and all dataset transformations, balancing strategies, and hyperparameter configurations were documented within the Jupyter Notebook to facilitate peer review and reproducibility.

5.3 Tools and Technologies

The system was implemented using Python 3.x due to its vast ecosystem of libraries for data science, machine learning, and computer vision, making it the ideal choice for rapid prototyping and research-oriented workflows. The Google Colab environment provided a cloud-based Jupyter Notebook interface with pre-configured dependencies, which enables seamless collaboration, ease of code execution, and access to hardware accelerators without local resource constraints.

Libraries for Data Handling & Visualization

- **Pandas**: Used for structured data manipulation, particularly for reading metadata, processing tabular datasets, and aligning waste classification outputs with forecasting inputs.
- **Numpy**: Provided fast numerical computation capabilities, essential for array-based operations and pre-processing image and time-series data.
- **Matplotlib**: Facilitated the creation of visual representations, such as trend graphs, detection overlays, and forecasting result plots, to support analysis and interpretation.
- **Yaml**: Enabled structured configuration management, particularly for model parameters and YOLO dataset configuration files.
- **cv2 (OpenCV)**: Used for image processing tasks including resizing, cropping, augmentation, and format conversion to prepare input data for classification.
- **Glob**: Assisted in dynamically retrieving file paths from large datasets, streamlining batch processing of images.
- **os & shutil**: Provided file system operations, such as creating directories, moving files, and organizing datasets for training and testing.

Deep Learning Frameworks

- **ultralytics (YOLOv8)**: It is implemented as the primary object detection framework for waste classification, leveraging pre-trained weights to achieve high detection accuracy while minimizing training time. YOLOv8's architecture is optimized for both inference speed and accuracy, making it well-suited for real-time waste detection in diverse environments.
- **TensorFlow/Keras**: It is used for implementing the Long Short-Term Memory (LSTM) network, enabling the forecasting of waste generation trends based on spatio-temporal metadata. TensorFlow/Keras was selected for its flexible API and ability to handle sequential data efficiently.

Supporting Tools

- **Google Colab**: Provided GPU acceleration (Tesla T4 and L4) to reduce model training and inference times, critical for iterative experimentation. Its integrated environment eliminated local hardware limitations and supported reproducibility across research collaborators.
- **Google Drive**: Used as the primary dataset storage and retrieval mechanism, allowing seamless integration with Colab notebooks for reading, writing, and versioning datasets without requiring manual uploads for each session.

6. Evaluation

This section presents a rigorous evaluation, an end-to-end analysis of the system's outcomes which links detection quality, forecasting accuracy, and policy utility that demonstrates how the approach answers

the research question and objectives. Only the most relevant results for YOLOv8, class-wise and aggregate mAP, precision/recall, and confusion analyses, for the LSTM, per-category RMSE/MAE with error profiles over time, and for policy generation, a semantic scoring framework (relevance, specificity, actionability, completeness, innovation) applied to LLM outputs conditioned on forecasts is reported.
Also, wherever necessary, visual aids have been provided to assess the robustness through PR curves, confusion matrices, error distributions, and policy score charts etc. which all together provide a practical evaluation of the solution.

Evaluation of YOLO Model Performance on Validation Data

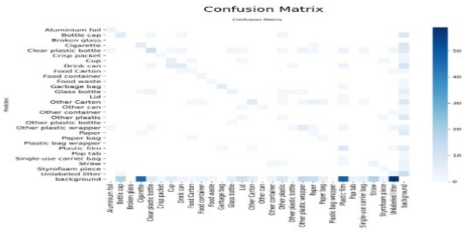

Fig. 3 - Confusion Matrix for YOLOv8s

From Fig. 3 we can see that the model correctly identifies certain classes with sparse strong diagonals, with frequent significant class confusion between visually similar waste types and misclassification into background suggesting the need for balanced dataset and clearer inter-class feature separation.

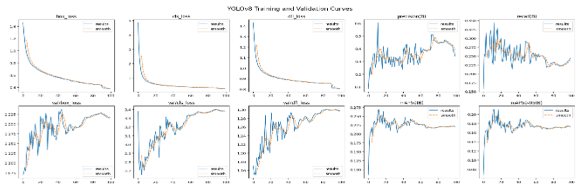
Fig. 4 - Training and validation curves of YOLOv8s

The above graphs in Fig. 4 shows that the steadily decreasing training losses alongside high, fluctuating validation losses suggest overfitting, while modest precision (0.5–0.6), low recall (<0.3), and low mAP (~0.25/0.18) indicate limited detection accuracy and poor localization across IoU thresholds.

	x	y	x2	y2	confidence	class	file	i
0	120.282677	424.086365	196.695038	456.329651	0.927362	Drink can	000000_jpg.rf.e3fe7e2f0223906d4a754b6650db1cc1...	0.0
1	244.527222	53.697983	276.535217	81.637512	0.373458	Other plastic wrapper	000000_jpg.rf.e3fe7e2f0223906d4a754b6650db1cc1...	0.0
2	542.956604	158.048492	640.000000	461.089935	0.759245	Other container	000001_JPG.rf.74374d550f5b0299284313ff86688c4a...	1.0
3	414.880890	457.555359	431.260773	481.614807	0.527324	Bottle cap	000001_JPG.rf.74374d550f5b0299284313ff86688c4a...	1.0
4	271.452271	221.220490	343.145081	243.144440	0.258690	Straw	000001_JPG.rf.74374d550f5b0299284313ff86688c4a...	1.0
...	...	...	...	...	...	...	...	...
269	53.253113	143.982910	548.764526	584.692993	0.517329	Food waste	food-waste14_jpg.rf.e80c4af8a2954441811f48afae...	154.0
270	248.372681	119.332825	401.118408	502.218933	0.527856	Food waste	food-waste18_jpg.rf.734167bd36bcb8c96ed42d2313...	155.0
271	259.065887	86.175713	370.380890	340.030701	0.902697	Food waste	food-waste4_jpg.rf.10c1f28c3ad15c42ad83eedea29...	156.0
272	487.878815	292.548889	602.471802	495.678894	0.632116	Food waste	food-waste4_jpg.rf.10c1f28c3ad15c42ad83eedea29...	156.0
273	257.625366	381.177002	576.693604	464.552612	0.975364	Other Carton	pizza_box9_jpg.rf.2962c1757e5a65801aed8a3190be...	157.0

274 rows × 8 columns

Fig. 5 - PBOX for Test Data prediction

The above Fig. 5 PBOX predictions show inconsistent confidence (0.25–0.97), with strong detection for well-represented classes like ‘Other Carton’ but multiple mid-confidence false positives, highlighting the need for threshold tuning and NMS refinement.

The above results prove that, while turning the models can help in making the results better, the underlying challenge arises from the overall highly imbalanced nature of the dataset, which restricts the model’s ability to generalize even with techniques like oversampling that carries the risk of introducing biasness because of duplicating existing patterns. Also, if the minority classes have very limited diversity, oversampling will learn from those patterns which fail in unseen variations as YOLOv8s heavily relies on diversity and not just quantity, affecting its ability to generalize[14].

6.1 Experiment 1:

For this Experiment, the evaluation of the YOLO-based waste classification models is done using model.val() for class-wise performance instead of solely relying on the mean evaluation summary. The reason behind this decision lies in the imbalanced nature of the dataset which skews the mean metrics, allowing dominant classes to disproportionately influence the overall scores, while under-represented categories remain under-predicted or ignored.

```
Evaluation Summary:
Mean Precision          : 0.299
Mean Recall             : 0.344
Mean AP@0.5             : 0.269
Mean AP@0.5:0.95        : 0.210
```

Fig. 6 - Evaluation Summary for YOLOv8s (All Classes)

The mean evaluation metrics in Fig. 6 reveal a mean precision of 0.299 and mean recall of 0.344, which indicates that the model retrieves relevant instances reasonably, there is a notable gap in accurately classifying them. The Mean Average Precision (mAP) at IoU=0.5 stands at 0.269, whereas the stricter mAP@[0.5:0.95] yields a lower score of 0.210, reflecting reduced performance under stricter localization and classification accuracy thresholds. Although these values provide an overview, they fail to represent the disparity in performance across individual waste categories.

Class	Images	Instances	Box(P	R	mAP50	mAP50-95):
all	158	462	0.299	0.344	0.269	0.21
Aluminium foil	4	5	0.586	0.6	0.667	0.566
Bottle cap	23	25	0.289	0.32	0.271	0.185
Cigarette	25	54	0.189	0.185	0.141	0.0639
Clear plastic bottle	21	25	0.293	0.6	0.389	0.343
Crisp packet	2	2	0.217	0.5	0.498	0.498
Cup	17	18	0.306	0.418	0.396	0.268
Drink can	11	12	0.193	0.583	0.307	0.249
Food Carton	5	5	0.0831	0.2	0.0874	0.0776
Food container	5	6	1	0.473	0.509	0.493
Food waste	3	3	0.623	0.579	0.564	0.214
Garbage bag	6	7	0.583	0.857	0.73	0.455
Glass bottle	9	12	0.274	0.473	0.195	0.149
Lid	14	16	0.349	0.188	0.117	0.107
Other Carton	10	14	0.227	0.643	0.341	0.258
Other can	1	1	0.26	1	0.332	0.332
Other container	3	4	0	0	0.0128	0.0102
Other plastic	15	17	0.0411	0.0588	0.0615	0.0599
Other plastic bottle	4	4	0	0	0.306	0.248
Other plastic wrapper	21	26	0.248	0.269	0.216	0.182
Paper	10	21	0.118	0.0952	0.0719	0.0601
Paper bag	4	7	0.644	0.286	0.247	0.207
Plastic film	41	69	0.407	0.199	0.199	0.145
Pop tab	2	3	0	0	0	0
Single-use carrier bag	7	8	0.282	0.125	0.0816	0.0393
Straw	13	22	0.281	0.182	0.117	0.0885
Styrofoam piece	5	6	0.419	0.333	0.364	0.362
Unlabeled litter	33	70	0.162	0.114	0.047	0.0228

Speed: 1.6ms preprocess, 6.1ms inference, 0.0ms loss, 5.0ms postprocess per image
Results saved to **runs/detect/val**

Fig. 7 - model.val() evaluation

The above Fig. 7 shows the class-wise breakdown shows large performance gaps, with some classes like Garbage Bag achieving high mAP@0.5 (0.73) and others like Other Container near zero (~0.012), driven by uneven data distribution and visual distinctiveness. Precision-recall patterns further reveal reliable detection for classes like Aluminum Foil (0.586/0.6) but poor results for small or noisy items like Cigarette (0.189/0.185).

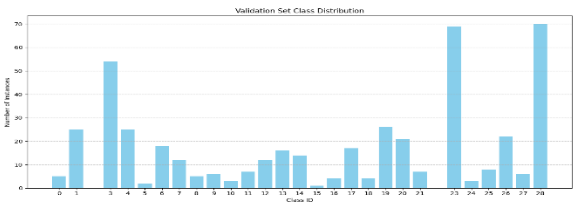

Fig. 8 - Validation class distribution

The validation class distribution in Fig. 8 reveals severe imbalance, where high-frequency classes dominate learning yet still vary in performance due to intra-class visual variability, while rare classes are overshadowed in detection accuracy.

This evaluation confirms that the aggregate mean metrics mask the challenges of imbalanced datasets. The model.val()-based class-wise analysis not only reveals strengths and weaknesses per category but also provides information about strategies such as data augmentation for minority classes, reweighting loss functions, and synthetic sample generation. From a practical perspective, these insights are vital for real-world deployment, where certain waste types can be rare but critical for training the data.

LSTM Forecasting - Experiment Evaluation (per-class and quantile-binned predictions)
The LSTM Forecasting model evaluation provides the following insights:

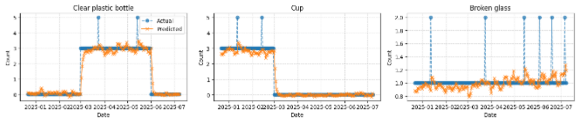

Fig. 9 - Actual vs Predicted for Top 3 Waste Types

The graphs in the above Fig. 9 show that the model effectively captures seasonal/periodic trends but struggles with underpredicting rare surges in Clear Plastic Bottles & Cup while being able to identify sharp drops to zero and periods of sustained high counts. For Broken Glass, the prediction closely follows actual counts, though it slightly under- or overestimate during sudden small peaks.

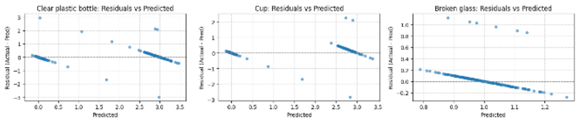

Fig. 10 - Residual vs Predicted Analysis

The Residual vs Predicted Graphs in Fig. 10 for the top 3 items shows that the Clear Plastic Bottle and Cup are modeled well for baseline trends but miss extreme fluctuations, while Broken Glass residuals are tightly clustered with slight overestimation at higher values.

Overall, these results show that the LSTM is effective at capturing consistent patterns that are in demand for frequently occurring waste types, but may require additional handling (e.g., event-based features or spike modeling) for rare, abrupt changes. This supports the use case of predictive waste management, where stable baseline forecasting is crucial, but extreme spikes in the deviations will be handled with complementary alert systems.

6.2 Experiment 2:

For this experiment of the LSTM model, we move away from the normal way of evaluating the model but focus on evaluating per class performance to gain more insights and leverage that for policy generation.

```
Evaluation per waste type:
Aluminium foil              | RMSE: 0.17 | MAE: 0.08
Bottle cap                  | RMSE: 0.14 | MAE: 0.07
Broken glass                | RMSE: 0.19 | MAE: 0.09
Cigarette                   | RMSE: 0.06 | MAE: 0.05
Clear plastic bottle        | RMSE: 0.45 | MAE: 0.20
Crisp packet                | RMSE: 0.39 | MAE: 0.13
Cup                         | RMSE: 0.36 | MAE: 0.15
Drink can                   | RMSE: 0.31 | MAE: 0.14
Food Carton                 | RMSE: 0.66 | MAE: 0.51
Food container              | RMSE: 0.10 | MAE: 0.04
Food waste                  | RMSE: 0.54 | MAE: 0.46
Garbage bag                 | RMSE: 0.63 | MAE: 0.55
Glass bottle                | RMSE: 0.29 | MAE: 0.15
Lid                         | RMSE: 0.56 | MAE: 0.46
Other carton                | RMSE: 0.55 | MAE: 0.46
Other can                   | RMSE: 0.42 | MAE: 0.30
Other container             | RMSE: 0.26 | MAE: 0.12
Other plastic               | RMSE: 0.28 | MAE: 0.12
Other plastic bottle        | RMSE: 0.25 | MAE: 0.14
Other plastic wrapper       | RMSE: 0.39 | MAE: 0.26
Paper                       | RMSE: 0.38 | MAE: 0.21
Paper bag                   | RMSE: 0.51 | MAE: 0.41
Plastic bag wrapper         | RMSE: 0.36 | MAE: 0.13
Plastic film                | RMSE: 0.23 | MAE: 0.16
Pop tab                     | RMSE: 0.57 | MAE: 0.43
Single-use carrier bag      | RMSE: 0.31 | MAE: 0.15
Straw                       | RMSE: 0.47 | MAE: 0.38
Styrofoam piece             | RMSE: 0.50 | MAE: 0.39
Unlabeled litter            | RMSE: 0.23 | MAE: 0.10
```

Fig. 11 - Evaluation of LSTM model per waste types

The above graph in Fig. 11 shows that the waste categories have different variability, when aggregated can hide important behavior, where some classes such as Food container and Cigarette exhibit low errors (e.g., RMSE ≈ 0.10/0.06; MAE ≈ 0.04/0.05), while others like Food Carton, Garbage bag, Pop tab, and Lid have markedly higher RMSE/MAE. Evaluating by class helps in identifying which waste types are forecastable with high confidence and which waste types need augmentation.

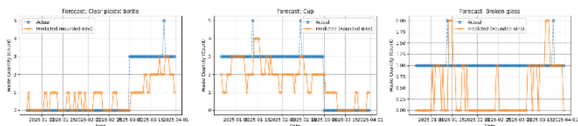

Fig. 12 - Forecasted Actual vs Predicted with Binning Code

The above graph in Fig. 12 indicates that for Clear Plastic Bottle and Cup, forecasts capture overall trends and shifts, though peaks are underestimated, which is still sufficient for adjusting collection frequency without overreacting to different anomalies. Broken Glass predictions preserve small but important rises despite timing lag, supporting targeted safety alerts for hazardous material.

The reason why we are using binning code instead of rounding or normalizing is because direct rounding of the continuous output will collapse signals which are small but meaningful & show variance to 0 eliminating all the low-volume classes. Also, normalized decimals (0-1) are also problematic for the downstream planning because they are not interpretable counts and are not comparable across classes with different scales. To avoid this calibration per-class quantile bins are done on validation predictions by building build class-specific bin edges from quantiles [0.0, 0.4, 0.7, 0.9, 0.98, 1.0] and mapping each continuous forecast to an ordinal level 0-4 via np.digitize. This avoids zeroing out rare events, normalize scale differences across categories for fair comparison and feed cleanly into policy rule through the LLM prompts.

The significance of this approach is that it builds a usable, per-class time series using the curated dataset by integrating results of image classification and metadata and calibrates forecasts to discrete levels that are comparable across categories and regions, which is helpful as planners need to allocate resources, schedule pickups, and trigger interventions.

Transformer policy evaluation - experiment & findings

The evaluation for the transformer model's suggestive policy generated framework is done by using five metrices which are : relevance, specificity, actionability, completeness, and innovation to assess and rank proposed solutions. These metrics were chosen to ensure that:

- Recommendations align with the problem context.
- Provides clear steps for execution.
- Feasible to implement.
- Cover the issue comprehensively.
- Bring fresh, creative ideas.

This structured multi-metric evaluation ensures the model balances creativity with feasibility, preventing the selection of ideas that might be vague and cannot be acted upon or too conventional to drive meaningful change.

6.3 Experiment 3:

This experiment shows why these scoring parameters work for the suggestive policies that are generated rather than relying on standard methods.

	place	policy_text	relevance	specificity	actionability	completeness	innovation	final_score
3	College	Limit single-use plastic: The park can phase o...	0.443805	0.344421	0.413007	0.460582	0.553662	0.443095
0	College	Implement a recycling program: The park manage...	0.492147	0.336457	0.406208	0.410053	0.473828	0.423739
2	College	Encourage reusable items: The park can provide...	0.389041	0.274918	0.361917	0.430047	0.623762	0.415937
13	Hotel	Limit single-use plastic: The park can phase o...	0.473786	0.373554	0.425239	0.412299	0.553662	0.447708
12	Hotel	Encourage reusable items: The park can provide...	0.443594	0.320332	0.406185	0.383452	0.623762	0.435465
10	Hotel	Implement a recycling program: The park manage...	0.535735	0.374346	0.424134	0.361949	0.473828	0.433999
20	Park	Implement a recycling program: The park manage...	0.732785	0.454409	0.568595	0.410535	0.473828	0.528031
23	Park	Limit single-use plastic: The park can phase o...	0.634275	0.441636	0.539255	0.441253	0.553662	0.522016
28	Park	Use eco-friendly products: The park can use ec...	0.691116	0.404383	0.520351	0.379473	0.541695	0.507404

Fig. 13 - Scoring parameters for generated policies

The above table in Fig. 13 shows that the Park policies rank highest (final ~ 0.528, 0.522, 0.507) with strong relevance (~0.73, 0.63, 0.69) and solid actionability (~0.57 - 0.52), linking forecasts (e.g., plastics) to actions like recycling programs and reducing single-use items. Hotel policies score moderately (final ~0.434 - 0.448) with some higher innovation (~0.62) but lower relevance/specificity (~0.47 - 0.37), while College policies are most novel (>0.55) yet less grounded (~0.39 - 0.49), indicating a need for tighter constraints.

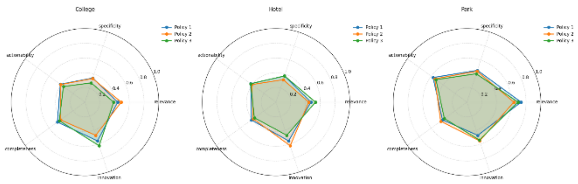

Fig. 14 - Radar Graphs for scoring parameters

The radar charts in the above Fig. 14 compare the five scores for the top three policies per place.

- Park shows the largest, well-rounded polygons, especially on relevance and actionability which indicates deployable, context-aware guidance.
- Hotels exhibit a trade-off between higher innovation on some policies but smaller specificity/completeness areas, imply that the policies need more local details.
- College polygons extend most on innovation, with comparatively smaller relevance and specificity.

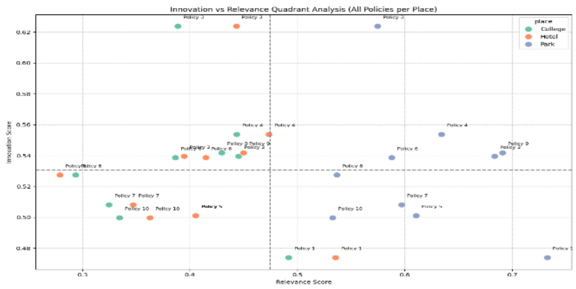

Fig. 15 - Innovation vs Relevance quadrant analysis for all policies

The above graph in Fig. 15 shows the quadrant plot which indicates that several park policies have high relevance and innovation which is essential for actionable strategies. College points lie more in the innovative region with less relevancy, while the Hotel clusters near the center with few innovative exceptions. This allows the planners and practitioners to pick policies that are both on-target and distinctive, and flags those needing refinement.
The reason why these scoring metrices work is because standard metrices (BLEU, ROUGE, perplexity) reward n-gram overlap or fluency, not decision usefulness. This approach not only ties up to our suggestive framework, the automated nature of it acts as recommendation providing not just policies but meaningful insights which can be utilized.

6.4 Discussion

Based on the evaluations for all the steps of the hybrid model of approach for this solution, the transformer-based approach to generate suggestive policies achieved promising results, aligning well with the research goal of developing a multi-dimensional, context-sensitive evaluation framework for sustainable policy recommendations. By incorporating relevance, specificity, actionability, completeness, and innovation metrices this study moves beyond simplistic single scoring assessment and offers more in-depth view of the quality of output which reflects real-world decision-making needs in a much better and more practical way. Its adaptability across various context helps in revealing scenario-specific performance patterns, enabling targeted analysis of where the model excels and where refinement is needed which compared to other LLM policy generation studies, allows for clearer identification of trade-offs[15] and provides actionable insight for iterative model improvement.

The study acknowledges that further refinements are possible such as integrating domain-specific constraints, combining automated scoring with expert ratings, and expanding underrepresented categories with validated data, these serve as the next steps that build on an already solid foundation. Data Imbalance in the image dataset was one of the key issues faced, which can be improved by applying more techniques (like adding class weights), that impacted the results of the classification and forecasting leading to generating decent prompts. Importantly, the current design demonstrates that with well-crafted prompts, structured evaluation metrics, and targeted data balancing, transformer models can produce not only creative but also contextually relevant and operationally feasible policy ideas.

In the broader context of prior work, this research makes a meaningful contribution by providing a scalable, adaptable, and transparent framework for assessing AI-generated sustainability policies. The findings confirm the potential of large language models as valuable partners in policy ideation when supported by robust evaluation strategies. Moreover, the approach allows for integrating constraint-aware generation and dynamic metric weighting in future iterations[15], which ensures that the outputs are increasingly aligned with both innovation goals and practical implementation needs.

7. Conclusion and Future Work

This study asked: How can a multi-modal AI approach be leveraged to proactively automate waste-management policy generation and promote sustainability across urban and rural regions by integrating currently isolated operations through data-driven suggestions? This question is answered by delivering a unified pipeline that:

- Detects waste in images with YOLOv8s.
- Convert detections plus metadata (lat, long, timestamp) into a spatio-temporal series for LSTM forecasting.
- Condition a transformer LLM (Zephyr-7B) to draft context-aware policies, with a rule-based filter and semantic scoring for selection.

This directly addresses the research's objectives to integrate real-time classification, regional forecasting, and policy optimization hybrid model which can be deployable in data-scarce environments and to evaluate performance with task-appropriate metrics (mAP, precision, recall for detection, RMSE & MAE for forecasting, and a five-dimension policy scoring).

The work shows:

- Feasibility of creating an end-to-end automation system from perception to prediction to prescription using one accessible dataset (images along with metadata), enabling location-agnostic deployment.
- Detection performance is constrained by class imbalance, with aggregate validation scores around mP = 0.299, mR = 0.344, mAP@0.5 = 0.269, and mAP@0.5:0.95 = 0.210, but on the other hand, per-class analysis reveals strong classes (e.g., Garbage Bag at AP@0.5 ~ 0.73) and weak classes where diversity is lacking.
- LSTM forecasting captures per-class trends and with the help of quantile-binning, it preserves all low-volume signals for operational use by avoiding rounding to 0 and losing less represented waste types which can provide valuable insights into the waste generation trends.
- Policy generation is practically ranked and audited via five semantic dimensions which include relevance, specificity, actionability, completeness, and innovation, with the park setting producing several high-relevance and high-innovation policies ready for implementation.

Together, these results combined, substantiate the system's novelty and practical value as it manufactures its own time series from images, forecasts next-day volumes per type for each location, and proposes ranked, explainable policies which shift the operations from being reactive to proactive.

7.1 Limitations

The primary constraints that affect the results arise from the data representation, not the architecture. The annotated waste image dataset used for this research was highly imbalanced and limited in visual diversity for several classes of waste. Even though oversampling helped by increasing the quantity of the underrepresented class, it cannot inject the missing variation, which suppresses per-class generalization and contributes to misclassifications. Forecasting results recognized most of the overall patterns of waste generation but show under-prediction on rare spikes for some categories[3] while baseline trends are captured, sudden deviations remain challenging. For generating policy, the LLM can be sometimes overly generic without structure, hence the current approach caters to the problems by providing structured prompt schemas, a light rule-filter, and post-hoc semantic scoring which can be further improved using tighter, domain-aware constraints during generation for immediate implementation. These are framed as avenues for strengthening an already functional pipeline, rather than gaps in the core design.

7.2 Future Work

Although this research shows a full proof end to end unified pipeline for automated suggestive waste policy generation, it still has some aspects which can be explored and worked upon to improve its efficiency, making it scalable and adaptable. Some of those future works are as follows:

Prioritizing diversity-oriented enrichment of minority classes (new scenes, lighting, viewpoints), targeted augmentation, re-weighting losses, and validated synthetic samples which are all suggested by the class-wise findings, to improve difficult categories beyond what duplication through oversampling can offer.

Adding other external datasets like climate, lifestyle of the people etc. that is supported by the current framework structure, which can be used as parameters for enabling the LLM to learn about a region, or specific days and provide more precise & actionable suggestive policies. Embedding domain constraints (regulatory, resource, ethical) directly into generation and combine automated semantic scoring with expert ratings to refine metric weighting per locality and extending under-represented policy types using validated data.

Leverage the system's cold-start, data-scarce readiness to roll-out incrementally and integrate with municipal workflows as a continuous recommender with human approval. This is commercially plausible as a region-agnostic, SaaS-style decision-support tool[3], which is grounded in one dataset and transparent, auditable metrics.

Integrating autonomous AI agents will be capable of monitoring model performance, identifying drift, and retraining models using new incoming data. These agents could operate in a self-learning loop by automatically adjusting parameters, augmenting datasets, and refining policies without human oversight when confidence thresholds are met, while still escalating cases for human intervention in critical or low-confidence scenarios. This hybrid approach would enable adaptive, context-aware waste forecasting and policy generation that improves over time[15] and responds to changing waste patterns and operational requirements.

Overall, this research's goal was set to develop a holistic, data-driven framework that could detect waste, forecast its occurrence, and automatically generate context-aware, actionable policies which would bridge the gap between perception, prediction, and prescription in waste management. The outcomes of the evaluations demonstrate that this goal was substantially achieved by the system which successfully integrated multiple AI components (classification, prediction and policy generation) into a unified automated workflow, delivering measurable detection accuracy of the various waste types from the images, reliable per-class forecasting for identifying trends and patterns of waste generation, and high-quality policy recommendations based on the forecasted waste data & ranked for operational deployment. While certain areas such as rare-event prediction and minority-class detection present opportunities for enhancement, the core proof-of-concept validates the feasibility and practical value of such an integrated solution. This cements the work as a strong foundation for further research, scaling, and eventual real-world adoption.

References

[1] P. F. Proença and P. Simões, "TACO: Trash Annotations in Context for Litter Detection," *arXiv:2003.06975 [cs]*, Mar. 2020, Available: https://arxiv.org/abs/2003.06975

[2] S. Kaza, L. C. Yao, P. Bhada-Tata, and F. Van Woerden, *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*. Washington, DC: World Bank, 2018. Available: https://openknowledge.worldbank.org/entities/publication/d3f9d45e-115f-559b-b14f-28552410e90a

[3] S. Ahmed, S. Mubarak, J. T. Du, and S. Wibowo, "Forecasting the Status of Municipal Waste in Smart Bins Using Deep Learning," *International Journal of Environmental Research and Public Health*, vol. 19, no. 24, p. 16798, Dec. 2022, doi: https://doi.org/10.3390/ijerph192416798.

[4] M. Córdova *et al.*, "Litter Detection with Deep Learning: A Comparative Study," *Sensors*, vol. 22, no. 2, p. 548, Jan. 2022, doi: https://doi.org/10.3390/s22020548.

[5] UNEP, "Global Waste Management Outlook 2024," *UNEP - UN Environment Programme*, 2024. https://www.unep.org/resources/global-waste-management-outlook-2024

[6] United Nations, "The 17 sustainable development goals," *United Nations*, 2025. https://sdgs.un.org/goals

[7] D. Reis, J. Kupec, J. Hong, and A. Daoudi, "Real-Time Flying Object Detection with YOLOv8," May 2023, doi: https://doi.org/10.48550/arxiv.2305.09972.

[8] M. S. Reddy *et al.*, "pLitterStreet: Street Level Plastic Litter Detection and Mapping," *arXiv.org*, 2024. https://arxiv.org/abs/2401.14719

[9] A. Namoun, A. Tufail, M. Y. Khan, A. Alrehaili, T. A. Syed, and O. BenRhouma, "Solid Waste Generation and Disposal Using Machine Learning Approaches: A Survey of Solutions and Challenges," *Sustainability*, vol. 14, no. 20, p. 13578, Oct. 2022, doi: https://doi.org/10.3390/su142013578.

[10] L. Huang, T. Cai, Y. Zhu, Y. Zhu, W. Wang, and K. Sun, "LSTM-Based Forecasting for Urban Construction Waste Generation," *Sustainability*, vol. 12, no. 20, p. 8555, Oct. 2020, doi: https://doi.org/10.3390/su12208555.

[11] R. Rafailov, A. Sharma, E. Mitchell, S. Ermon, C. D. Manning, and C. Finn, "Direct Preference Optimization: Your Language Model is Secretly a Reward Model," *arXiv.org*, May 29, 2023. https://arxiv.org/abs/2305.18290

[12] L. Tunstall *et al.*, "Zephyr: Direct Distillation of LM Alignment," *arXiv.org*, Oct. 25, 2023. https://arxiv.org/abs/2310.16944

[13] T. Schick *et al.*, "Toolformer: Language Models Can Teach Themselves to Use Tools," *arXiv:2302.04761 [cs]*, Feb. 2023, Available: https://arxiv.org/abs/2302.04761

[14] N. Crasto, "Class Imbalance in Object Detection: An Experimental Diagnosis and Study of Mitigation Strategies," *arXiv.org*, Mar. 11, 2024. https://arxiv.org/abs/2403.07113

[15] G. Piatti *et al.*, "Cooperate or Collapse: Emergence of Sustainable Cooperation in a Society of LLM Agents." Accessed: Aug. 10, 2025. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2024/file/ca9567d8ef6b2ea2da0d7eed57b933ee-Paper-Conference.pdf

[16] "Dublin City Council Litter Management Plan 2020-2022 Working Together to Keep Dublin Clean Acknowledgement." Available: https://www.dublincity.ie/sites/default/files/2021-01/dublin-city-litter-management-plan-2020-2022.pdf