Urban Solid Waste Management Using Machine Learning: A Case Study of Dega Town, Ethiopia.

Gudeta Tesema Mamo [1]*, Tsegaye Hordofa Gudeta [1], Misgaye Workneh Abdisa [1]

[1] *College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China*

---***---

Abstract - *Accurate forecasting of waste generation is crucial for effective urban waste management, especially as cities grow and urbanize. This study evaluates the performance of three machine learning models, Decision Tree, Random Forest, and Long Short-Term Memory networks, in predicting waste generation trends in Dega Town, Ethiopia, using data from 2014 to 2023. The models were assessed based on Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, and R-squared.*

The Random Forest model performed the best, achieving an R^2 of 0.9847, indicating it explained 98.47% of the variance in the data. It also had the lowest MAE (1801.62), RMSE (3681.04), and MSE (13,550,064), making it the most accurate model for forecasting waste generation. In contrast, the Decision Tree model showed an R^2 of 0.9773, with MAE of 2162.43, RMSE of 4481.44, and MSE of 20,083,317, suggesting moderate accuracy but higher error rates. The LSTM model had the lowest performance, with an R^2 of 0.9658, MAE of 2923.58, RMSE of 5500.17, and MSE of 30,251,926, highlighting challenges in capturing long-term dependencies. These results indicate that Random Forest is the most effective model for waste generation forecasting in urban areas. The study also emphasizes incorporating waste-type data (e.g., food, plastics, paper) to improve predictive accuracy. The findings provide valuable insights for waste management authorities to optimize resource allocation and effectively plan future waste volumes.

***Key Words*:** Urban Solid Waste Management, Machine Learning, Decision Trees, Random Forest, LSTM, Waste Generation

1. INTRODUCTION

Urbanization and population growth have created significant challenges for solid waste management (SWM) in many developing countries, including Ethiopia. Rapid urban expansion often outpaces the capacity of local authorities to manage waste effectively. Dega Town, an emerging urban area in Ethiopia, exemplifies these challenges, with waste generation rates increasing due to a growing population and urban migration. Effective waste management is crucial for public health, environmental protection, and urban sustainability. Poor waste management practices can lead to health hazards, pollution, and decreased quality of life for residents[1]. In Dega Town, inadequate waste collection and disposal systems have resulted in littering, illegal dumping, and increased incidence of diseases related to poor sanitation. Current waste management strategies in Dega Town are largely reactive and lack adequate foresight and data-driven approaches. The town suffers from inadequate waste collection infrastructure, inconsistent collection schedules, and limited recycling initiatives. As a result, a significant portion of generated waste remains uncollected or improperly disposed of, exacerbating environmental degradation and health risks.

Predictive analytics is a powerful tool that uses historical data, statistical algorithms, and machine learning techniques to forecast future trends. In the context of waste management, predictive analytics plays a crucial role in optimizing waste collection, improving resource allocation, and enhancing the overall efficiency of waste management systems. By analyzing patterns in waste generation, predictive models provide actionable insights that help waste management authorities make informed decisions regarding collection schedules, resource distribution, and policy development [2]. Studies have demonstrated that implementing predictive analytics leads to significant cost savings and improves service delivery by allowing authorities to optimize routes and minimize operational expenses [3]. Furthermore, by forecasting waste generation trends, predictive models enable targeted public awareness campaigns by identifying areas with the highest waste production, ensuring that educational efforts are focused where they are most needed [4].

The association between waste generation and demographic, socio-economic, and cultural factors is welldocumented. As urban populations rise, so does the volume of waste generated. According to [5], cities in low-income countries produce significantly less waste per capita compared to those in high-income countries. However, rapid urbanization in these regions is expected to exacerbate waste management challenges. Urbanization not only increases the total volume of waste but also alters its composition. For instance, the shift toward consumeroriented lifestyles typically leads to a rise in packaging waste and organic waste [6]. Studies show that urban areas with higher population densities tend to experience greater waste generation per capita [7], a trend that is especially pronounced in rapidly growing cities across the globe.

Urban solid waste generation (USWG) has evolved dramatically over the past decades, influenced by a confluence of factors, including demographic shifts, economic development, urbanization, technological advancements, and changing consumption patterns. Understanding these historical trends is essential for developing effective waste management strategies in the face of growing urban populations and increasing waste volumes. As populations grow, particularly in urban areas, the amount of waste produced tends to increase. Urbanization has been a significant driver of waste generation, as evidenced by the rapid expansion of cities worldwide. According to the United Nations [8, 9], the global urban population is projected to reach 68% by 2050, up from 55% in 2018. This urban migration often leads to higher waste production due to consumption behaviors associated with urban lifestyles.

The relationship between population density and waste generation is particularly noteworthy. Research shows that cities with higher population densities generate more waste per capita than rural areas. For example, Ribić et al. (2017)[10] found that densely populated urban areas in Europe produced significantly more waste, primarily due to increased packaging and consumer goods. As metropolitan areas continue to expand and populations grow, the infrastructure required for effective waste management often struggles to keep up with the rising waste volumes.

Economic development plays a critical role in shaping waste generation patterns. As countries transition from lowincome to middle-income and eventually high-income status, waste generation tends to increase. This pattern is particularly evident in countries undergoing rapid economic growth, such as China and India. In these developing nations, economic growth is often accompanied by a rise in consumption levels, which leads to a sharp increase in waste generation. Kaza et al. (2018) [11] reported that waste generated in low-income countries is substantially lower than in high-income countries, primarily due to lower consumption rates. However, as these countries urbanize and develop, waste production rises significantly. Conversely, high-income countries tend to produce more waste per capita due to higher consumption levels of goods and services.

The composition of waste also changes with economic development. As economies grow, the types of waste generated shift from organic and biodegradable materials to more complex waste streams, including plastics, electronics, and hazardous materials. This shift presents additional challenges for waste management systems, which must adapt to handle new waste streams, requiring more sophisticated recycling and disposal methods.

Technological advancements have significantly shaped modern waste generation trends. The widespread use of plastics, for example, has drastically altered the waste landscape since the mid-20th century[12, 13]. Due to their versatility and convenience, plastics are now ubiquitous in packaging and consumer products. This increase in plastic waste has led to significant environmental concerns, as plastics are non-biodegradable and accumulate in landfills and the natural environment.

Among the most exciting developments in waste management is the integration of machine learning into predictive analytics. Machine learning techniques offer powerful capabilities for modelling the complex relationships between various socio-economic, demographic, and environmental factors that influence a waste generation. Among the most used methods are regression models, which can be classified into linear regression, multiple regression, and non-linear regression.

Linear regression helps establish a simple relationship between waste generation and independent variables, such as population size or income level. Kaza et al. (2018)[11] used linear regression to explore waste generation patterns across different cities, finding that factors such as population density and economic status significantly impacted waste volumes. Multiple regression, which accounts for multiple independent variables simultaneously, provides a more nuanced understanding of waste generation. Zhou et al. (2022)[14] applied multiple regression to analyze socioeconomic factors affecting waste production in urban China, revealing that income and educational attainment were key predictors of waste generation.

Non-linear regression is used when relationships between variables are not linear. Abdallah et al. (2020)[15] applied non-linear regression to model waste generation in Tehran, achieving more accurate predictions by incorporating economic and seasonal factors. Time series regression, which analyzes historical data to identify trends, seasonal variations, and cyclical patterns, has also proven useful. Beigl et al. (2008) [16] used time series regression to study waste generation in Vienna, uncovering seasonal trends that could inform waste management strategies.

Decision trees and RF are other machine-learning techniques commonly used in waste management. Decision trees model decisions and their possible outcomes, while random forests combine multiple decision trees to enhance prediction accuracy[17]These methods have been successfully applied in various studies to predict waste generation, waste composition, and waste management efficiency. Almeida et al. (2023)[18] used random forests to examine waste generation factors in Brazil, achieving better accuracy than traditional regression models.

Understanding the factors influencing waste generation is key to developing effective management strategies. Demographic characteristics, such as population density, age distribution, and household composition, significantly affect waste generation patterns. Studies by Voukkali et al. (2024) [19] emphasize that urban areas with higher population densities tend to generate more waste per capita. Economic indicators, such as income levels, also play a significant role. Higher income levels are typically associated with increased consumption and waste production [20]. Cultural attitudes toward waste disposal, including the level of public awareness and participation in recycling programs, also influence waste generation patterns [21, 22].

Urbanization, particularly in developing countries, increases the demand for goods and services, which leads to higher waste production. As urban populations grow, infrastructure needs for waste management often lag, exacerbating the challenges of waste collection, disposal, and recycling. Cultural and socio-economic factors also shape the types of waste generated. In town areas, for example, the prevalence of single-use plastics and packaging materials has increased dramatically due to consumer convenience and marketing strategies. This shift poses challenges for recycling systems and waste management infrastructure. This study aims to evaluate the urban solid waste management system in Dega Town by using machine learning techniques to develop predictive models that accurately forecast waste generation trends, considering various socio-economic and demographic factors. By understanding these trends, waste management authorities can implement more effective and sustainable practices that improve waste management efficiency and promote environmental sustainability.

2. Materials and methods

2.1 Study Area

Dega Town is in the western part of Ethiopia, nearly 548 km from Addis Ababa, the capital city. This strategic location places it within the administrative boundaries of the Oromia region, making it a vital local hub for various socio-economic activities. Several neighboring areas, including Northern Chora Woreda to the north, Dabo Hanna Woreda to the east, Meko Woreda to the west, and Alge Sachi Woreda to the south border Dega Town. The surrounding districts collectively contribute to the unique demographic and economic landscape of Dega Town, which Contributes significantly to broader regional growth.

Spatially, Dega Town is positioned between latitude 8°37′31" and 8°31′6" North and longitude 36°5′43" to 36°6′48" East. This location situates the town in a region characterized by a diverse topography, ranging from rolling hills to flat plains, alongside various climatic conditions that can affect waste generation and management practices. The altitude of Dega Town varies between 1,700 and 2,300 meters above sea level, which further influences local environmental conditions, agricultural practices, and population density.

The Dega Town altitude and regional features create a unique microclimate that affects the types and volumes of waste generated. For instance, agricultural activities in the surrounding areas contribute significantly to organic waste, while urbanization fuels the production of packaging and industrial waste. Understanding these dynamics is essential for developing effective waste management strategies tailored to the specific needs of Dega Town. In recent years, Dega Town has experienced population growth and urban expansion, leading to increased pressure on waste management systems. This growth has been accompanied by changes in consumption patterns and waste generation, necessitating a comprehensive waste management approach incorporating local characteristics. The town's socioeconomic activities, such as commerce, trade, and agriculture, further complicate waste management, highlighting the importance of integrating machine learning methods to predict and optimize waste handling.

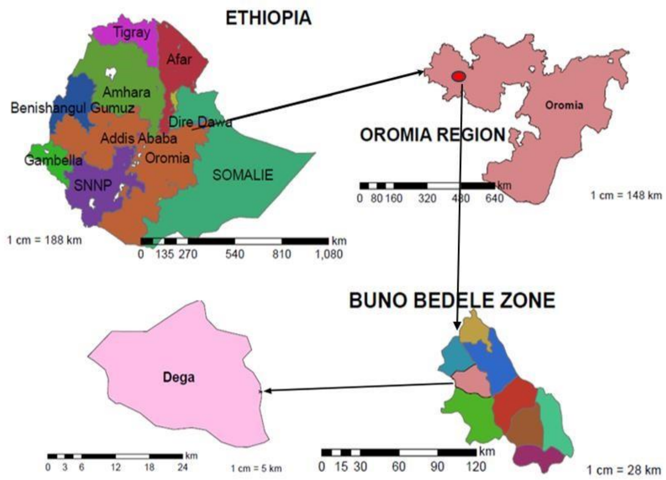

Fig – 1: Shows study area map

2.2 Data Sources and Collection

The waste data for this study were obtained from the waste management authorities in Dega Town. Collection methods included direct observations, waste audits, and reports from waste management teams. Monthly and yearly data on nine different waste types were compiled from 2014 to 2023. Municipal staff conducted regular waste audits in designated areas, the foundation for estimating waste volumes. These audits measured and categorized waste collected from residential, commercial, and industrial sectors. Additionally, official Dega Town waste management department reports provided supplementary information on waste generation trends, disposal methods, and recycling activities. To ensure reliability, the data were cross-verified for consistency. A thorough review of the records was performed to enhance data accuracy, addressing discrepancies such as missing or inconsistent entries through consultations with local authorities and waste collection services. The finalized data were organized into a structured database, with separate columns for each waste type, waste volume (in kilograms), and the corresponding period for each year in the dataset.

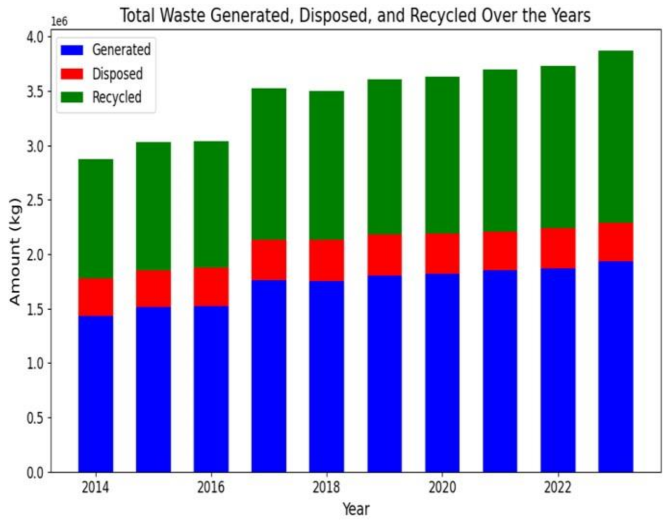

Chart -1: Illustrates yearly data trends

2.3 Data Preprocessing

Data preprocessing is essential for ensuring the quality and reliability of the dataset before analysis. The collected data were subjected to a rigorous cleaning process. This involved identifying and addressing missing values, inconsistencies, and outliers. Missing values were handled through interpolation methods, which estimate the missing data points based on existing values. Outliers were identified using statistical techniques and were either corrected or removed to maintain the integrity of the dataset. Normalization techniques were applied to ensure the data features were on a consistent scale. This step was particularly important given that the dataset included various waste types measured in different units (e.g., kilograms for weight). Normalization helped improve the performance of the Random Forests model by ensuring that no single feature dominated the others due to differences in scale. The cleaned and normalized data were then structured into a comprehensive database. Each entry included columns for each waste type, waste volume (in kilograms), demographic factors, and the corresponding time for each year in the dataset. This structured approach facilitated efficient analysis and model training.

3. Machine Learning Models

In this research, three different machine learning models were utilized to predict waste generation in Dega Town, each selected for its unique strengths and capabilities in handling different aspects of the data. The models employed include Decision Trees, Random Forests, and LSTM. These models were selected to address the complexity of waste generation patterns, which multiple socio-economic, demographic, and temporal factors can influence. The following subsections provide an in-depth investigation of each model, outlining their theoretical underpinnings, methodologies, and roles in urban solid waste management.

3.1 Decision Trees

Decision Trees are the most extensively used tools for forecasting and predictive modeling. The fundamental goal of this approach is to develop a model that predicts the value of a target variable based on multiple input variables or features. Decision trees' versatility and interpretability make them powerful tools for a wide range of applications in classification and regression tasks.

Decision Trees can be generally classified into two categories. The first category is classification, where the objective is to assign data points to specific classes or categories. For example, a Decision Tree might predict whether a piece of waste belongs to a specific category, such as recyclable or non-recyclable. The second category is regression, where the goal is to predict a continuous real number as the output. In this case, Decision Trees are used to predict values like the amount of waste generated or the total cost associated with waste management. In the context of waste management, Decision Trees are particularly useful for identifying and predicting key patterns in waste disposal. For example, they are commonly applied to pinpoint illegal waste disposal locations by analyzing factors such as geographic features, demographic data, and waste disposal behaviors. By leveraging these variables, Decision Trees can identify regions with a higher likelihood of illegal dumping, allowing authorities to target enforcement efforts more effectively.

Furthermore, Decision Trees play a crucial role in analyzing waste production behaviour patterns. Decision Trees can reveal underlying trends and help predict future waste production by examining historical waste generation, disposal, and recycling data[23]. This is valuable for municipalities and waste management companies to plan for waste collection, recycling strategies, and resource allocation. Beyond waste identification and behaviour prediction, Decision Trees are also used for more specific forecasting tasks, such as estimating future waste generation, classifying different types of waste, and optimizing waste compression processes. Decision Trees are an essential tool in the realm of waste management, offering a range of applications from forecasting waste generation to identifying illegal disposal practices[23]. Their ability to handle both categorical and continuous data and their simplicity and interpretability make them an asset in enlightening the effectiveness and sustainability of waste management systems.

3.2 Random Forests

Random Forest (RF) is an ensemble machine-learning algorithm that leverages the power of multiple decision trees to make predictions. As an extension of the bootstrap aggregation (bagging) method, Random Forest combines the outputs of various individual decision trees, each trained on a random subset of the data, to create a robust and accurate model. This ensemble approach mitigates the weaknesses of single decision trees, such as overfitting and high variance, and enhances the model's generalization ability.

Random Forests are highly versatile and can be applied to classification and regression problems. RF is commonly used in classification tasks to categorize data into distinct classes, while they predict continuous outcomes in regression tasks[24, 25]. The algorithm excels with structured datasets, where the relationships between features and target variables are complex and non-linear. It is particularly effective when the underlying data contains a mix of categorical and numerical features.

Although Random Forests are predominantly used for classification and regression problems, they also have utility in time series forecasting. However, to apply Random Forests to time series data, the time series must first be transformed into a supervised learning problem. This transformation involves creating lag features, which allow the model to learn from past values to predict future outcomes. This approach can be particularly useful in forecasting solid waste generation, where historical data plays a crucial role in predicting future trends.

When applying RF to time series forecasting, evaluating the model using walk-forward validation rather than traditional k-fold cross-validation is essential. Walk-forward validation is more suitable for time series data because it respects the temporal order of the data, preventing future values from "leaking" into past data. Alternatively, k-fold cross-validation may result in optimistically biased results, as it does not account for the sequential nature of time series data[26, 27]. This study uses RF regression as a forecasting model for predicting solid waste generation. By utilizing lag features, we aim to capture the data's temporal dependencies and improve our forecasts' accuracy. This approach allows us to model the patterns of waste generation over time, considering past behaviour to predict future waste volumes effectively.

3.3 Long Short-Term Memory Networks

Long Short-Term Memory (LSTM) networks are specialized Recurrent Neural Networks (RNN) designed to address the vanishing gradient problem, which occurs in traditional RNNs when long-term dependencies become difficult to learn. LSTMs incorporate memory cells that allow them to store and retrieve information over extended periods, making them well-suited for sequential data where past information is important for predicting future outcomes. Unlike standard RNNs, LSTMs use three key gates: input, forget, and output gates, which regulate the flow of information, enabling the model to remember or forget data at each step selectively.

The advantage of LSTMs over traditional models is their Capability to capture and model complex, nonlinear relationships in data [28, 29]. This is especially useful in time series forecasting, where patterns are often not linear and can change over time. In applications such as waste generation forecasting, LSTMs can analyze historical data to identify trends, seasonal fluctuations, and other temporal dependencies, providing accurate predictions even in the presence of non-linear interactions between various factors like population growth, economic conditions, and consumption behaviors. LSTMs excel in tasks that require learning from long data sequences, making them particularly effective for forecasting real-valued time series, such as predicting waste volumes over time. By leveraging their memory capabilities, LSTMs can account for long-term trends while adapting to new information, ensuring that they provide reliable predictions. This makes LSTM networks a powerful tool for optimizing waste management systems, enabling better resource planning and more effective decision-making[30].

3.4 Model testing and validation

The effectiveness of regression algorithms in predicting waste generation was evaluated by testing the trained models with a separate validation dataset that was not part of the training process. This method ensures that the model's performance is evaluated on data they have not previously encountered, providing a more accurate estimate of their ability to generalize to new, unseen data. The validation process consisted of two stages: parameter optimization and the training and prediction phases.

A grid search technique with tenfold cross-validation was employed in the parameter optimization phase. This methodology systematically explored different combinations of hyperparameters to identify the best parameters for each tree-based model. Grid search ensures that the models are finely tuned, enabling them to perform at their highest capacity by testing multiple configurations and selecting the one that minimizes error and improves model robustness[30].

After fine-tuning the model's parameters, the training and prediction process began. The best-performing hyperparameters were applied to train the models using the given data. Once trained, the models made predictions on the validation dataset. To assess their performance, various error metrics were used, including Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2). MSE and MAE help measure how far the predictions deviate from actual values, while R^2 indicates how well the model explains the variability in the data.

By comparing the training and test scores, we gained an understanding of the model's generalizability and its ability to perform well on new, unseen data. This comprehensive testing and validation process allowed for a clear assessment of each model's predictive capabilities and error rates, ensuring that the most suitable model was selected for forecasting future waste generation.

$$MAE = \frac{\sum_{i=1}^{n}(Y_i - X_i)^2}{n} \quad (1)$$

Where Y_i is the predicted value, X_i is the actual value, and n is the total number of data points.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2 \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y}_i)^2} \quad (3)$$

Where $\hat{Y}_i$ the model is predicted value of waste generation rate, Y_i is the actual (data) value and $\overline{Y}_i$ is the mean, and n is the number of observations.

The Sum of Squared Errors (SSE) was expressed as a percentage based on the Root Mean Squared Error (RMSE). Both SSE and R^2 were used as key performance measures for evaluating the training and testing datasets. In most cases, modifications to the model's parameters and structure led to lower errors in the training phase compared to the testing phase.

4 Results and Discussion

Table 1 showcases the performance of three predictive models: Decision Tree, Random Forest, and LSTM. It provides a summary of how well each model performs based on key evaluation metrics, including Mean Absolute Error (MAE), MSE, RMSE, and R^2. These metrics play a vital role in assessing the accuracy and reliability of the model's predictions.

Table -1: Presents a comparison of the three models across the four evaluation metrics

Model	MAE	MSE	RMSE	R^2
Decision Tree	2162.43	20,083,317	4481.44	0.9773
Random Forest	1801.62	13,550,064	3681.04	0.9847
LSTM	2923.58	30,251,926	5500.17	0.9658

The Decision Tree model presented performance with an R^2 value of 0.9773, meaning that it was able to explain 97.73% of the variance in the data. This shows that the Decision Tree successfully captured the major trends in waste generation for Dega Town. The relatively low MAE of 2162.43 and RMSE of 4481.44 further suggest that the model's predictions were fairly accurate, with moderate deviations from actual values. However, the MSE of 20,083,317 was relatively high, indicating that the model had some large prediction errors. This is typical for Decision Tree models, which can sometimes overfit the data, especially excellent performance, suggesting that Random Forest was able to more effectively capture complex relationships in the data compared to the Decision Tree model. Additionally, the MAE of 1801.62 and RMSE of 3681.04 were the lowest among the models, reflecting that the Random Forest made the most accurate predictions. The MSE of 13,550,064 was also the smallest, indicating that the model was better at minimizing errors and outliers. The strong performance of Random Forest can be attributed to its ensemble approach, which aggregates the outputs of multiple decision trees trained on random subsets of the data. This helps decrease overfitting and increases the model's generalizability. Given its superior accuracy, Random Forest stands out as the best model for forecasting waste generation in Dega Town, where multiple factors are at play.

The LSTM model, while capable of capturing sequential dependencies in data, performed less effectively than the

Decision Tree and RF models. The R^2 value of 0.9658 indicates that the model was able to explain 96.58% of the variance, which is lower than the other two models. The MAE of 2923.58 and RMSE of 5500.17 were the highest, indicating that the predictions made by the LSTM model had larger deviations from the actual values. Moreover, the MSE of 30,251,926 was significantly higher than that of both the Decision Tree and RF models, highlighting the challenges the LSTM model faced in accurately forecasting waste generation. The lower performance of the LSTM model could be due to several factors, such as underfitting, insufficient model complexity, or inability to capture long-term dependencies in the data properly. LSTMs are generally effective for sequential data, but they may require more extensive finetuning or additional features to capture the nuances in waste generation trends.

A vital observation in this research is the exclusion of waste type as a specific feature in the predictive models. Diverse types of waste (like organic, plastics, paper, metals, glass, and electronics) are generated at varying rates and follow distinct generation patterns. The absence of these details could have limited the models' ability to capture the full complexity of the waste generation character in Dega Town. Including waste type as a separate feature could have significantly improved the models' predictive accuracy.

Food waste, for example, can fluctuate seasonally, often reaching higher levels during holidays or festivals. Plastic waste tends to reflect shifts in consumption patterns, particularly those linked to packaging. Organic waste, meanwhile, is usually connected to regional agricultural activities. By accounting for these variations, the model could more accurately predict waste generation, not just in total, but for each type of waste[31-33]. Moreover, the inclusion of waste types would allow waste management authorities to design more targeted strategies for resource allocation, recycling initiatives, and waste reduction efforts.

The RF model emerged as the best performer in this study, demonstrating the highest accuracy and reliability in predicting waste generation in Dega Town. Its ability to handle multiple variables and complex relationships made it particularly well-suited. Decision Trees, while effective, presented slightly lower predictive accuracy, likely owing to their susceptibility to overfitting. LSTM models, while valuable for handling time-series data, struggled with underfitting in this case, possibly due to insufficient model complexity or inadequate feature engineering.

The findings from this study underscore the need for a nuanced approach to waste management, one that not only predicts total waste generation but also accounts for the types of waste being generated. Incorporating waste-type data into the model could lead to better-targeted waste management strategies, optimizing resource allocation, and improving waste collection schedules. Furthermore, Random Forest's superior performance suggests that it could be a reliable tool for waste management authorities to forecast future waste trends, enhance operational efficiency, and promote sustainable practices. The results of this research provide valuable insights for municipal waste management authorities in Dega Town and similar urban areas. The RF model's ability to predict waste generation trends with high accuracy offers a foundation for improving resource planning and optimizing waste collection strategies. Authorities can leverage this model to anticipate future waste volumes, allocate resources more effectively, and reduce operational costs. Furthermore, the model could serve as a tool for guiding public awareness campaigns, identifying areas with high waste generation, and promoting more sustainable waste management practices.

Future work should focus on integrating waste-type data into predictive models to enhance their accuracy. Furthermore, improvements to the LSTM model, for instance, adjusting its architecture or incorporating more training data, could make it more effective for forecasting waste generation trends. The use of hybrid models, which combine the strengths of multiple machine learning techniques, could further improve forecasting accuracy. With continued refinement and testing, these predictive models have the potential to revolutionize urban waste management systems, providing data-driven solutions to meet the challenges of growing urban populations and increasing waste production.

4. CONCLUSIONS

This study quantitatively assessed urban solid waste management, particularly addressing the prediction of waste generation in Dega Town, Ethiopia, using machine Learning Approaches. Prediction of waste generation trends was done using three main methods: decision tree, random forest, and LSTM networks. The evaluation of these models has been carried out through four metrics: MAE, MSE, RMSE, and R^2. Among them, the Random Forest model exhibited the highest accuracy and reliability in predicting waste generation.

The RF model achieved an R^2 score of 0.9847, which means that it was able to actually explain more than 98% of the fluctuation in the data of waste generation. This indicates that it is able to capture the underlying patterns and relationships in the data, making it an optimal model for waste generation forecasting in urban environments. With an MAE of 1801.62 and an RMSE of 3681.04, the model results also provide further evidence for its performance, specifically, that predictions were not just accurate but consistent as well. Random Forest uses an ensemble learning method and builds multiple decision trees by training them on random datasets. It aggregates the results from those multiple decision trees, yet it performs better than the previous models. Doing so minimizes overfitting, improves generalization, and significantly increases the model's performance when making predictions on new/ahead data. Because decision trees are simple models that are easy to interpret, they are more likely to be overfitted by creating complex structures when interacting with complicated datasets. For this particular study, the decision Tree was calculated to have an R^2 of 0.9773, still strong but inferior to that of the Random Forest. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) values were larger than those from RF, suggesting that the Decision Tree model had somewhat larger deviations in its outputs. However, due to its high interpretability, it can significantly help in understanding the foremost waste generation factors. Decision trees can also be very beneficial for decisionmakers because they need to understand how features in the data relate to one another, and decision trees give clear rules for making decisions — an important requirement in creating targeted waste management strategies.

The LSTM model, despite being well-suited for time-series forecasting, faced difficulties in this application, resulting in the lowest performance among the three models. The R^2 value of 0.9658 for the LSTM model indicates that it was able to explain only 96.58% of the variance in the data. Additionally, the MAE, RMSE, and MSE values were notably higher than those for the other models, suggesting that the LSTM was less effective in capturing the trends in waste generation. The relatively poor performance of the LSTM model could be attributed to a variety of factors, including underfitting, where the model did not thoroughly learn the complexities of the data, or insufficient model complexity, which limited its ability to capture long-term dependencies in the time series effectively. LSTM models are generally effective for capturing temporal dependencies in sequential data. Still, in this case, they did not outperform simpler models like Random Forest, likely due to the need for more fine-tuning and the possibility of missing key factors in the data.

One critical aspect that was not incorporated into the models but could have greatly improved their accuracy is the consideration of waste type. In the current models, total waste generation was considered, and different kinds of waste, such as food waste, plastics, paper, glass, and organic waste, are generated at different rates and influenced by distinct factors. Waste types exhibit different patterns of generation, and understanding these variations could improve the precision of waste generation forecasts. Including waste type data would allow the models not only to predict total waste generation but also to forecast the specific contributions of different waste categories. This would be vital for optimizing waste collection and recycling efforts, as diverse waste types require different management tactics.

Including the type of waste would allow municipalities and waste management authorities to adapt their approaches according to the unique challenges presented by each waste type. For example, some materials, such as plastics, may require more advanced recycling programs, or organic waste could be diverted for composting. More granular predictive modelling would also help to anticipate future demand for wasted disposal infrastructure to ensure cities are prepared to pick up not just more waste but also a different type of waste. Urban waste management from a machine learning perspective: With so much variation in how models perform, the results suggest the application of machine learning techniques may improve urban waste management practice. The Random Forest model significantly outperformed the other models, indicating that it is the most appropriate model for predicting waste generation in Dega Town, providing the best accuracy and prediction validity. With this model in place, a waste management authority can estimate future waste volumes and respond to this by optimizing resource allocation and developing more focused strategies around waste collection. Such information enables improved planning and more sustainable waste management systems that are better equipped to deal with the increasing volumes of waste resulting from urbanization and population expansion.

This study presents numerous opportunities for future research and advancements in waste generation modelling. The most important of these is that we need to include data regarding the waste type to improve the accuracy and precision of model predictions. This overview provides a more robust understanding of the relationships that drive these disposal pathways, allowing model and program designers to more successfully and appropriately develop complex recycling and disposal pathways that better account for the propensity to recycle for different waste types. One more encouraging route for future work would be to refine the LSTM model further. Existing models did not account for the complexity of input data and failed to capture long-term dependencies to achieve better performance. However, this model can help LSTM predict waste generation even for longer periods with better training data, changes in LSTM structure, and better feature engineering. Furthermore, the LSTM model can be optimized using hyperparameter tuning techniques such as changing the number of epochs, layers, and units of the model that would enable the model to gather insight from its learned patterns from past data to provide a higher degree of predictive power. Finally, hybrid models may be applied to benefit from the additional power of RF and LSTM. So, hybrid models accounting for both the interpretability and accuracy of RF and the inherent ability of LSTM to manage time series could potentially offer a more holistic approach to predicting waste generation. The first trend can be predicted using RF, while LSTM can make more accurate predictions by capturing the sequential relationships for a longer time.

The use of machine learning models in waste management forecasting represents a significant advancement in the field, offering the potential for more accurate, data-driven decision-making. The RF model stands out as a powerful tool for predicting future waste trends, but further work on incorporating waste type data and refining LSTM models will undoubtedly improve the ability to forecast waste generation and guide sustainable waste management strategies in urban areas. As cities continue to grow and urbanization accelerates, the importance of predictive models like these will only increase, helping municipalities stay ahead of waste management challenges and ensuring more sustainable cities for the future.

ACKNOWLEDGEMENT

The first author sincerely appreciates the Government of China for supporting higher education and Tongji University IESD for its academic resources and guidance. Thanks to the Ethiopian government and Dega Town authorities for cooperating during data collection. Deep gratitude to family, friends, and colleagues for their support and to Ebrahim Aman for his invaluable assistance.

References

[1] L. Hirpe and C. Yeom, "Municipal solid waste management policies, practices, and challenges in Ethiopia: a systematic review," *Sustainability,* vol. 13, no. 20, p. 11241, 2021.

[2] R. E. V. Sesay and P. Fang, "Circular Economy in Municipal Solid Waste Management: Innovations and Challenges for Urban Sustainability," *Journal of Environmental Protection,* vol. 16, no. 2, pp. 35-65, 2025.

[3] N. Alahmari, R. Mehmood, A. Alzahrani, T. Yigitcanlar, and J. M. Corchado, "Autonomous and Sustainable Service economies: Data-Driven optimization of Design and Operations through Discovery of Multi-perspective parameters," *Sustainability,* vol. 15, no. 22, p. 16003, 2023.

[4] F. Fasano, A. S. Addante, B. Valenzano, and G. Scannicchio, "Variables influencing per capita production, separate collection, and costs of municipal solid waste in the Apulia region (Italy): An experience of deep learning," *International Journal of Environmental Research and Public Health,* vol. 18, no. 2, p. 752, 2021.

[5] C. Pharino and C. Pharino, "Waste characteristics and practices," *Challenges for Sustainable Solid Waste Management: Lessons from Thailand,* pp. 1528, 2017.

[6] D. Reike, W. J. Vermeulen, and S. Witjes, "The circular economy: new or refurbished as CE 3.0?—exploring controversies in the conceptualization of the circular economy through a focus on history and resource value retention options," *Resources, conservation and recycling,* vol. 135, pp. 246-264, 2018.

[7] L. Izquierdo-Horna, R. Kahhat, and I. Vázquez-Rowe, "Reviewing the influence of sociocultural, environmental and economic variables to forecast municipal solid waste (MSW) generation," *Sustainable production and consumption,* vol. 33, pp. 809-819, 2022.

[8] U. N. O. f. P. Services, "Global State of National Urban Policy 2021 Achieving Sustainable Development Goals and Delivering Climate Action," 2021.

[9] X. Li *et al.*, "Global urban growth between 1870 and 2100 from integrated high resolution mapped data and urban dynamic modeling," *Communications Earth & Environment,* vol. 2, no. 1, p. 201, 2021.

[10] B. Ribić, N. Voća, and B. Ilakovac, "Concept of sustainable waste management in the city of Zagreb: Towards the implementation of circular economy approach," *Journal of the Air & Waste Management Association,* vol. 67, no. 2, pp. 241-259, 2017.

[11] S. Kaza, L. Yao, P. Bhada-Tata, and F. Van Woerden, *What a waste 2.0: a global snapshot of solid waste management to 2050*. World Bank Publications, 2018.

[12] S. E. Vergara and G. Tchobanoglous, "Municipal solid waste and the environment: a global perspective," *Annual review of environment and resources,* vol. 37, no. 1, pp. 277-309, 2012.

[13] A. Lakhouit, "Revolutionizing Urban Solid Waste Management with AI and IoT: A review of smart solutions for waste collection, sorting, and recycling," *Results in Engineering,* p. 104018, 2025.

[14] B. Zhou, F. Qi, M. F. Riaz, and T. Ali, "An Analysis of the Factors behind Rural Residents' Satisfaction with Residential Waste Management in Jiangxi, China," *International Journal of Environmental Research and Public Health,* vol. 19, no. 21, p. 14220, 2022.

[15] M. Abdallah, M. A. Talib, S. Feroz, Q. Nasir, H. Abdalla, and B. Mahfood, "Artificial intelligence applications in solid waste management: A systematic research review," *Waste Management,* vol. 109, pp. 231-246, 2020.

[16] P. Beigl, S. Lebersorger, and S. Salhofer, "Modelling municipal solid waste generation: A review," *Waste management,* vol. 28, no. 1, pp. 200-214, 2008.

[17] T. Shaikhina, D. Lowe, S. Daga, D. Briggs, R. Higgins, and N. Khovanova, "Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation," *Biomedical Signal Processing and Control,* vol. 52, pp. 456-462, 2019.

[18] T. A. d. C. Almeida, E. F. Felix, C. M. A. de Sousa, G. O. M. Pedroso, M. F. B. Motta, and L. P. Prado, "Influence of the ANN Hyperparameters on the Forecast Accuracy of RAC's Compressive Strength," *Materials,* vol. 16, no. 24, p. 7683, 2023.

[19] I. Voukkali, I. Papamichael, P. Loizia, and A. A. Zorpas, "Urbanization and solid waste production: Prospects and challenges," *Environmental Science and Pollution Research,* vol. 31, no. 12, pp. 1767817689, 2024.

[20] V. H. A. d. M. Vieira and D. R. Matheus, "The impact of socioeconomic factors on municipal solid waste generation in São Paulo, Brazil," *Waste Management & Research,* vol. 36, no. 1, pp. 79-85, 2018.

[21] B. Gu *et al.*, "Promoting Chinese urban residents' participation in source separation and recycling," *Waste Management,* vol. 139, pp. 290-299, 2022.

[22] L. Xiao, G. Zhang, Y. Zhu, and T. Lin, "Promoting public participation in household waste management: A survey based method and case study in Xiamen city, China," *Journal of cleaner production,* vol. 144, pp. 313-322, 2017.

[23] J. C. Bijos, V. M. Zanta, J. Morató, L. M. Queiroz, and K. P. Oliveira-Esquerre, "Improving circularity in municipal solid waste management through machine learning in Latin America and the Caribbean," *Sustainable Chemistry and Pharmacy,* vol. 28, p. 100740, 2022.

[24] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert systems with applications,* vol. 134, pp. 93101, 2019.

[25] N. Horning, "Random Forests: An algorithm for image classification and generation of continuous fields data sets," in *Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences, Osaka, Japan*, 2010, vol. 911, pp. 1-6.

[26] V. Cerqueira, L. Torgo, and I. Mozetič, "Evaluating time series forecasting models: An empirical study on performance estimation methods," *Machine Learning,* vol. 109, no. 11, pp. 1997-2028, 2020.

[27] I. K. Nti, O. Nyarko-Boateng, and J. Aning, "Performance of machine learning algorithms with different K values in K-fold cross-validation," *International Journal of Information Technology and Computer Science,* vol. 13, no. 6, pp. 61-71, 2021.

[28] D. Niu, F. Wu, S. Dai, S. He, and B. Wu, "Detection of long-term effect in forecasting municipal solid waste using a long short-term memory neural network," *Journal of Cleaner Production,* vol. 290, p. 125187, 2021.

[29] 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, 2020.

[30] M. Y. Shams, A. M. Elshewey, E.-S. M. El-Kenawy, A. Ibrahim, F. M. Talaat, and Z. Tarek, "Water quality prediction using machine learning models based on grid search method," *Multimedia Tools and Applications,* vol. 83, no. 12, pp. 35307-35334, 2024.

[31] C. M. Keske, M. Mills, T. Godfrey, L. Tanguay, and J. Dicker, "Waste management in remote rural communities across the Canadian North: Challenges and opportunities," *Detritus,* vol. 2, no. 1, p. 63, 2018.

[32] G. Kennedy, *An ontology of trash: The disposable and its problematic nature*. State University of New York Press, 2012.

[33] W. Chaabane, "Solid waste management in tourism destinations in Tunisia: diagnostic and improvement approaches," Universität Rostock, 2020.