Predictive Modeling of Hospital Waste Generation Using Machine Learning Based on Patient Inflow

Muhammad Arshad[1], Muhammad Umer Farooq[1], Najeeb Ullah Khan[1], Kifayat Ullah[2], Bilal Ur Rehman[2], Salman Ilahi Siddiqui[2], Muhammad Farooq[2], Muhammad Kashif [2]
[1]Department of Industrial Engineering, Jalozai Campus, UET Peshawar, Pakistan
[2]Department of Electrical Engineering, UET Peshawar
*Correspondence: kifayat.bangash@uetpeshawar.edu.pk
Citation | Arshad. M, Farooq. M. U, Khan. N. U, Ullah. K, Rehman. B. U, Siddiqui. S. I, Farooq. M, Kashif. M, "Predictive Modeling of Hospital Waste Generation Using Machine Learning Based on Patient Inflow", IJIST, Vol. 07 Issue. 03 pp 1612-1629, July 2025
Received | June 11, 2025 **Revised |** July 23, 2025 **Accepted |** July 24, 2025 **Published |** July 26, 2025.

Effective hospital waste management is a key to the security of the environment and the provision of healthcare hygiene. This paper develops a predictive analytics model to forecast the amount of daily hospital waste generated based on patient inflow using a linear regression model. Real-time data from 60 days were gathered in a tertiary healthcare institution, which informed the number of patients and the resultant waste in kilograms. The data is taken from Lady Reading Hospital (LRH), Peshawar, Pakistan. The model obtained an R value of 0.88 in the training and 0.81 in the validation datasets and a Root Mean Square Error (RMSE) of 130.52 kg. The predictor of patient volume was established as significant through statistical validation via ANOVA, and the model was found to be within the key regression assumptions through the residual analysis. The findings emphasize that predictive modeling within a hospital waste planning system is viable, and a cost-efficient and explainable option can be used in operational forecasting. The offered method contributes to enhanced resource distribution, risk aversion, and adherence to the sustainable healthcare objectives.
Keywords: Predictive Modeling, Hospital Waste Estimation, Linear Regression, Patient Inflow, ANOVA Validation, Healthcare Forecasting, Waste Management Optimization

Introduction:

Effective hospital waste management is essential not only for maintaining hygiene but also as a crucial factor in operational efficiency and environmental sustainability. In today's context of rising patient admissions and constrained resources, healthcare facilities urgently require systems capable of predicting future waste generation. Such predictive systems would enable proactive waste management, shifting the approach from reactive responses to strategic planning and resource optimization. Predictive analytics has become the tool that is used to address these needs. By predicting waste quantities based on the trend of patient inflow, hospitals can manage staffing, arrange collection schedules, and maintain compliance with regulations. This paper, therefore, explores the role of predictive modeling in enhancing hospital waste management infrastructure and facilitating a shift from reactive to proactive waste handling practices. Hospital waste typically includes both biological and non-biological materials that are rendered unusable and require proper disposal [1]. It is generally divided into medical, domestic, and infectious waste. Medical waste is generated in diagnosis, treatment, and immunization procedures, whereas infectious waste contains material that can cause disease through contact with a patient [2]. Hospital waste can be very diverse in quantity and complexity, which is dependent on the activity of the facility. Hospitals, laboratories, and home-based medical treatment are the sources of health waste through dialysis and insulin injections [3]. When not managed properly, such waste can pose significant health risks and disrupt hospital operations.

Figure 1. Hospital Waste [4]

A large percentage of biomedical waste, which is estimated to be 75-90%, is similar in properties to that of municipal waste. The remaining 10-25% is dangerous and can carry diseases like HIV, hepatitis B and C, tuberculosis, cholera, etc. [5][6]. According to estimates by the World Health Organization (WHO), unsafe syringe practices have led to approximately 2 million cases of HIV and 21 million cases of hepatitis B infections [7]. This highlights the importance of accurate prediction and the creation of effective disposal plans to prevent the threat to the general population. Hazardous medical waste, if not segregated and disposed of properly, can threaten healthcare workers, waste handlers, and the surrounding population [8][9]. Infectious and hazardous items are syringes, intravenous bottles, gloves, soiled dressings, and chemical-containing containers [10]. When these materials are poorly classified or allowed to accumulate uncontrolled, operating costs and safety liabilities can be significantly higher. Despite the implementation of hospital waste regulations, many healthcare facilities are still using manual ways of estimating the quantity of waste and planning disposal. Hospital waste, although limited in quantity, is highly infectious and poses significant risks if mismanaged [11]. Inaccurate forecasting of the waste usually results in overflows, delayed pick-ups, and wastage of resources. Proper management of biomedical wastes thus relies on credible waste volume trends data to make evidence-based decisions.

Researchers in Pakistan report that hospitals generate approximately 2.0 kg of waste per bed per day, of which 0.1 to 0.5 kg is categorized as hazardous [12]. Without predictive tools, hospitals are forced to estimate needs based on trial and error, leading to resource wastage and a lack of safety standard compliance. The response to this should be that healthcare facilities developed written policies that fit the nature of their activities to maintain uniformity and safe management of waste [13]. However, policy alone is insufficient without accurate forecasting models. The use of predictive analytics allows hospitals to predict the daily waste generation based on historical data regarding the inflow of patients and machine learning algorithms. The danger of hazardous waste is spread to everyone who encounters the waste: hospital staff, visitors, and waste operators. The problem of diseases that are transmitted by body fluids, especially Hepatitis B, Hepatitis C, and AIDS, is a critical issue in the context of healthcare waste [14]. Waste trend anticipation will allow hospitals to implement interventions on time, minimize risks of exposure, and guarantee the safety of workers.

The contribution of this research paper is the development of a predictive model that uses Machine Learning Algorithms (MLA), especially linear regression, to estimate the amount of hospital waste based on the daily arrival of patients. The simplicity, computational efficiency, and strong correlation between patient volume and waste generation make this modeling framework an ideal choice for implementation in hospital settings. By incorporating predictive analytics, the approach aims to improve the accuracy of waste forecasts, support the timely operational planning process, and help reduce the health risk that comes with improper waste management.

This study aims to leverage machine learning for predicting hospital waste generation, analyze the model's performance to streamline planning activities, and reduce reliance on manual estimations by implementing data-driven decision-making approaches.

The novelty of this study lies in its exclusive use of a daily patient inflow as a single predictive variable, which allows finding an accurate estimation of hospital waste in real-time operation scenarios that are lightweight and scalable.

Literature Review:

Predictive Techniques for Healthcare and Waste Forecasting:

Predictive modeling is an essential element of healthcare since it enables the forecasting of disease progression, needs, and processes. The current literature review summarizes the empirical research that uses empirical methods (statistical and machine learning) in healthcare settings, with particular importance to the hospital waste generation estimation. Rath et al. examined the predictability of the COVID-19 case trend in India with reference to the findings of the World Health Organization data by regression models. Through the introduction of multiple linear regression and linear regression, the research achieved strong relationships between confirmed, active, and recovered cases. The R2 value is 0.99 for all the predictive models, which depicts high accuracy. These results earn the relevance of regression models in the prediction of time-dependent healthcare data and level them up to the predictive hospital job, such as accurate estimation of waste [15]. Badawy et al. conduct a systematic review of machine learning and deep learning predictive analytics in healthcare, examining 41 studies published between 2019 and 2022.

This review supports the inclusion of predictive tools powered using AI into hospital practices, such as the prediction of the capacity of generating waste [16]. The machine learning model proposed by Haque et al. utilizes hospital electronic health record data to predict patient readmissions using algorithms such as Random Forest and XGBoost. The accuracy of XGBoost is better, thus demonstrating the machine learning abilities in predicting the developments in hospital operations. Although not the primary focus of the study, the approach used for predicting patient readmissions could likely be generalized to forecast hospital waste patterns by leveraging patient- and service-level datasets. Mustafa Alhanaqtah [17] compares three methods (regression, time-series ARIMA, and growth rate) to forecast the amount of waste in the Jordanian household. Regression analysis illustrates how input variables such as population and income influence waste volume. While ARIMA offers slightly better performance for short-term forecasting, regression is more effective in identifying cause-and-effect relationships. This makes it a potentially suitable approach for predicting hospital waste based on patient inflow and demographic data [18]. Jayadi et al. use linear regression with increments to predict and estimate the amount of waste produced based on the SIMASKOT data, which is a municipal waste management system. It also updates its models every year that an observation is available and thus improves accuracy over the long term. Compared with traditional regression, the method offers better performance and is especially suitable for the waste streams that are variable, like metals. Though the technique was based on urban data, it can be adopted to make predictions on hospital wastes so long as there is a regular updating of patient flow statistics [19].

The proposed research focuses on predictive analytics in the waste forecasting of hospitals; however, it only uses the daily patient inflow as the sole predictor. Unlike the existing studies, where an analysis is based on large amounts of healthcare data or complex machine learning models, this work isolates a single operational variable and applies linear regression for simplicity and interpretability. This makes the framework particularly practical to operate in data-limited settings. Even though it is simple, the model achieves high predictive accuracy, therefore showing that linear regression can be used as a convenient and scalable alternative for hospitals with limited computational capacity.

Data-Driven Models for Hospital and Municipal Waste Estimation:

Effective waste management is based on models that use empirical parameters like volume of patients served, size of facility, and usage patterns of the services. In the current review, issues regarding the studies utilizing the methods of regression and machine learning to forecast the volume of hospital and urban garbage, taking into consideration the real-life applicability, are summarized. Shakoor et al. examine waste management practices and occupational exposure at the Pakistan Institute of Medical Sciences in Islamabad. Although the study primarily focuses on occupational health, it also presents quantitative data on hazardous and non-hazardous waste volumes and conducts trend analysis to account for variations in waste loads. The case study highlights the importance of waste volumes in the functioning of the waste system and, in turn, the value of predictive methods in the process of optimising waste management [20]. Çetinkaya et al. proposed a regression-based model of forecasting medical waste volumes in the province of Aksaray, Turkey. The demographic determinants, such as the age groups of patients and the GDP per capita, were input variables with a determination coefficient (R^2) of 0.979. The paper concludes that accurate forecasting is essential for developing sustainable medical waste management plans and emphasizes that population structure and economic conditions are the primary factors influencing the volume of waste generated [21].

Arub et al., in their study, describe the analysis of the waste generation in 17 teaching hospitals located in Lahore, Pakistan, using linear regression to explore how the various hospital parameters influence waste volumes and identify variables that determine infectious waste and general waste. The findings demonstrate that the bed counts, the number of patients, and overall waste production have high correlations (R2 up to 0.9995 in the case of general waste). Surveys, hospital records, and field observations were used when collecting data. The results also demonstrate the effectiveness of simplistic regression methods of waste prediction, especially within urban environments with active population growth [22]. Altin et al. employed two other sophisticated modelling techniques, Kernel-Based Support Vector Machine (SVM) and Deep Learning, to predict medical waste in one of the Turkish private hospitals. Those input variables were inpatient admissions, surgery operations, and the duration of stay in the ICU. The accuracy of both models was tolerable, and Deep Learning was the best performer (R2 = 0.466). The research confirms that the multiple determinants can be fitted well to complex machine-learning algorithms rather than traditional regression models, and therefore, the model is appropriate in forecasting hospital waste [23]. Chien et al. introduced a new method called Seasonal Z-number Regression (SZR) to forecast hospital wastes in Taiwan. SZR combines fuzzy logic and dealing with uncertainty, seasonal analysis, and support vector regression. SZR outperformed neural networks and time series each time compared. This result indicates that SZR can be more stable in terms of waste prediction instrument, particularly in places where seasonal variation plays a major role in the generation of waste [24]. Konyalioglu et al. developed a hybrid computing system, which integrated a nonlinear grey Bernoulli model with a nature-inspired firefly algorithm-based optimizer in forecasting medical waste generation in Istanbul.

The framework outstripped classical regression methods and other similar models that recorded an error margin of less than 3.5%. The results support the effectiveness of hybrid modeling in capturing complex time-scale behaviors, such as post-COVID fluctuations in hospital waste flows, thereby demonstrating the potential of this approach for handling highly nonlinear predictive tasks [25]. Al-Omran and Khan presented a study of the creation of medical waste in the Bahraini hospitals using the monthly data of both the public and private hospitals. Detailed regression modeling by machine learning showed that the ensemble voting regressor is the most accurate predictor that explains over 90 % of the variance. Inpatient admissions, surgical procedures, and the resident population were the major factors that caused predictive performance. The findings prove the potential of machine-learning technologies to facilitate accurate prediction and planning of operations in the environments of public and commercial healthcare facilities [26]. Nopiah et al examined the factors that led to the production of medical waste in Malaysia through correlation and multiple regression analysis. They recorded strong linear correlations across the values of waste production and population, and gross domestic product (R > 093). Complementary regression modeling helps to clarify the role of every factor; the most important factors are population growth and economic activity. Such a framework with simple modifications (the most important of which is the replacement of population dynamics with patient inflow) can be easily adopted by hospital waste prediction [27].

The current research differs from the previous modeling of hospital wastes by the exclusion of complex demographic and infrastructural variables, and thus only focuses on the predictive variable of the inflow of patients. Existing works often rely on multi-variable regression methods, hybrid models, or sophisticated machine-learning frameworks to predict waste using economic indicators, regional characteristics, or facility-based data. In contrast, this paper emphasizes simplicity, using daily records of operations collected at a steady 60-day interval, and using this data to create a linear model, which can be scaled and easily updated. This makes the approach highly adaptable to smaller healthcare facilities or developing regions where detailed input variables might not exist. In addition, the study also fills a gap by offering a fully validated, statistically tested (ANOVA, residuals, RMSE) model based on localized, time-specific hospital data.

Table 1. Dataset from LRH Hospital, Peshawar

Author	Year	Model Type	Input Features	R^2 / Accuracy	Country
Rath et al. [15]	2020	Multiple Linear Regression	Confirmed, active, recovered cases	0.99	India
Çetinkaya et al. [21]	2020	Multiple Linear Regression	Age groups, GDP per capita	0.979	Turkey
Arub et al. [22]	2020	Linear Regression	Beds, patients, hospital parameters	0.9995	Pakistan
Badawy et al. [16]	2023	Systematic Review (ML/DL)	Clinical data, imaging, diagnostic features	Not specified	Egypt

Haque et al. [17]	2023	XGBoost, Random Forest	EHR: diagnosis, treatments, demographics	Highest with XGBoost	USA
Altin et al. [23]	2023	Deep Learning, SVM	Inpatients, surgeries, and ICU stay	0.466	Turkey
Nopiah et al. [27]	2023	Correlation & Multiple Regression	Population, GDP	r > 0.93	Malaysia
Alhanaqtah [18]	2024	Regression, ARIMA, Growth Rate	Population, income	ARIMA better short-term	Jordan
Jayadi et al. [19]	2024	Incremental Linear Regression	SIMASKOT urban waste data	Better than traditional regression	Indonesia
Shakoor et al. [20]	2024	Trend Analysis	Hazardous and non-hazardous waste volumes	Not specified	Pakistan
Chien et al. [24]	2024	Seasonal Z-number Regression	Fuzzy logic, seasonal data	Outperformed other models	Taiwan
Al-Omran and Khan [26]	2024	Ensemble Voting Regressor	Inpatients, surgeries, and population	>0.90	Bahrain
Konyalioglu et al. [25]	2025	Hybrid Grey Model + Firefly	Time-series waste data	<3.5% error	Turkey

Methodology:

Research Design and Approach:

The model employed in this research adopts a quantitative approach for the predictive analysis of hospital waste generation. A linear relationship was used to determine the relationship between patient inflow and the generation of waste per day by using a linear regression model. This method is chosen for its simplicity, ease of interpretation, and suitability for predicting continuous numerical outcomes. The work was organized according to the acquisition of real-world hospital data, model creation, training, validation, and subsequent testing. The new strategy also focused on the use of data to perform more efficient planning and operation of hospital waste through forecasting.

Data Source and Study Context:

The data for this study were collected from Lady Reading Hospital, where a series of observations was conducted over 60 consecutive days. The results are presented using two primary variables: the number of patients attended to per day and the corresponding amount of waste generated, measured in kilograms. The information was gathered by visiting the sites with the help of hospital workers and taking notes accordingly in trip logs. An in-depth visualization of this data set can be represented in Table 1, and the daily tendencies of the patient inflow and waste production are illustrated in Figure 3, with the confirmed positive tendency between the two. The study setting depicted a real-life in-hospital facility and warranted the usual fluctuation of the patient volumes and levels of waste, thus offering a reasonable ground for creating a realistic and workable predictive model.

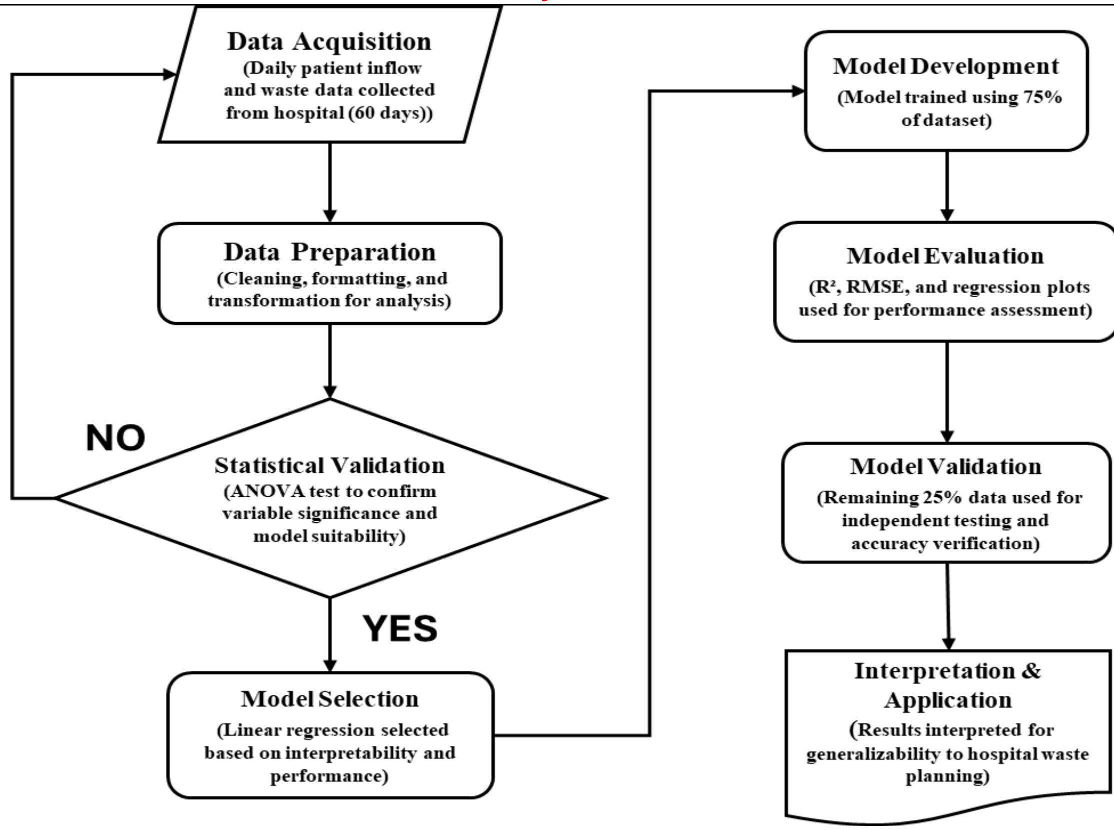

Figure 2. Research Flowchart

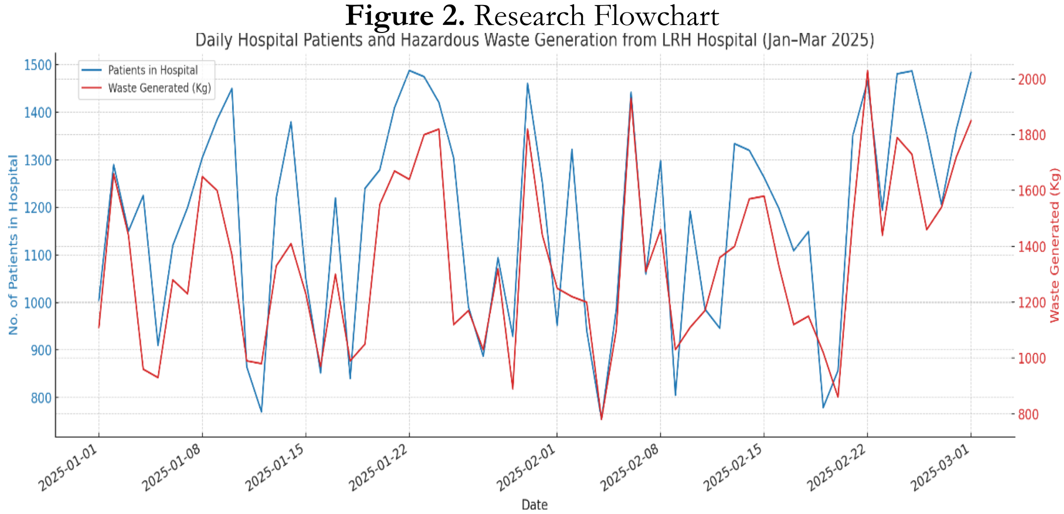

Figure 3. Daily trends of patient inflow and waste generation

Statistical Validation Using ANOVA:

An analysis of Variance (ANOVA) was conducted to ascertain the level of statistical validity in the correlation between the number of patients and the amount of waste generated. This aims to determine whether fluctuations in patient inflow significantly contribute to variations in waste generation. The test was taken on the entire 60-day data, and regression was carried out after this. It required breaking down the overall variance of waste between variances that were explained by the regression and those that were attributable to chance. Additionally, a lack-of-fit test is incorporated to assess whether a linear model adequately captures the trend in the data without introducing systematic errors. This step is crucial as it evaluates the robustness of the data and its suitability for use in predictive modeling.

Table 2. Dataset from LRH Hospital, Peshawar

Hospital	Waste Type	Date	No. of Patients	Waste Generated (Kg)	Date	No. of Patients	Waste Generated (Kg)	Date	No. of Patients	Waste Generated (Kg)
Lady Reading Hospital	Hazardous	01-Jan-25	1005	1110	21-Jan-25	1488	1670	09-Feb-25	805	1030
		02-Jan-25	1290	1660	22-Jan-25	1475	1840	10-Feb-25	1192	1340
		03-Jan-25	1150	1440	23-Jan-25	1412	1820	11-Feb-25	986	1170
		04-Jan-25	1225	1360	24-Jan-25	941	1120	12-Feb-25	946	1360
		05-Jan-25	910	960	25-Jan-25	1301	1540	13-Feb-25	1334	1400
		06-Jan-25	1120	1280	26-Jan-25	991	1170	14-Feb-25	1320	1570
		07-Jan-25	1010	1230	27-Jan-25	887	1030	15-Feb-25	1263	1580
		08-Jan-25	1200	1650	28-Jan-25	1094	1320	16-Feb-25	1198	1550
		09-Jan-25	1385	1600	29-Jan-25	929	890	17-Feb-25	1109	1120
		10-Jan-25	1450	1700	30-Jan-25	1461	1820	18-Feb-25	1449	1550
		11-Jan-25	865	980	31-Jan-25	1252	1440	19-Feb-25	779	1020
		12-Jan-25	770	930	01-Feb-25	952	1250	20-Feb-25	857	860
		13-Jan-25	1310	1630	02-Feb-25	1322	1620	21-Feb-25	1351	1530
		14-Jan-25	840	920	03-Feb-25	941	1200	22-Feb-25	1466	2030
		15-Jan-25	880	890	04-Feb-25	754	780	23-Feb-25	1194	1440
		16-Jan-25	1052	1340	05-Feb-25	986	1100	24-Feb-25	1187	1430

		17-Jan-25	1220	1360	06-Feb-25	1442	1990	25-Feb-25	1481	1790
		18-Jan-25	840	1050	07-Feb-25	1060	1310	26-Feb-25	1355	1460
		19-Jan-25	1279	1550	08-Feb-25	1298	1460	27-Feb-25	1206	1540
		20-Jan-25	1410	1650	09-Feb-25	805	1030	28-Feb-25	1364	1720

Table 5. Comparative analysis against existing studies

Author	Year	Model Type	Input Features	R^2 Score	Root Mean Square Error (RMSE)	Context
Arub et al.	2020	Linear Regression	Beds, patient visits, hospital capacity	Up to 0.9995	Not reported	Teaching Hospitals, Pakistan
Çetinkaya et al.	2020	Multiple Linear Regression	Age group distribution, GDP per capita	0.979	Not reported	Medical Waste, Turkey
Altin et al.	2023	Deep Learning, SVM	Inpatients, surgeries, and ICU stays	0.466 (Deep Learning)	Not reported	Hospital, Turkey
Haque et al.	2023	XGBoost, Random Forest	EHR data: diagnosis, treatments, demographics	Highest with XGBoost	Not reported	Hospital Readmissions, USA
Al-Omran & Khan	2024	Ensemble Voting Regressor	Inpatients, population, surgeries	>0.90	Not reported	Public/Private Hospitals, Bahrain
Chien et al.	2024	Seasonal Z-number Regression	Seasonal patterns, fuzzy logic, and hospital waste	Outperformed other models	Not reported	Hospital, Taiwan
Konyalıoğlu et al.	2024	Grey Bernoulli + Firefly Algorithm	Time series waste data	Error < 3.5% (high accuracy)	Not reported	Medical Waste, Istanbul
Proposed Study	2025	Linear Regression	Patient inflow	0.88 (train), 0.81 (Val)	130.52 kg (Val)	Hospital, Pakistan

Unlike more complex models such as Support Vector Regression or Neural Networks, linear regression is easy to train, requires minimal computational resources, and quickly adapts to the 60-day dataset, making it suitable for real-time implementation in hospital systems. This simplicity and efficiency make it particularly effective for operational decision-making in healthcare settings [12]. In this study, the number of patients served per day is treated as the independent variable, while the resulting amount of hospital waste generated (measured in kilograms) is considered the dependent variable. The overall linear relationship between hospital activity and waste generation has been established in previous studies, further confirming the suitability of this algorithm for the current analysis [13].

Computational Efficiency of the Selected Model:

One of the key advantages of using linear regression in this study is its low computational demand. The model efficiently trains on the 60-day dataset without requiring specialized hardware or extended processing time. This efficiency led to its application in real-time or resource-constrained areas like the hospital, where speed and decipherable outcome are a necessity. As opposed to more complicated machine learning techniques such as Support Vector Machines or Neural Networks, linear regression offered good results based on much lower computational demands, which made it simple to integrate into hospital management systems without sacrificing much precision.

Data Acquisition and Preprocessing:

Patient and Waste Volume Data Collection:

The structured observational method was applied to collect data. The amount of waste generated per day in kilograms was documented every day regarding the number of patients coming in on that day over a duration of 60 days. The hospital records had the information entered manually, and subsequently, informative tables were prepared by digitizing the same. Every entry explained a single day of operations in the hospital and contained two main characteristics: the number of patients and the amount of waste produced in total.

Data Cleaning and Transformation Procedures:

The consistency and accuracy of the dataset were maintained by cleaning it before developing a model. To eliminate the distortion of null values, all rows containing the values of this nature were eliminated. Duplicates were reviewed and removed, column names were standardized, string operations were performed, and numeric fields were checked to ensure they were correctly typed. Outliers were identified and handled during data cleaning before model training to ensure robust regression results. The new index feature, too, was introduced to preserve the time order of the data points. The model employed a single-variable approach, where the input factor (x) represented the number of patients seen per day, and the output variable (y) corresponded to the total waste generated, measured in kilograms. While the model could be made more complex by incorporating additional variables such as department-wise or categorized waste, the chosen approach remained simpler and more generalizable to the entire hospital setting [28].

Predictive Modeling Framework:

Justification for Linear Regression Selection:

Linear regression was selected as the modeling method due to its ease of interpretation, low computational cost, and its effectiveness in defining linear relationships between continuous variables. The patient number was an independent variable, and the amount of waste produced per day was taken as a dependent variable in this study. The approach was also adopted in instances of cases involving speedy, corroborative planning predictions in operations. The modeling process is presented in Figure 4, and the main parts in the process are data input, training the models, performance evaluation, validation, and output of making predictions, which is in a structured and repeatable process.

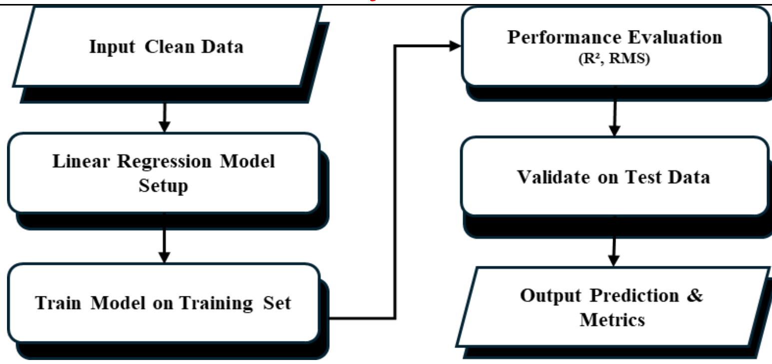

Figure 4. Predictive Modelling Flowchart

Model Formulation and Mathematical Representation:

The predictive model applied in the present study was an ordinary linear regression equation referring to a linear prediction between an independent variable and a dependent variable:

$$\mathbf{Y = mx + c}$$

Where:

Y is our forecast for the day-to-day waste, in kilograms.

x is several patients reported on a particular date.

m is the rate of regression or slope of the regression line, and it is a measurement of the average increase in waste per extra patient.

C is the interception, which is the approximate amount of waste generated even in the absence of a patient, which is a result of constant, or background, waste related to hospital activities.

Historical data pairs of patient inflow and corresponding waste quantity were used to train the model by calculating the optimal values of *m* (slope) and *c* (intercept). These values were applied by minimizing the error of predictions by the least squares method, which provides the line of best fit by minimizing the sum of squared differences between actual and predicted values. The formulation supported the continuity of the forecast and planning purposes since it enabled the model to generalize the trend of the waste output per unit of volume of patients.

Computational Tools and Libraries Utilized:

The Python programming language was used in applying the modeling process. Important libraries were Pandas to deal with data, NumPy to do numerical tasks, Matplotlib to visualize data, and Scikit-learn to create a linear regression model. The development environment used was Jupyter Notebook, which allowed for step-by-step analysis, plotting, and evaluation of the models in a clear and visually organized format. 3

Model Training and Evaluation:

Training Dataset Configuration:

The Table was then split into 2 parts: 75% of the data (i.e., 45 out of 60 days) was to be used to train the model, whereas the remaining 25% (15 days) was considered for validation purposes. This led to model training such that it could recognize some underlying patterns in the training data, and the test was done on a blinded portion, such that there was no bias during evaluation. The application of the hold-out validation methodology was useful in determining how generalizable and robust the predictive model can be in settings of realistic forecasting.

Visualization and Regression Line Plotting:

Once the model had been fitted, the scatter plot was created to observe the relationship between the patient inflow and waste output. The plot was used to determine the linearity of the relationship and the scatter of points on the plot; a regression line was also overlaid to verify the model fit visually. The plot facilitated in explaining the interpretability of the plot as to whether the waste generation improved with the number of patients, which implies the support of the linear model.

Performance Assessment Metrics (R^2, RMSE):

Coefficient of determination (R^2) and Root mean square error (RMSE) were used to check the performance of the model. The R^2 value described the extent of explaining the variance in waste generation by the variable of patient volume, whereas the value of RMSE described the mean error of prediction in kilograms. The result of the training was calculated as $R^2 = 0.88$, which indicates a good fit of the variables. RMSE was also computed to determine the extent to which predictions differed from actual values.

Model Validation and Testing:

Validation Dataset Setup:

The other 15 days of data were utilized as a validation set to measure the model's generalizability. This dataset was not involved in the training process and served as an objective means of evaluating the performance of the models. The validation data set contained unseen values of patient inflows and actual values of the corresponding waste quantity.

Statistical Accuracy and Error Analysis:

The model was used here to generate predictions on a validation set and compared with actual waste values. The score given during the validation R^2 was 0.81, meaning the model had a strong relationship with prediction. The RMSE on the validation was also close to the training, meaning that the model had a slight overfit with small variability in the model's behavior on additional data.

Interpretation of Results and Generalizability:

The test findings indicated that the linear regression model held the potential to predict the daily burden of hospital waste daily depending on the number of patients admitted. The model proved to have high predictive power and was accurate in most instances; therefore, it makes it appropriate to be part of the healthcare planning systems.

The fact that it could be generalized to related situations in hospitals promoted its possible adoption in the practical process of waste management.

Results and Discussion:

Statistical Validation Using ANOVA:

To assess the statistical significance of the correlation between daily patient inflow and the amount of waste disposed of in the hospital, an extension of the analysis of variance (ANOVA) was conducted. The findings indicated an F-value of 384.13 and a P-value of 0.000, representing a strong connection linking the two variables. This affirms the fact that the changes in the generation of waste largely depend on changes in the number of patients. Moreover, the lack-of-fit test recalculates a P-value = 0.219 that exceeds the required value of P = 0.05 level, denoting that the linear modeling is appropriate to the observed data without a significant systematic failure. The relatively low pure error confirms the consistency and credibility of the dataset. Such results confirm that the set of data can be used with regression modeling, and it makes a good background for predictive analysis.

Table 3. ANOVA Test for Statistical Significance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	4825009	4825009	384.13	0.000
Number of Patients in Hospital	1	4825009	4825009	384.13	0.000
Error	58	728524	12561		
Lack-of-Fit	55	714424	12990	2.76	0.219
Pure Error	3	14100	4700		
Total	59	5553533			

Residual Analysis:

Residual analysis was done to validate that the parameters used in the linear regression model fulfilled major assumptions for valid inference. Normal probability plot revealed the fact that residuals were equally distributed on a straight line; thus, the errors were independently distributed according to a normal distribution. The plot of residuals against fitted value demonstrated no pattern and a random scatter, which indicates the same variance under all the levels of predicted variables. The plot of the histogram of residuals had an approximately symmetrical bell-shaped pattern, which further confirmed the assumption of normality. The residuals by the order of observations graph also showed no trend pattern of time-based data, validating the non-occurrence of autocorrelation or temporal bias. All these outcomes prove that the regression model meets the assumptions related to linearity, homoscedasticity, normality, and independence, which further contributes to the reliability of this model as the predictor.

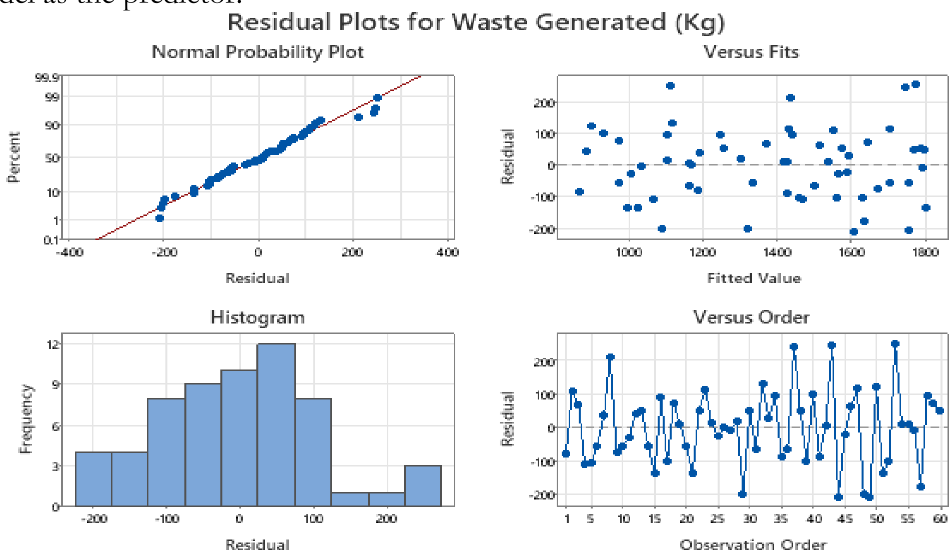

Figure 5. Residual Plots for valid inference

Model Training Results:

The training data consisted of 45 days of data (data used to fit the model) with the inflow of patients during a day as an independent variable, and the volume of hospital waste generation (in kilograms) as a dependent variable. The resulting regression equation was (y = 1.27x - 93.06), where y took the value of the predicted waste and x was the capacity of the patients in the hospital. The resulting R^2 of 0.88 is an indication that 88% of the variance in the waste generated can be attributed to patient inflow during the training stage.

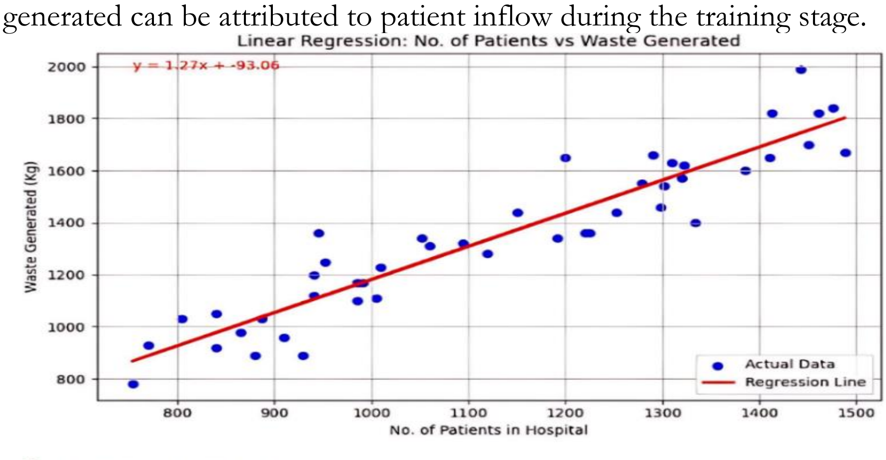

Figure 6. Model Training Output

Figure 6 depicts the scatter plot that illustrates an excellent linear relationship since the actual data (blue dots) and the regression line (red) are illustrated. The plotted data fall within a range of approximately 800 to 1500 patients per day, with corresponding waste generation values ranging from around 850 to 1950 kilograms. Most of the points are rather close to the regression line, which proves the minimal residual error and the high model compliance with the data trend. For instance, one of the data points indicates nearly 1450 patients, implying nearly 1900 kg of waste, which is quite a close figure to the modeled prediction. In the same way, the predicted and actual value of waste of 900 patients is nearly 1050 kg, thus indicating that the accuracy of the model is consistent over the observed range. The graphic distribution of the data points, along with the regression line and the large value of $R^{2,}$ demonstrates that the fitted model can reflect the actual relationship between the inflow of patients to a hospital and the amount of waste produced by it. These findings qualify the robustness of the model and confirm its usability in the forecasts of operations.

Model Validation Results:

The generalizability of the model was evaluated using the remaining 15 days of data, which served as the validation set. Using the same trained regression equation $y=1.27x-93.06y$, the model was tested on the unseen validation dataset. The validations were done within the R^2 of 0.81, meaning that 81 per cent of the variance in the hospital waste generation was explained by the model. The average prediction difference between physical and forecasted waste was calculated to be 130.52 kg, which is equivalent to the Root Mean Square Error (RMSE) of the model. Table 3 shows the summary of the model performance over all training and validation phases.

Table 4. Summary of regression model performance during training and validation phases.

Metric	Value
Regression Equation	$y = 1.27x - 93.06$
R^2 Score (Training)	0.88
R^2 Score (Validation)	0.81
RMSE (Validation) [kg]	130.52

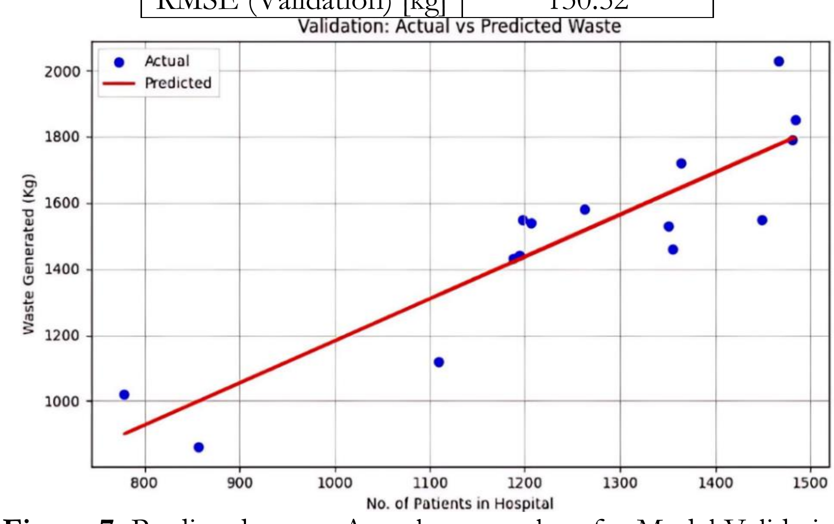

Figure 7. Predicted versus Actual waste values for Model Validation

Figure 7 is the scatter diagram of the predicted and actual values of waste in the validation set. The blue points are real-life waste data, whereas the red line is a regression line that shows the predicted outputs. Even though there is a slight deviation between the prediction and the plotted points, the latter lies close to the former throughout the range of input, about 780 kg to 1480 kg, as well as in the range of waste, about 900 kg to more than 2000 kg. An example of this is that the model produces an accurate forecast of 1450-1500 kg of waste when the available patients exceed 1300, and thus, the application is highly consistent even during days of large capacity. The fact that the overall clustering of the line confirms that the model will have valuable predictive accuracy when implemented on new data that has not been previously seen.

Discussion of Findings:

The findings in this study corroborate the fact that linear regression can be a safe method and a practical way for predicting hospital waste in terms of patient inflow. During training, the model achieved an R^2 score of 0.88, indicating a strong fit. In the validation phase, the R^2 score was 0.81, and the Root Mean Square Error (RMSE) was calculated to be 130.52 kg, reflecting a reasonably accurate predictive performance on unseen data. These metrics illustrated that the model managed to fit in with the real relationship between the patient volume and the size of the waste; therefore, it proved suitability for the short-term operational forecasting in health care environments.

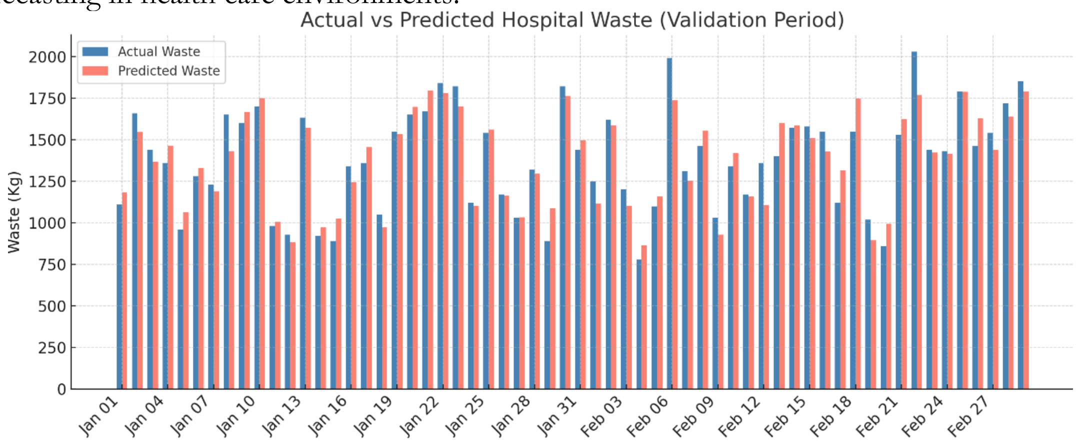

Figure 8. Comparison of actual versus predicted hospital waste using linear regression

Figure 8 depicts the prediction-actual correlation of waste production throughout the 60-observation period. Figure 8 indicates that there is little fluctuation in the number on most days, even when the flow of patients fluctuates. Increase in quantities of waste above 1900 kg in data collected on higher patient load days, January 2 and January 3. The predicted trend closely aligned with the observed waste quantities, and the model's stability demonstrated its robustness across varying operational conditions. Additional support on these findings comes along with ANOVA and residual diagnostics tests, which confirmed the validity and reliability of the regression model obtained. Nonetheless, the model is built on the strictly linear relationship between waste and patient inflow. In practice, during hospital operations, nonlinear patterns can be caused by other factors like seasonal variation, the nature of medical services, or a change in the activity of a department. In these circumstances, linear models might become more inaccurate, meaning that even more versatile methods can be tested in the future. However, linear regression is very appropriate in less resourceful situations, such as hospitals, where fast decision-making is of the essence since it is very simple and cheap in terms of computation. To contextualize the model's performance, Table 4 provides a comparative analysis against existing studies in the domain of hospital and medical waste forecasting:

Although the past research findings, such as those of Arub et al. and Çetinkaya et al., have yielded a higher regression R^2 coefficient, they relied on multiple input variables, which are subjected to complicated data-collection processes. In contrast, the proposed model achieved competitive accuracy using only a single, easily obtainable input variable, including daily patient count. Its minimalistic design enhances the model's practicality and makes it highly adaptable to hospitals that have limited access to large amounts of operating data or advanced computational tools. Overall, findings validate the effectiveness of linear regression in the context of hospital waste prediction and establish the proposed methodology as a computationally effective and applicable in practice method usable in real healthcare settings.

Conclusion:

This study presents a predictive modeling framework for estimating hospital waste generation based on daily patient inflow, utilizing a linear regression approach. The model was constructed and tested against 60 days of real-time operational information that came along with a tertiary care hospital. Assessment indicators showed great accuracy with an R^2 value of 0.88 in training and 0.81 in validation, as well as a Root Mean Square Error (RMSE) of 130.52 kg. The fact that the inflow of patients is a significant predictor and that the measure of residual indicated that the regression assumptions were accepted was confirmed with the use of ANOVA and residual analysis. The findings indicate the stability and replicability of the model in terms of application in the estimation of healthcare wastes in the real world.

These findings support the fact that there is a clear relationship between patient volume and the generation of waste in hospital settings, which can be used to plan and prevent or intervene properly. The solution is both transparent and computationally fast and can be applied with the lowest technical cost-of-addition, so the model can be applied in typical decision-making and the optimization of operations. Pre-determining the quantity of waste by advising hospitals on potential waste will help in the advanced planning of traffic of waste collection, enhancement of hygiene conditions, and mitigation of danger to personnel and patients. In addition, it also increases the capability of healthcare managers to coordinate their operations with the regulations on waste management and the prevention of infection.

Besides its operational advantages, the given system is likely to contribute to the main objectives covered by the Sustainable Development Goals (SDGs), especially SDG 3 (Good Health and Well-being) and SDG 12 (Responsible Consumption and Production). The model allows creating safer environments in the hospital, limits the exposure to hazardous waste, and allows us to use the resources more efficiently. The fact that it is also easily integrated into hospital information systems and capable of adaptation to different sizes of facilities enhances its further potential for being used more widely. Overall, the study contributes a practical and scalable tool for predictive waste management in healthcare, offering a foundation for more advanced data-driven interventions in the future.

Future Work:

As the current study has shown the efficiency of the linear regression in predicting the quantity of hospital waste in patients' inflow, there are still a couple of improvement options. It can be investigated in future studies to use more advanced algorithms of machine learning, like Support Vector Regression, Random Forest, or Gradient Boosting, which may result in higher accuracy of prediction, especially in the case of non-linear or complex operational scenarios. These strategies can put better focus on the secret structures of the information and will improve the desirability of the model within different inpatient settings.

Further research can also include the involvement of more predictive characteristics than the number of patients (departmental inflow, seasonal variations, or occupancy levels, the day-of-the-week effect, etc.). This would result in more detailed and local forecasts of waste. This would also enable the hierarchy of dynamic update and continuous learning models with integration into real-time hospital management systems that could enable the hospitals to adjust waste planning daily. Such extensions would further enhance the applicability of the model to decision-makers and enhance its importance in operation planning and sustainable management of healthcare waste.

References:

[1] W. A. Rutala and C. G. Mayhall, "Medical Waste," *Infect. Control Hosp. Epidemiol.*, vol. 13, no. 1, pp. 38–48, Jan. 1992, doi: 10.2307/30146966.

[2] A. A. S Altin, "Determination of Hospital Waste Composition and Disposal Methods:

a Case Study," *Polish J. Environ. Stud.*, vol. 12, no. 2, pp. 251–255, 2003, [Online]. Available: https://www.pjoes.com/Determination-of-Hospital-Waste-Composition-r-nand-Disposal-Methods-a-Case-Study,87553,0,2.html

[3] World Health Organization, "Working together for Health: World Health Report 2006," *World Health*, vol. 19, no. 3, p. 237, 2006.

[4] P. D. Himani S Bansod, "Biomedical Waste Management and Its Importance: A Systematic Review," *Cureus*, vol. 15, no. 2, p. 2, 2023, [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9981497/

[5] M. Tsakona, E. Anagnostopoulou, and E. Gidarakos, "Hospital waste management and toxicity evaluation: A case study," *Waste Manag.*, vol. 27, no. 7, pp. 912–920, 2007, doi: https://doi.org/10.1016/j.wasman.2006.04.019.

[6] B. A. Z. Alagöz and G. Kocasoy, "Treatment and disposal alternatives for health-care waste in developing countries - A case study in Istanbul, Turkey," *Waste Manag. Res.*, vol. 25, no. 1, pp. 83–89, Feb. 2007, doi: 10.1177/0734242X07069497.

[7] W. H. O. and U. N. C. F. (UNICEF), "Water, sanitation and hygiene in health care facilities Status in low- and middle-income countries and way forward," *Geneva World Heal. Organ.*, 2015, [Online]. Available: https://iris.who.int/bitstream/handle/10665/154588/9789241508476_eng.pdf

[8] K. I. Enkhtsetseg Shinee, Enkhjargal Gombojav, Akio Nishimura, Nobuyuki Hamajima, "Healthcare waste management in the capital city of Mongolia," *Waste Manag.*, vol. 28, no. 2, pp. 435–441, 2008, doi: https://doi.org/10.1016/j.wasman.2006.12.022.

[9] W. C. B. Jr., "Basic Hazardous Waste Management," *Taylor Fr.*, p. 488, 2001, doi: https://doi.org/10.1201/9781420032604.

[10] H. Chandra, "Hospital waste: An environmental hazard and its management," *C Newsl. ISEB India*, vol. 5, no. 3, pp. 80–5, 1999, [Online]. Available: https://isebindia.com/95_99/99-07-2.html

[11] C. Fluke, "Handling hazardous waste," *J. Heal. Mater. Manag.*, vol. 6, no. 2, pp. 70–73, 1988.

[12] S. K. H. and S. Shahab, "Hospital and biomedical waste management," *Community Med. Public Heal. 4th Ed Karachi Time Publ*, pp. 426–437, 2003.

[13] Y. H. and M. Shigemitsu, "Proper disposal (management) of medical wastes–infection prevention and waste management (Clean Hospital Project) at Hiroshima City, Asa Hospita," *Rinsho Byori*, pp. 26–31, 2000, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/10901042/

[14] S. Ahmed, "Hospital waste sickening public and environment," *Dly. Times*, vol. 7, 2004.

[15] A. R. T. Smita Rath, Alakananda Tripathy, "Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model," *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 5, pp. 1467–1474, 2020, doi: https://doi.org/10.1016/j.dsx.2020.07.045.

[16] N. R. & H. A. H. Mohammed Badawy, "Healthcare predictive analytics using machine learning and deep learning techniques: a survey," *J. Electr. Syst. Inf. Technol.*, vol. 10, no. 40, 2023, doi: https://doi.org/10.1186/s43067-023-00108-y.

[17] M. M. Haque, "Advancing Healthcare Outcomes with AI: Predicting Hospital Readmissions in the USA," *J. Med. Heal. Stud*, vol. 4, no. 5, pp. 94–109, 2023.

[18] O. J. Alhanaqtah, "Forecasting of the Waste Generation in Jordan: Alternative Econometric Approaches," *Int. J. Econ. Financ.*, vol. 16, no. 2, p. 12, 2023, [Online]. Available: https://ccsenet.org/journal/index.php/ijef/article/view/0/49651

[19] N. R. H. Puguh Jayadi, "Forecasting Waste Generation with Increment Linear Regression Technique: A Case Study of SIMASKOT Application," *J. Comput. Sci. Adv.*,

vol. 2, no. 5, 2024, [Online]. Available: https://journal.ypidathu.or.id/index.php/jcsa/article/view/1369

[20] M. A. Abdul Shakoor, "Quantitative and Qualitative Assessment of Hospital Waste Management at Capital's Health Care Facility, Pakistan," *Bull. Bus. Econ.*, vol. 13, no. 1, p. 3, 2024, [Online]. Available: https://bbejournal.com/BBE/article/view/679

[21] A. Y. Çetinkaya, S. L. Kuzu, and A. Demir, "Medical waste management in a mid-populated Turkish city and development of medical waste prediction model," *Environ. Dev. Sustain.*, vol. 22, no. 7, pp. 6233–6244, Oct. 2020, doi: 10.1007/S10668-019-00474-6/METRICS.

[22] S. R. A. Samita Arub, "Assessment of Waste Generation Rate in Teaching Hospitals of Metropolitan City of Pakistan," *Civ. Eng. J.*, vol. 6, no. 9, pp. 1809–1821, 2020, doi: 10.28991/cej-2020-03091584.

[23] F. Ö. Fatma Gül Altin, İbrahim Budak, "Predicting the amount of medical waste using kernel-based SVM and deep learning methods for a private hospital in Turkey," *Sustain. Chem. Pharm.*, vol. 33, p. 101060, 2023, doi: https://doi.org/10.1016/j.scp.2023.101060.

[24] S.-Y. C. Hsing-Chin Chien, Ting-Yu Lin, Kuo-Ping Lin, Trang Cam Hoang, Hoa Anh Tran, Hsiang-Yun Wang, Hui-Mei Huang, "Developing seasonal z-number regression for waste-disposal forecasting in a Taiwanese hospital," *Appl. Soft Comput.*, vol. 152, p. 111196, 2024, doi: https://doi.org/10.1016/j.asoc.2023.111196.

[25] T. O. Aziz Kemal Konyalıoğlu, "Forecasting medical waste in Istanbul using a novel nonlinear grey Bernoulli model optimized by firefly algorithm," *Waste Manag. Res.*, vol. 43, no. 5, p. 5, 2025, doi: https://doi.org/10.1177/0734242X241271065.

[26] A. A. Khadija Al-Omran, "Integrated Environmental Assessment of Medical Waste Management in the Kingdom of Bahrain," *Sustainability*, vol. 15, no. 3, p. 2397, 2023, doi: https://doi.org/10.3390/su15032397.

[27] and C. J. K. Z. M. Nopiah, F. Zulkipli, N. E. A. Basri, "A Study of Waste Generation Prediction using Correlation and Multiple Regression Analyses," *Int. J. Ind. Eng. Technol. Oper. Manag*, vol. 1, no. 2, pp. 54–57, 2023.

[28] O. R. Kashif Hameed, "Types of Hospital Waste and Waste Generation Rate in Different Hospitals of Faisalabad City, Pakistan," *J. Basic Appl. Sci.*, vol. 13, no. 1, 2017, [Online]. Available: https://setpublisher.com/index.php/jbas/article/view/1679