

Article

Multi-layer perceptron artificial neural network for environmental risks prediction of SW-induced pollution in Dar es Salaam, Tanzania

Emmanuel KazuvaDepartment of Geography, Tourism and Hospitality, The Open University of Tanzania, Dar es Salaam, Box 23409, Tanzania;
ekazuvae@gmail.com

CITATION

Kazuva E. Multi-layer perceptron artificial neural network for environmental risks prediction of SW-induced pollution in Dar es Salaam, Tanzania. *Advances in Analytic Science*. 2024; 5(2): 3018. <https://doi.org/10.54517/aas.v5i2.3018>

ARTICLE INFO

Received: 22 October 2024
Accepted: 3 December 2024
Available online: 11 December 2024

COPYRIGHT

Copyright © 2024 by author(s). *Advances in Analytic Science* is published by Asia Pacific Academy of Science Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license. <https://creativecommons.org/licenses/by/4.0/>

Abstract: Many metropolitan areas face significant environmental challenges posed by improper disposal and management of solid waste. As a result, environmental risks have emerged as a pressing concern, prompting dedicated research efforts. This study on environmental risk prediction of Dar es Salaam SW coincides with a mounting governmental effort over rising pollution levels from inadequate SW management. Using the multi-layer perceptron artificial neural network (MLP-ANN) model, it effectively examines the prevailing conditions and forecasts waste generation rates (WGRs) and environmental risk index (ERI) associated with SW pollution. As confirmed with 94.5% prediction accuracy and 86.5% success rate of the MLP-ANN model, WGRs in Dar es Salaam have doubled in less than two decades. Besides, over 40% of the overall generated SW is left unattended. Consequently, the ERI exhibits a consistent upward trajectory throughout the assessment period, with intermittent fluctuations between Level II and III but a persistent overall increase. Projections indicate an escalation of ERI to Level IV by 2025/26 and to a critical threshold (Level V) by 2038. The key indices such as pressure, state, and impact are anticipated to reach critical thresholds ahead of the comprehensive ERI. This underscores the imperative for timely interventions and the urgency of addressing SW management issues to curb the escalating environmental risks in Dar es Salaam and other metropolises with similar challenges.

Keywords: waste generation rates; environmental risks; artificial neural network; predictive modelling; Dar es Salaam

1. Introduction

Solid waste-induced pollution is a significant environmental challenge in many metropolitan areas, including Dar es Salaam, which has a population of over six million and a rapidly growing economy [1–3]. The city generates large amounts of solid waste (SW) that are often disposed of in environmentally unfriendly ways [3–5], leading to serious public health and environmental consequences. Policymakers and researchers face major challenges in addressing this issue [2,6–8].

In recent years, there has been growing interest in using artificial intelligence (AI), specifically artificial neural networks (ANN), to predict and aid the reduction of risks associated with environmental pollution from mishandled SW [9]. ANNs are machine learning algorithms modelled on the human brain, enabling them to process complex environmental data and make accurate predictions about future pollution levels [10]. The multi-layer perceptron (MLP) is a type of ANN that has been successfully used in predicting air, water, and land pollution [11–14]. It is recommended for time series problems due to its stochastic input-output matching capabilities. In this study, the MLP-ANN was used to predict annual generation rates of six waste streams (organic-food-based, plastic, organic-paper-based, glass, metal,

and ‘others’) from four major sources: Residential, industrial, commercial, and institutional and services wastes. A linear regression model was also used to measure the environmental risk index and predict future environmental risk levels in Dar es Salaam over the next 20 years.

Environmental risk assessment aims to provide systematic procedures for predicting potential risks from environmental management scenarios, aiding decision-makers in implementing policies to reduce environmental impacts [15], mitigate greenhouse emissions, and promote sustainability [16]. In this regard, environmental risk prediction (ERP) is a step in ERA aiding risk management. It uses specific risk factors and models to estimate the likelihood of adverse outcomes due to changes in environmental conditions from both, natural and anthropogenic forces [17]. For the prediction model to be reliable it requires adequate discrimination, calibration, face validity, and environmental usefulness [18–20]. Thus, a basic understanding of the models is vital before applying them in environmental practice.

The study by Abbasi and Hanandeh [21] identified five categories of modes/methods for the forecasting of SW situations: Descriptive statistical methods, regression analysis, time series analysis, material flow model, and AI models. Other studies have used non-linear multivariate regression analysis to identify dependent parameters for predictor variables [22,23]. Robust AI algorithms, considered state-of-the-art, offer accurate and reliable predictions of waste generation and ERI due to their flexibility and non-linear, non-parametric structure [24]. MLPs provide a promising approach to predicting risks associated with waste disposal and identifying effective mitigation strategies.

This study aims to develop an MLP-based model to predict environmental risks from solid waste-induced pollution in Dar es Salaam. The model will be trained using historical data on waste generation, disposal, environmental quality, and demographic and economic indicators. By analyzing this data, the MLP identify patterns and relationships between environmental factors, enhancing the understanding of pollution drivers in the city and serving as a tool for predicting and mitigating risks associated with solid waste-induced pollution.

2. Material and methods

2.1. Geographical description of the study area

The study was conducted in Dar es Salaam, Tanzania’s largest and main commercial city. As shown in **Figure 1**, this region lies between latitudes 6°35’–7°10’ S and longitudes 39°02’–39°30’ E. Being a coastal area, much of its eastern region bordering the Indian Ocean consists of lowlands with elevations ranging from –3 m below sea level to 268 m above sea level (**Figure 1**). Dar es Salaam covers a total area of 1800 km², encompassing both water and land mass. Its population has grown significantly from 0.8 million in 1978 to over 5.4 million in 2022. According to World Population Estimates (WPE) for 2024, the population of this city is projected to reach 8.2 million people, reflecting a 30% increase from 2022 data [25]. The region experiences an average annual temperature of 30 °C and a humidity level of 80%. It has two rainy seasons: One from October to December and the other from March to May. Although the city is not prone to natural weather hazards, floods which are

exacerbated by inadequate storm drains and blocked drainage systems caused by improperly disposed waste are common in this city [3,26].

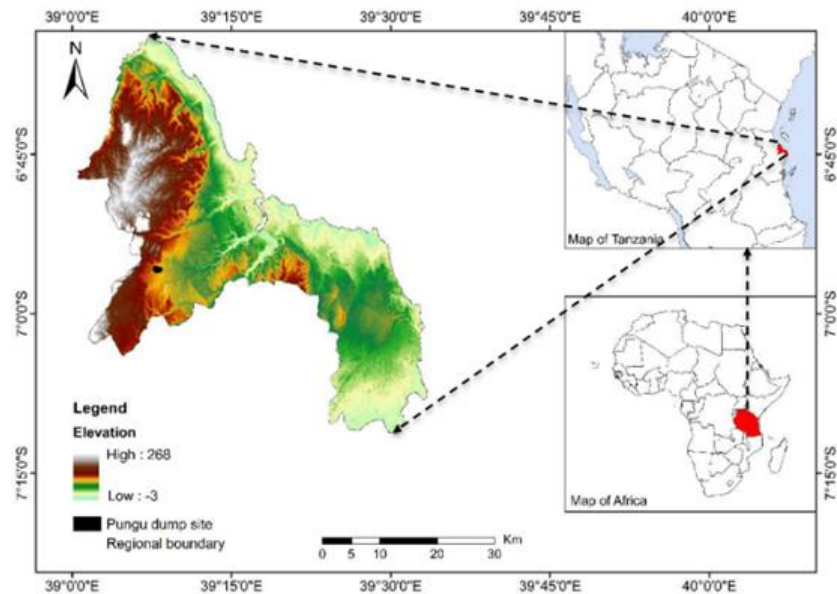


Figure 1. Study area map.

2.2. Computing waste generation rates and environmental risk index

The major aspects of this section were to determine waste generation rates (WGR) [2,5], Environmental Risk Index (ERI) [15,17,18] and subsequent environmental risk level (ERL) [7,8,16] which is the essential data source for WGR and ERI predictions. For WGR, the primary data on actual waste collection from 2006 to 2022 were obtained from the sampled sites using procedures similar to those previously applied and well-documented in the author's previously published report [2]. This aided the determination and drawing of inferences of average per capita generation rates in the city for 17 years, somewhat is crucial as data input to WGR predictive models.

On the other hand, to compute ERI, the driving force-pressure-state-impact-response, making a DPSIR model. The central idea of the DPSIR model is that socioeconomic development is commonly a driving force (D) that exerts pressure (P) which changes the natural beauty and state (S) of the environment. From the changed environmental state, the impacts (I) on the ecological environment and human health are experienced. The impacts provoke society's responses (R) through preventive, adoptive, or curative measures [27]. This model is considered a valuable tool for reporting and addressing complex environmental issues particularly those related to human activities and has been used by several national and international organizations, including the United Nations (UN)¹ in various environmental assessment and management initiatives, and by the United States Environmental Protection Agency [28] in the sustainable Puerto Rico initiative [29,30]. Thus, the DPSIR model has proved to be an effective tool for developing indicators and reporting the state and consequences of environmental degradation in urban planning, environmental impact assessment (EIA), land resource evaluation, health, wetland ecosystem evaluation, and other related environmental assessment [31–35]. In this study, the DPSIR model and

other supporting computation software including the analytical hierarchy process (AHP) and the expert questionnaire method (EQM)-Delphi technique were adopted in similar procedures as previously used and well vindicated in own published report [2]. Likewise, these techniques were used to organize risk indicator systems and quantify them in a hierarchical model to identify ERI which was a crucial procedure and served as data input to ERI predictive models.

2.3. Waste generation rate and environmental risk index prediction

The multi-layer perceptron (MLP) [36] which is a class of artificial neuron network (ANN) [11] is recommended among the best models that are applied to predict solutions to time series problems due to their stochastic input-output matching capabilities. In the current study, the multi-layer perceptron artificial neural network (MLP-ANN) was used to validate and verify the prediction of annual generation rates of the six waste streams (organic-food-based, plastic, organic-paper based, glass, metal, and ‘others’) as grouped by Kazuva and Zhang [5]. However, for ease of analysis, the model was introduced in four major sources-based waste categories: Residential, industrial, commercial, and institution and services wastes. The obtained prediction results are accurate and statistically significant as all independent variables showed a strong linear relationship with dependent variables. Besides, the linear regression model was used for the environmental risk index to predict the environmental risk level for Dar es Salaam SW.

2.3.1. The architecture of the MLP-ANN model

The architecture of MLP-ANN as regression predictive models was developed using the libraries of open-source Python programming language software—3.7 version. Specifically, the study utilized Scikit-learn² to construct and train the MLP-ANN models as well as for performance evaluation metrics, NumPy and Pandas for data manipulation and preprocessing, and Matplotlib and Seaborn for visualization of data distributions and model performance metrics.

In the first step, the researcher completed an Excel dataset and saved it in Comma-Separated Values (CSV) format for each waste stream. This step was done by assigning the average values for missing ones. The first row of the CSV files comprised independent variables and the names of the dependent variables. The independent variables for each waste stream were selected based on the weight of a particular element to the quantity of annual SW generated of each waste stream. Thus, the parameters used for the SW prediction include population [36], urbanization rates [37], gross domestic product (GDP), number of industries (NoI), number of hospitals (NoH), number of schools/colleges (NSC), and the number of core food markets (NFm). The association principle here is that: (1) The population of an area has significant impacts on the increased volume of generated solid waste [38]; (2) urbanization rates indicate the direction of population size and influence waste generation rates of an area [38,39]; (3) The GDP which measures the economic activities, size of the economy and the growth rate of a country is a primary factor for WGR of a particular area [40]; (4) For NoI, NoH, NSC, and NFm which are among the primary waste generators were also considered independent variables with significant contributions to the total volume of SW generated. For this reason, the

amount of waste generated from these sources (industries, selected hospitals, health care service facilities, education institutions and major food market centres) was used as important parameters with input data for the prediction of annual WGR.

The second step involved the normalization of a standard value which was applied to both inputs and outputs following reading the data. In the third step, as shown in **Figure 2**, the researcher randomly divided the data into 70% for training and 30% for testing. The fourth step involves the choice of the MLP regression model parameters which was constructed with the training data. The fifth step involved the prediction of the output value with test data by using the MLP-ANN predictive model, the R^2 , and mean squared error (MSE) values calculated from the test data. The principle in this calculation is that, if R^2 is less than “0.90”, then the procedure is to be repeated. But if it is greater than “0.90”, then the prediction results are acceptable and the model is saved. Thus, the ideal R^2 value is the one closest to “1.00” while the acceptable MSE is the one closest to “0.00”.

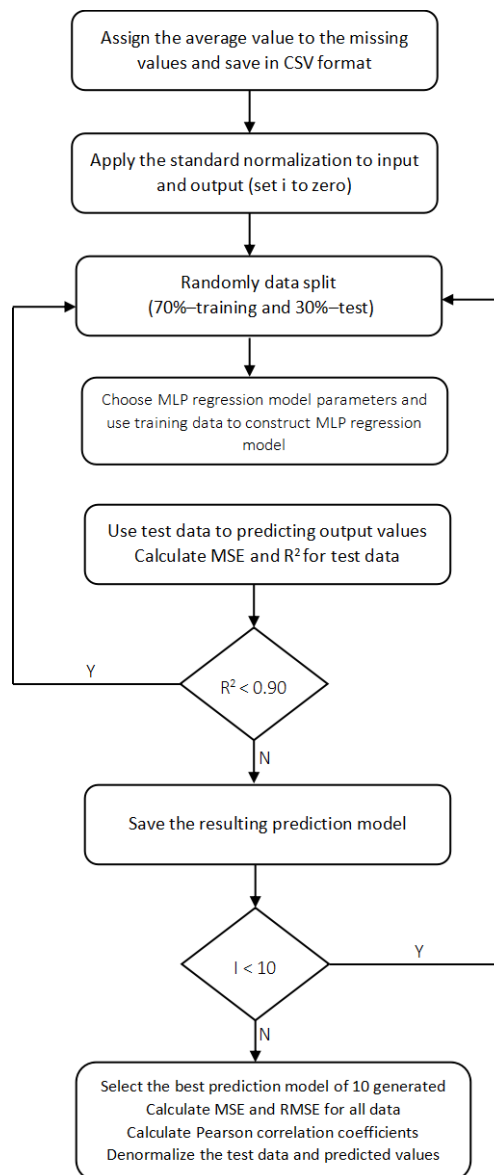


Figure 2. Flowchart of the multi-layer perceptron artificial neural network (MLP-ANN) method for identifying the best model for predicting WGRs in Dar es Salaam.

The random weight initialization tends to affect the performance of ANN [41]. To avoid or reduce this kind of shortfalls of the model, studies recommend repetitions of the algorithms in a reasonable number of attempts to find the optimal regression model with ideal MSE and R^2 (equal or close to “0.00” and “1.00”, respectively). In this case, scholars commonly use ten repetitions in the cross-validation of the ANN [42,43]. For this purpose, the study applied the same approach.

In the last step, test data and predicted values are normalized, and the actual output and predicted values are printed to compare the performance of the model. The main components of MLP-ANN application in the prediction of WGRs for residential, industrial, commercial, and institution and services are also shown in **Figure 2** above. The MSE and R^2 values which are the performance indicators of the model were calculated for all data using the optimal identified model of the ten repetitions/attempts. To find the effect of explanatory variables on results the Pearson Coefficients (P) between each dependent variable and the independent variables were calculated. The similarity shown in parity plots of predicted data against observed data is a sign of the validity of the predictive model used.

2.3.2. MLP-ANN model application

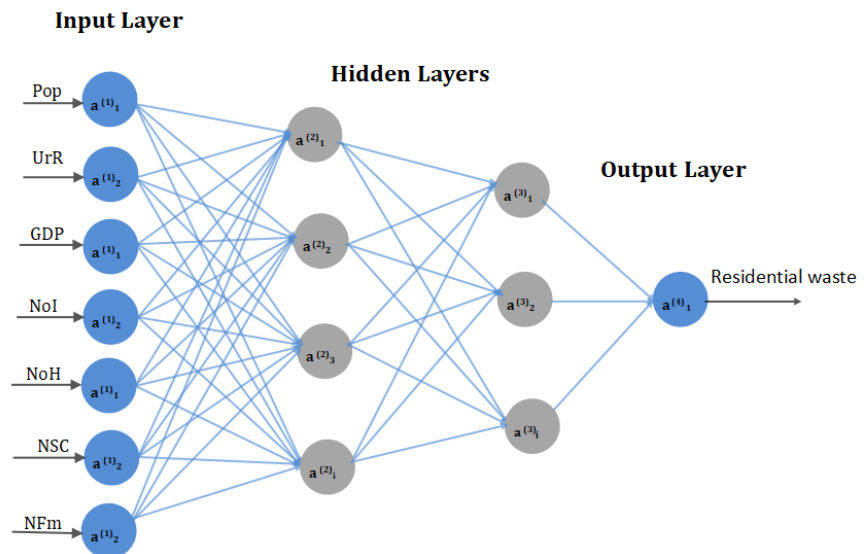


Figure 3. Architecture of the MLP-ANN with two hidden layers for predicting residential waste in Dar es Salaam.

The neurons of the hidden layer ($a_i^{(2)}$ and $a_j^{(3)}$) are defined such that $(i, j) \in \{(10, 5), (20, 10), (40, 10)\}$; GDP: Gross domestic product; NoH: Number of hospitals; NSC: Number of schools; NFm: Number of food markets.

The MLP contains several layers of neurons, whereby the first layer, also known as the lowest layer is an input layer that receives information from external sources/neurons. The last layer, also known as the highest layer is an output layer receiving the processed information as the problem response/problem solution. The intermediate layers which separate the input and output layers are known as hidden layers [44]. For comprehensive WGRs prediction with various aforementioned waste streams, each of the four streams/sources was treated as an independent entity in the model using seven independent variables introduced above (POP, UrR, GDP, NoI,

NoH, NSC, and NFm). **Figure 3** is an illustration of the MLP-ANN model architecture for residential waste. A similar model architecture was employed for other waste streams only by customizing the attributes of the particular stream with the selected independent variables.

As highlighted by Dursun [44] the use of perceptron learning helps to adjust weights for each representation of an input vector of dependent variables. The input layer receives an external activation vector and transfers the calculated values (Equation (1)), by weighted links to the first hidden layer. The same algorithm (Equation (1)) is applied for the hidden layers. The response at the output layer is calculated with data from hidden layers using the algorithm shown in Equation (2) and the sigmoidal activation function used for calculations is represented in Equation (3).

$$a_i^{k+1} = g \left(\sum_{j=0}^n w_{ij}^{(k)} x_j \right) \quad (1)$$

$$h_w(x) = a_i^{(k+1)} = g \left(\sum_{j=0}^n w_{ij}^{(k)} x_j \right) \quad (2)$$

$$g(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

Since the main structure of the MLP network is experimentally defined, the specialists' and experts' experiences were considered important in defining the process.

By using the gradient descent (GD) of the backpropagation algorithm—also known as the square error loss function, it was possible to modify the weight of the units outside the network and whenever necessary, to reduce the value between the estimate and the actual value [45]. This as shown in (Equation (4)) is dealing with the value of training data with the cost function $J(w)$. To accept and use the GD, the general expressions in (Equation (5)) must be factual. In this equation (Equation (4)), the cost function is partially derivative regarding the weight vector, and this value is multiplied by the learning rate. At this point, the result is subtracted from the weight vector to complete the update.

$$J(w) = \frac{1}{2m} \sum_{j=1}^m (h_w(x^i) - y^i)^2 \quad (4)$$

$$w_{ij}^{(i)} := w_{ij}^{(i)} - \alpha \frac{\partial}{\partial w_{ij}} \quad (5)$$

At this stage, the Root Mean Square Error (RMSE) and the Correlation Coefficient (R) were calculated and used as ANN metrics. For calculating the RMSE, the square root of the Mean Square Error (MSE) value (which is a common metric) was taken. Various mathematical techniques use the MSE in its easiest form to process metrics. According to Wolpert and Macready [45], MSE is used in machine learning

to evaluate the performance of classifiers. The MSE and RMSE calculations were done using the following equations:

$$MSE = \frac{(q_1 - a_1)^2 + \dots + (q_n - a_n)^2}{n} \quad (6)$$

$$RMSE = \sqrt{\frac{(q_1 - a_1)^2 + \dots + (q_n - a_n)^2}{n}} \quad (7)$$

where “ q ” is the estimated values; and “ a ” is the actual values.

Finally, the Correlation Coefficient (RC) was calculated using Equation (8) to evaluate the statistical relationship between the actual and estimated values.

$$R_{x,y} = \frac{\sum_{i=1}^n (x_i - x)(y_i - y)}{\sqrt{\sum_{i=1}^n (x_i - x)^2} \sqrt{\sum_{i=1}^n (y_i - y)^2}} \quad (8)$$

where “ x ” is the value of the variable in the dataset and “ y ” is the estimated value by the model.

The RC was used to present the weight of the linear relationship between two data samples as determined by the two-sided values with the minimum and maximum limit -1 and 1 , for negative and positive correlation, respectively. A negative correlation indicates the extent to which one variable increases as the other decreases, while a positive correlation indicates the extent to which the two variables increase or decrease. Some scholars grouped correlation into five groups whereby a value of 0 means there is no correlation, $\pm 0.1-0.3$ represents weak positive or negative correlation, $\pm 0.4-0.6$ represents moderate positive or negative correlation while $\pm 0.7-0.9$ is a strong positive or negative correlation [41,45]. Exactly -1 or $+1$ represents a perfect downhill (negative) or perfect uphill (positive) correlation. This study adopted the four correlation classes which are commonly used and depict that, the (0) value means no correlation (0) , $\pm(0.1-0.3)$ means weak negative or positive correlation, $\pm(0.3-0.5)$ indicates moderate positive or negative correlation and $\pm(0.5-1.0)$ represent a strong positive or negative correlation.

2.3.3. Performance verification of MLP-ANN and evaluation of prediction results

The study surveyed several structures of the MLP-ANN with different numbers of neurons in three layers as shown in **Figure 3**, two activation functions (logistic and tanh) and different training algorithms to estimate the annual WGRs for all waste streams. The combinations of Solver: 1bfgs/sgd/adam, with the activation functions of logistic and tanh sized $(10-5)$: $(20-10)$: $(40-20)$ were used to carry out the prediction experiments for WGRs. The best ANN-MLP predictive models are decided by performing 135 experiments for each waste stream, making a total of 540 for the four streams considered. Then, from the two model accuracy indices, i.e., MSE: Mean squared error and R^2 : The coefficient of determination the ideal networks were selected (ref. **Figure 3** for residential waste). **Table 1** summarizes the solver, activation functions, number of neurons in the hidden layers, the MSE and R^2 , the optimal results obtained and the model with the lowest MSE and highest R^2 that were selected to

represent be best MLP-ANN model for prediction of WGRs in Dar es Salaam. Similar procedures were followed for the other three waste streams.

Table 1. Specification and performance measures of an optimal MLP-ANN for data testing at each stage for all waste streams.

Waste stream	Solver	Activation function	Hidden layer neurons	Statistical indices for Measuring of accuracy			
				MSE test	R ² test	All MSE	All R ²
Residential	sgd	tanh	10-5	0.004	0.98	0.4233	0.98
Industrial	adam	tanh	20-10	0.012	0.96	0.0155	0.97
Commercial	lbfgs	logistic	40-20	0.006	0.97	0.0941	0.90
Institution & Services	adan	tanh	20-10	0.005	0.97	0.0766	0.94

After obtaining the prediction results, it was essential to validate the results to clear any ground of doubts while making the obtained results. In model validation, the existing/observed data on WGRs and ERL were compared with the acquired WGRs and ERL probability results. In this task, a receiver operating characteristic curve—a simple and reliable tool for the validation of prediction results was employed [46]. From the ROC, results of the success rate were obtained using the training data set, and the prediction accuracy was calculated using the validation dataset that was not used in the training process. The Area Under the Curve (AUC) for ML regression and ANN were calculated in “R” statistical software using training and testing data for success rate and prediction rate of accuracy assessment, respectively.

3. Results

3.1. Waste generation rates and environmental risk index

The waste categories as highlighted by Kazuva and others were adopted in the current study, where over 57% of the waste stream from the entire city was found to be organic waste [2,5]. Other categories are plastics waste (13%), paper waste (6%), glass waste (2%), metals waste (1%) and other assorted forms, making 21%. The characteristics of each category have extensively been studied in previous studies [2,3,5]. As indicated in **Table 2** the rate of SW generation in the city of Dar es Salaam has been increasing significantly from 3930 MT/day in 2006 to over 6080 MT/day in 2019 and 6570 MT/day in 2022. This increase is approximately 54.7% in just 14 years and about 67% in 17 years, respectively. The average amount of SW generation and its percentage composition for each stream is presented in **Table 2**.

Table 2. Average SW generated and composition per stream from 2006 to 2022 (MT/day).

Year	Organic (57.21 %)	Plastic (13.08 %)	Paper (6.12 %)	Glass (2.32 %)	Metal (1.02 %)	Others (20.25 %)	Total SW Generates (100 %)
2006	2248.35	514.04	240.52	91.18	40.09	795.83	3930.01
2007	2354.19	538.24	251.84	95.47	41.97	833.29	4115.00
2008	2452.02	560.61	262.30	99.44	43.72	867.92	4286.01
2009	2548.13	582.58	272.58	103.33	45.43	901.94	4453.99
2010	2618.51	598.67	280.11	106.19	46.69	926.84	4577.01

Table 2. (Continued).

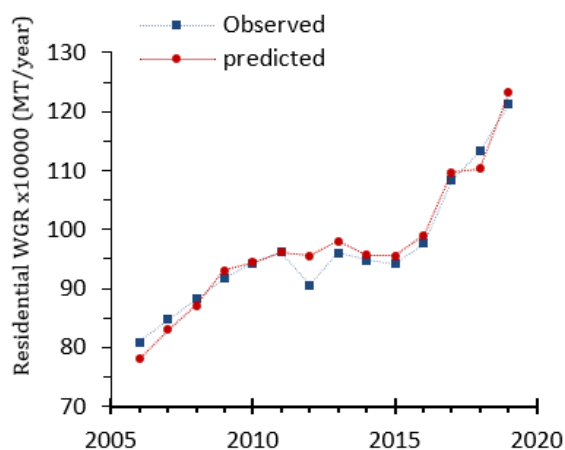
Year	Organic (57.21 %)	Plastic (13.08 %)	Paper (6.12 %)	Glass (2.32 %)	Metal (1.02 %)	Others (20.25 %)	Total SW Generates (100 %)
2011	2671.13	610.71	285.74	108.32	47.62	945.47	4668.99
2012	2515.52	575.13	269.10	102.01	44.85	890.39	4397.00
2013	2666.56	609.66	285.25	108.14	47.54	943.85	4661.00
2014	2634.52	602.33	281.83	106.84	46.97	932.51	4605.00
2015	2616.21	598.15	279.87	106.09	46.64	926.03	4572.99
2016	2711.75	619.99	290.09	109.97	48.35	959.85	4740.00
2017	3008.11	687.75	321.79	121.99	53.63	1064.75	5258.02
2018	3204.90	732.74	342.84	129.97	57.14	1134.41	5602.00
2019	3480.66	795.79	372.34	141.15	62.06	1232.01	6084.01
2020	3502.58	800.80	374.69	142.04	62.45	1239.77	6122.32
2021	3615.57	826.63	386.77	146.62	64.46	1279.76	6319.82
2022	3759.13	859.45	402.13	152.44	67.02	1330.58	6570.75
Total	48,607.84	11,113.28	5199.79	1971.18	866.63	17,205.20	84,963.92

Based on the actual data, it is evident that the generation rate of SW increased significantly for the last 17 years while the response to it was inadequate. This trend made the ERI continue rising from medium to relatively high level, somewhat increasing ecological environment and human health risks [2]. As clearly shown in the national population projection report for 2013–2035 [25,47], the population of this city is fast growing at 4% as the average annual growth rate. As the statistical relationship between population growth rates and increased waste generation rates is shown in MLP-ANN model results (**Table 1**); waste generation rates are expected to grow even faster.

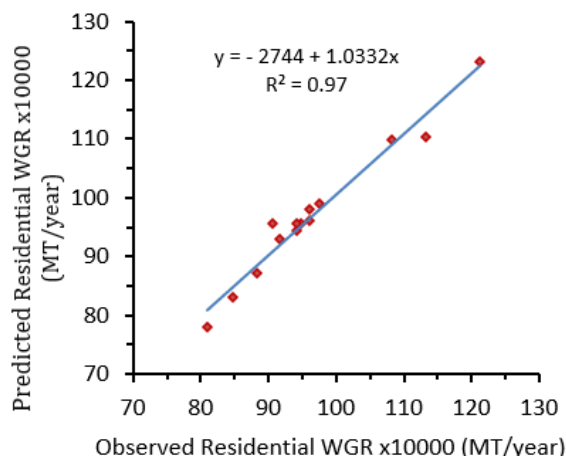
3.2. MLP-ANN prediction results of WGRs

3.2.1. WGRs prediction for individual factors

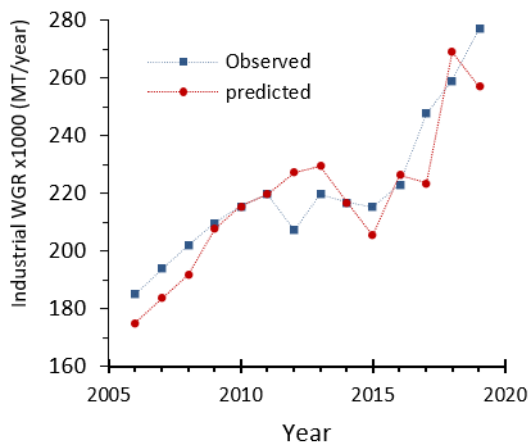
While all six waste streams (organic, plastic, paper, glass, metal, and ‘others’) were processed, the discrete predictive models were developed based on four major waste sources of the mentioned six waste categories which are residential, industrial, commercial, and institution and services wastes. The use of scatter plot—an important and powerful tool for showing the degree of correlation between two variables [48] was used. In this plot, when the data form a graph resembling a straight line, the correlation between the two variables is considered high. In other words, the more the slope value is close to 1.0, the stronger the correlation between the two variables. **Figure 4** is the scatter plots of MLP-ANN predictive models that are graphically present by comparing the actual and predicted WGRs for the four identified waste streams. For ease of graphical staging, only data for 2006–2019 were used.



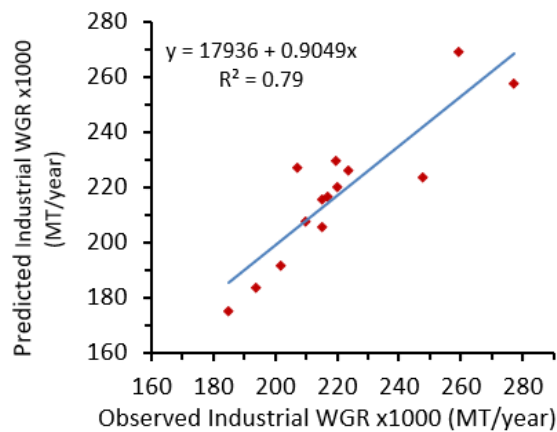
(a1)



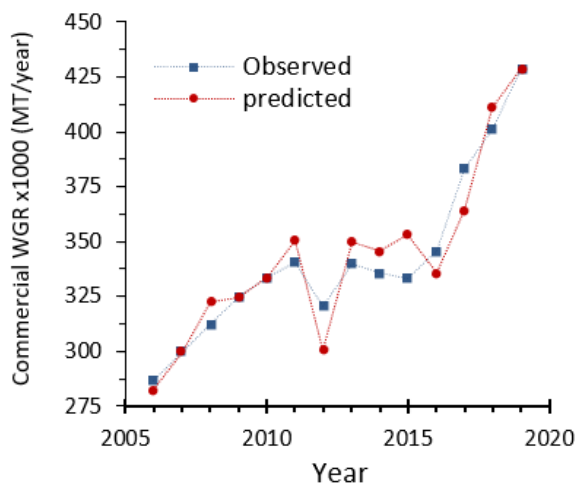
(b1)



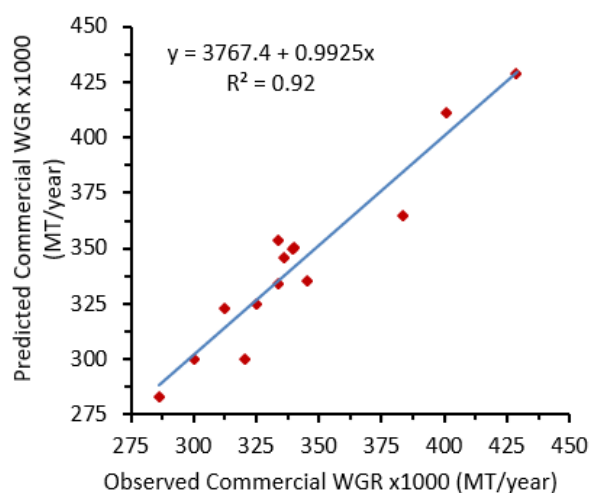
(a2)



(b2)



(a3)



(b3)

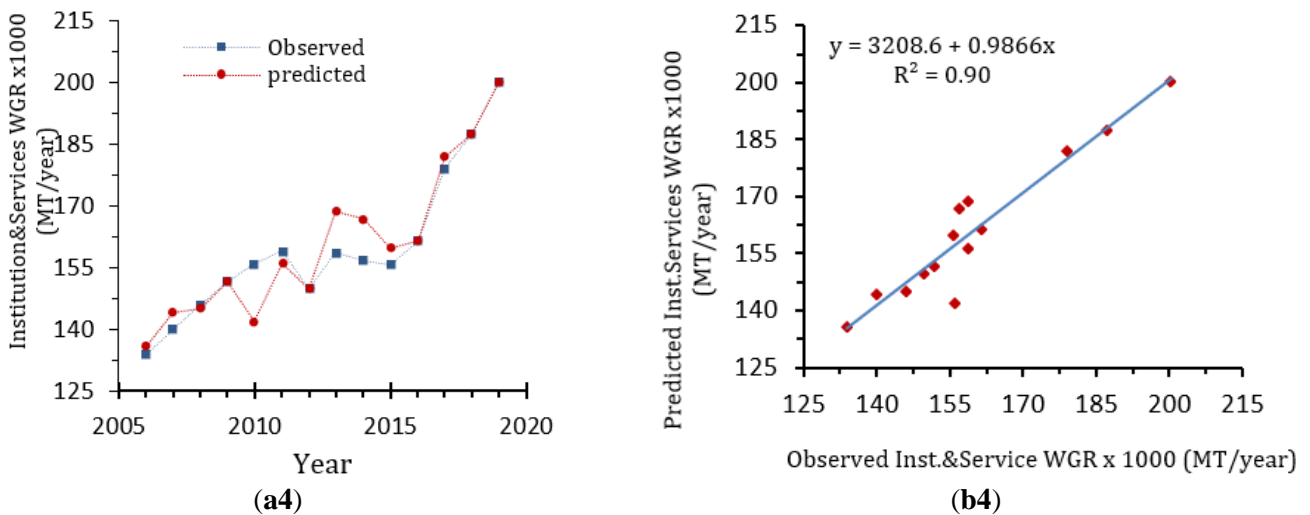


Figure 4. Results of the MLP-ANN model for annual waste generation rates (WGRs) for residential, industrial, commercial and institutional/services waste, (a) comparison of observed and predicted WGRs; (b) scatter plot of observed against predicted WGRs.

Apart from the results presented in **Figure 4** above, **Table 3** presents the Pearson Coefficient for the independent variables to dependent variables (residential, industrial, commercial, institution, and services waste). For instance, under residential waste, except for the school numbers parameter ($P = 0.49$) which is a moderate positive correlation, other independent variables showed a strong positive correlation with annual residential waste generation rates in Dar es Salaam (i.e., population ($P = 0.95$), urbanization ($P = 0.94$), GDP ($P = 0.92$), hospital numbers ($P = 0.63$), and industrial numbers ($P = 0.72$)). Correlation for all variables is presented in **Table 3**.

Table 3. Pearson Coefficients between independent variables and dependent variable (annual Residential, industrial, commercial and institution and service waste generation rates (WGRs)).

Type of waste/Dependent variables	Independent variables	Pearson Coefficient (P)	Correlation type
Residential	Population	0.95	Strong positive
	Urbanization	0.94	Strong positive
	GDP	0.92	Strong positive
	Hospital numbers	0.63	Strong positive
	Industrial numbers	0.72	Strong positive
	School numbers	0.49	Moderate positive
	Markets numbers	0.90	Strong positive
Industrial	Population	0.92	Strong positive
	Urbanization	0.95	Strong positive
	GDP	0.91	Strong positive
	Hospital numbers	0.70	Strong positive
	Industrial numbers	0.98	Strong positive
	School numbers	0.92	Strong positive
	Markets numbers	0.67	Strong positive

Table 3. (Continued).

Type of waste/Dependent variables	Independent variables	Pearson Coefficient (<i>P</i>)	Correlation type
Commercial	Population	0.97	Strong positive
	Urbanization	0.94	Strong positive
	GDP	0.93	Strong positive
	Hospital numbers	0.41	Moderate positive
	Industrial numbers	0.87	Strong positive
	School numbers	0.87	Strong positive
	Markets numbers	0.90	Strong positive
Institution & services	Population	0.92	Strong positive
	Urbanization	0.91	Strong positive
	GDP	0.90	Strong positive
	Hospital numbers	0.96	Strong positive
	Industrial numbers	0.60	Strong positive
	School numbers	0.95	Strong positive
	Markets numbers	0.45	Moderate positive

Generally, all selected independent variables are positively correlated with annual WGRs. The MLP-ANN model was confirmed powerful for both linear and non-linear relationships to classical regression models. The capability of MLP-ANN models in the prediction of annual WGRs in Dar es Salaam has been illustrated with high accuracy (> 85%) in estimating WGRs of the considered waste streams. The adaptive and flexible characteristics of MLP-ANN models together with statistical accuracy indices converged in determining an optimum architecture and information paradigm to predict WGRs after analyzing and training datasets. The study successfully identified independent variables that are highly correlated with dependent variables for residential, industrial, commercial, and institution and services waste datasets. The differences between observed/actual and predicted WGR values are not statistically significant regarding low MSE and high R^2 values. Population size, urbanization trend, and GDP are the most influential parameters for residential, industrial, and commercial WGRs. On the other hand, hospital and school/college numbers are the most effective factors for the institution and services WGRs.

3.2.2. Prediction of comprehensive WGRs

By using actual SW generation data and all indicators presented as independent variables for annual waste generation rates under the MLP-ANN model, SW generation has increased from an average generation of 3930 MT/day equivalent to 1.4 mil. MT/year in 2006 to about 5884 MT/day equivalent to 2.2 mil. MT/year in 2019 and 6570 MT/day equivalent to 2.4 mil. MT/day in 2022. From **Figure 5**, the use of one power simulation to calculate the trendline of the comprehensive waste generation rates (CWGRs) helped to obtain the equation.

$$y = 152.4 x - 30200 \text{ (passes the significance test, } R^2 = 0.8630)$$

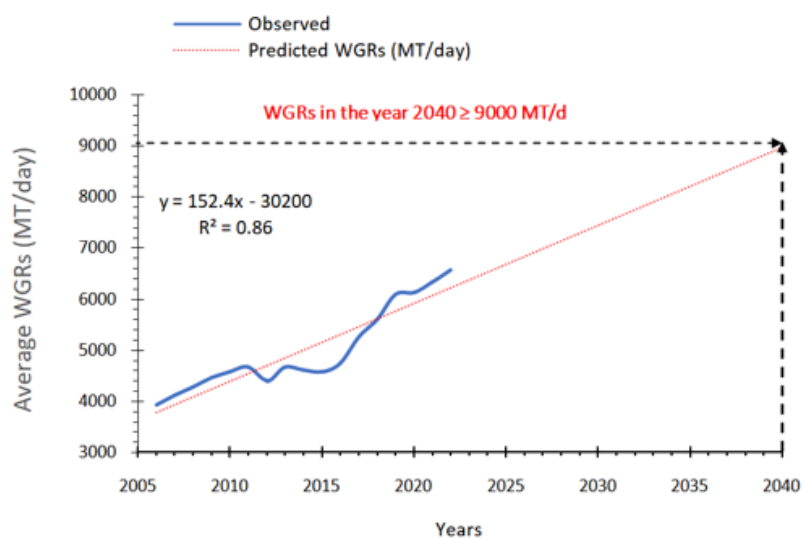


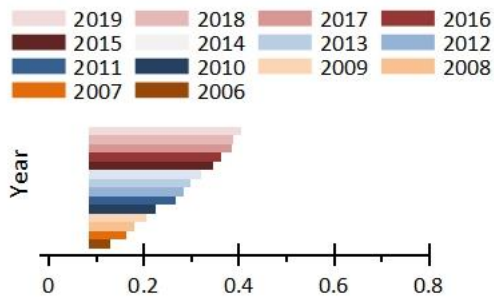
Figure 5. Average SW generation rates and prediction.

As shown in **Figure 5**, the generation rates of SW in Dar es Salaam increased year by year from 2006 to 2022, increasing by 152.4 MT every year. From this trend, and as shown by the prediction line, it is estimated that in 2040 (18 years from 2022) the amount of SW generated will exceed 9000 MT/day. Unless some improvements are made, the available management approaches will be unable to keep pace with this generation rate.

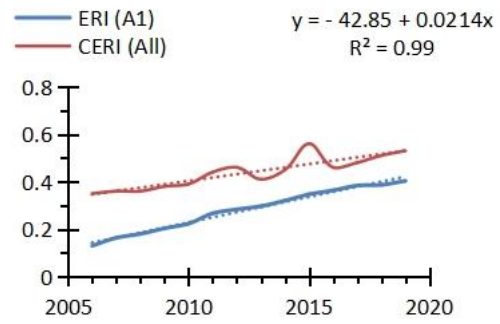
3.3. Environmental risk of SW-induced pollution

From the previous study by Kazuva and others [2], the ERI for Dar es Salaam SW was found to be a function of different factors grouped into five major risk indices of the DPSIR model, namely driving force, pressure, state, impacts, and response indices. These indices were subordinated by different indicators as the subsets of a comprehensive ERI system and were displayed as B- to E-layers.

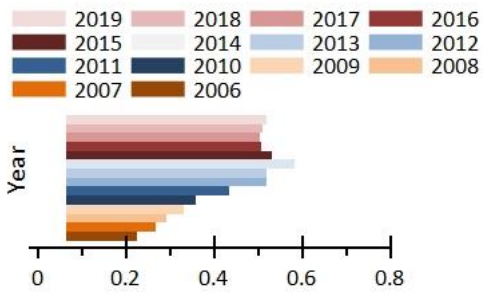
The result shows each factor of the five considered at different magnitudes has a significant contribution to the Comprehensive Environmental Risk Index (CERI) and so to the rise of ERL of Dar es Salaam SW. **Figure 6** compares the trend of each risk factor (index) with that of CERI. The correlation of the two variables is shown by the R-squared value (R^2) where they all indicate a strong positive correlation to the environmental risk level of the city. Similarly, the trendlines extracted equations show the trend of the particular index throughout the assessment period. As for the case of WGRs graphs, 2006–2019 data are inferences used. Despite the decrease in risk value for the response index (**Figure 6i,j**) it was not enough to immediately reverse the upward trend of ERI for other indices (A1–A4). As a result, there has been an increase from 0.22 to 0.45 risk values from 2006 to 2019, respectively. The scatter plot in **Figure 6l** shows the relationship and the degree of correlation between the compounded ERI for A1 through A4 and the ERI for response (A5) index. From this plot, the data resemble a straight line and the generated R-squared value ($R^2 = 0.85$) is strong evidence to justify that the correlation between the two variables is strong. Thus, for sustainable environmental risk reduction from SW in the city of Dar es Salaam, special attention to the response index is of particular importance.



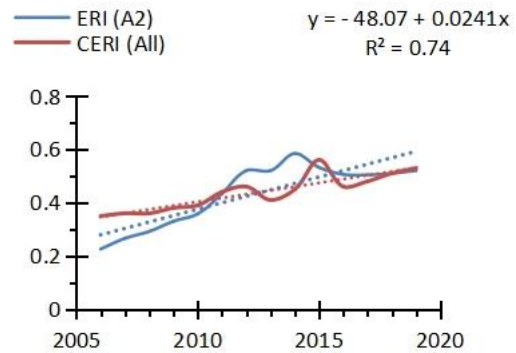
(a)



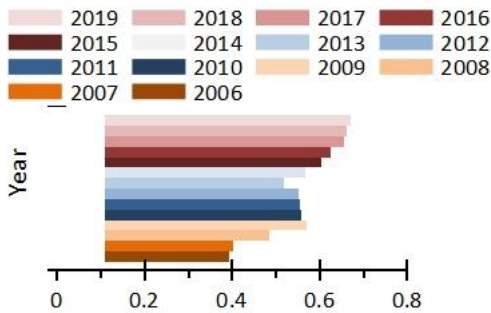
(b)



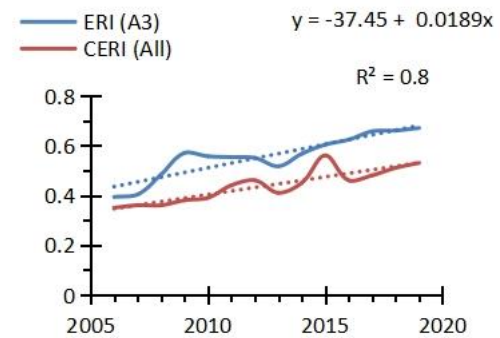
(c)



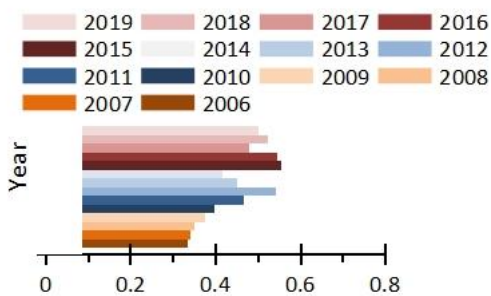
(d)



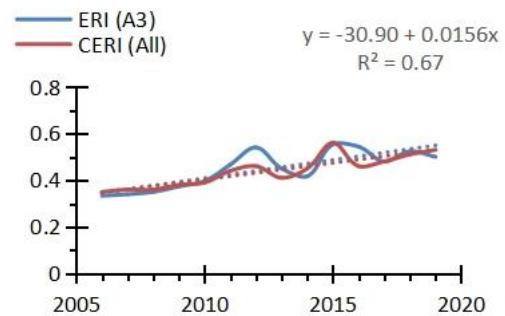
(e)



(f)



(g)



(h)

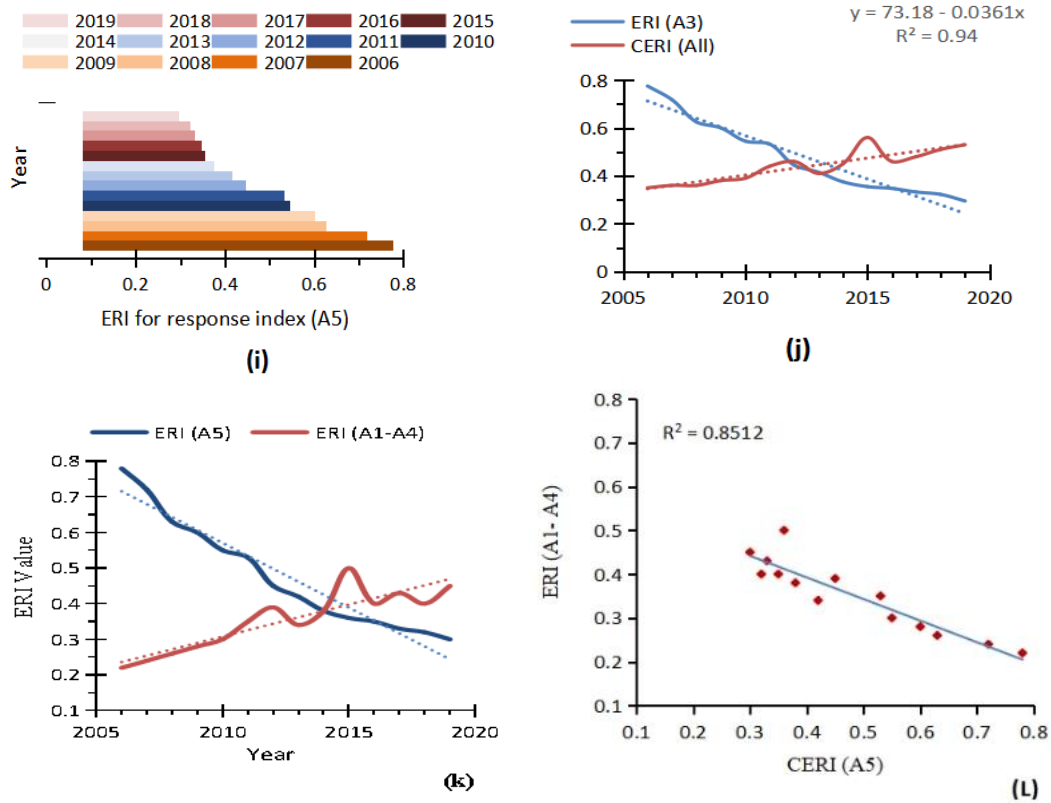


Figure 6. Comparison of ERI for all indices and their influence on the environmental risk level (ERL), (a) trend of driving forces index (A1); (b) comparison of driving forces index (A1) with the CERI; (c) trend of pressure index (A2); (d) comparison of pressure index (A2) with the CERI; (e) trend of state index (A3); (f) comparison of state index (A3) with the CERI; (g) trend of impact index (A4); (h) comparison of impact index (A4) with the CERI; (i) trend of response index (A5); (j) comparison of response index (A5) with the CERI; (k) comparison of ERI value for A5 and the compounded A1–A4; (l) scatter plot of ERI value for A5 and the compounded A1–A4.

3.4. Overall ERI and risk prediction for Dar es Salaam SW

The CERI for Dar es Salaam SW rose steadily from 2006, reaching level III in 2011 and peaking in 2015. Such a trend indicates a substantial increase in the risk value due to external pressures, needing restoration action [49]. The declines in 2013 and 2016 suggested improved management actions, such as improved management capacities, environmental publicity and education. However, other factors dominated by economic variables halted sustained progress. Despite fluctuations, the index remained high by 2019, indicating persistent environmental challenges despite intermittent improvements driven by government and community actions. Such a condition highlights the need for sustained efforts to get rid of all associated risks [50].

As the previous study indicates [2], the study categorizes ERI for Dar es Salaam SW into five levels based on risk degree and weight. These are Level I (0.10–0.2) extremely low, Level II (0.2–0.4) relatively low, Level III (0.4–0.6) Medium risk level, and Level IV (0.6–0.8) which is a relatively high risk. The ERI is considered extremely high at level V (> 0.8), marked as a critical threshold. The key indices—pressure, state, and impacts—exert significant influence, with pressure and state nearing level IV in 2015, while impact approached medium risk. Despite the differences, all preceded the

CERI's level. The trendline projections suggest the pressure index will hit the critical point (0.8 risk value) by 2022, with state and impact following in 2025 and 2030, respectively (**Figure 7**). On the other hand, the study used one power simulation to calculate the trendline of the CERI as shown in **Figure 7**. In this Figure the significance equation ($Y = 0.0141x + 0.332$); pass the significance test, $R^2 = 0.7871$, was obtained; meaning that the CERI of Dar es salaam SW increased year by year from 2006 to 2019 by about 0.014 annually. Therefore, it is estimated that in 2038 the environmental risk of Dar es Salaam SW will reach the critical point ($RV \geq 0.8$).

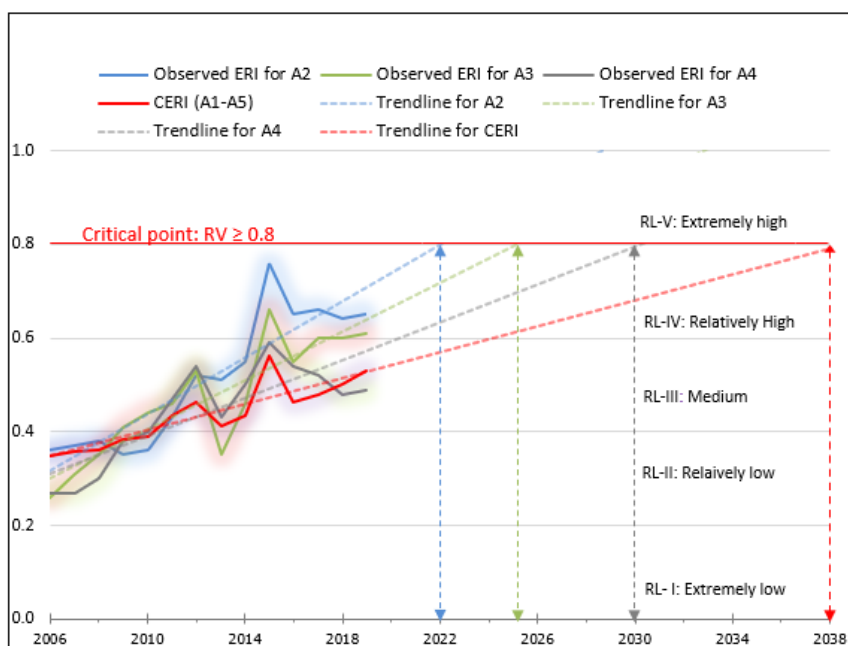


Figure 7. Risk level and prediction of pressure (A_2), state (A_3), impact (A_4) and the CERI for Dar es Salaam SW.

3.5. Validation of prediction results

The study aimed to assess solid waste generation rates (WGRs) in Dar es Salaam and their environmental impacts. This was accomplished by presenting the prediction results of both WGR and ERL. The predictions were validated through comparison with observed data using the receiver operating characteristic [46] curve. The MLP-ANN and MLR models were employed, achieving over 70% prediction accuracy. MLP-ANN demonstrated superior accuracy, with a 96.5% prediction rate and 92.5% success rate (**Figure 8a,b**, respectively), while MLR yielded an 86.0% prediction rate and 95.0% success rate (**Figure 8c,d**, respectively). These models underscore the identified indicators' reliability in predicting environmental risk levels, affirming their significance for decision-making. Responding to these indicators can proportionally reduce environmental risk levels in Dar es Salaam. The ROC curve analysis validated the models' accuracy, with MLP-ANN exhibiting the highest predictive capability. The study emphasizes the importance of these models for informed decision-making and environmental management strategies in the region.

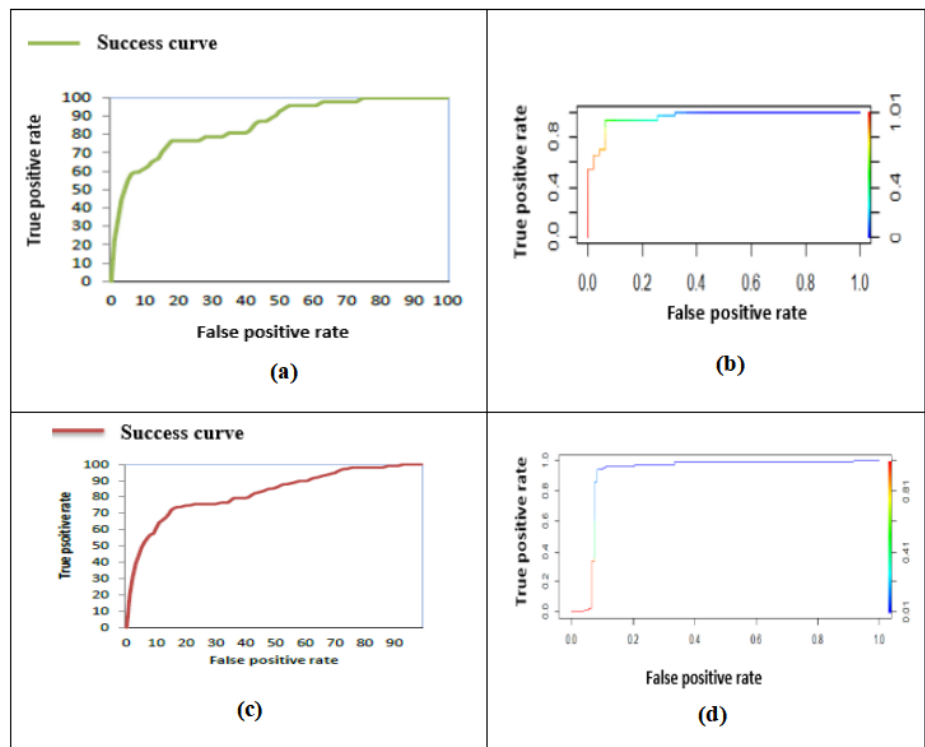


Figure 8. AUC of success and predictive rate for WGRs and ERI, (a) AUC curve of MLP-ANN model for WGRs; (b) WGRs predictive rate; (c) AUC curve of MLR model for ERI; (d) ERI prediction rate.

4. Conclusion

This study provides a comprehensive analysis of solid waste management in Dar es Salaam, focusing on waste generation rates and environmental impacts, particularly the Environmental Risk Index (ERI). Waste generation rates have doubled in less than 20 years, with MLP-ANN models achieving high prediction accuracy and success rates. Over 40% of generated waste remains unmanaged, contributing significantly to environmental hazards. The ERI, derived from the DPSIR framework, reveals pressure, state, and impact indices as critical contributors to pollution, with ERI levels rising from Level II (2006–2010) to Level III (2011–2019) and projected to reach Level V by 2038.

The study highlights the potential of MLP-ANN models for predicting waste generation and assessing environmental risks. Future research could focus on advanced AI techniques, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, to better capture temporal and spatial patterns. Additionally, integrating machine learning with optimization methods like Genetic Algorithms (GAs) or Particle Swarm Optimization (PSO) may enhance resource allocation, optimize waste collection, and support sustainable interventions in waste management. These advancements would contribute to adaptive and efficient waste management systems for Dar es Salaam and other urban areas.

Acknowledgments: The author of this article is grateful for the support received from his employer, The Open University of Tanzania, which played a pivotal role in the success of this study. Additionally, he is grateful to GnG Eco-Cities Ltd., the

Environmental Management argent for funding this study. Furthermore, the author would like to acknowledge the cooperation received from community members who were involved in the survey. Their generosity and active participation in the study are highly valued.

Funding: This work was supported by the Environmental Management Argent “GnG Eco-Cities” under the Project of Waste to Energy Technologies of the City of Dar es Salaam, Tanzania; Grant N0. (GnG-W2R-T-II-2023).

Data availability: The data used to support the findings of this study are included and the sources are well-referred.

Consent to participate: The author registers the consent to participate in this paper.

Consent to publish: The author declares (i) that this is the original research article which in full or any part whatsoever, has not been published, accepted for publication or under editorial review for publication elsewhere; and that our institute’s representative “Department of Geography, Tourism and Hospitality of the Open University of Tanzania” is fully aware of this submission; (ii) Upon acceptance for publication, the author confirms and grants the Advances in Analytic Science (AAS) Journal the right to publish this article in accordance with the journal’s copyright policy.

Conflict of interest: The author declares no conflict of interest.

Notes

- ¹ <https://www.un.org>
- ² <https://scikit-learn.org/stable/>

References

1. Nabegu AB. An analysis of municipal solid waste in Kano metropolis, Nigeria. *Journal of Human Ecology*. 2010; 31(2): 111–119. doi:10.1080/09709274.2010.11906301
2. Kazuva E, Zhang J, Tong Z, et al. The DPSIR model for environmental risk assessment of municipal solid waste in Dar es Salaam city, Tanzania. *International journal of environmental research and public health*. 2018; 15(8): 1692. doi: 10.3390/ijerph15081692
3. Kazuva E, Zhang J, Tong Z, et al. GIS-and MCD-based suitability assessment for optimized location of solid waste landfills in Dar es Salaam, Tanzania. 2021; 28: 11259–11278. doi: 10.1007/s11356-020-11213-0
4. Kazuva E. Determinants of Individuals’ willingness to use Economic Instruments foe Solid waste Management in Dar es Salaam. *International Journal of Environmental Science and Natural Resources*. 2017; 4(4). doi: 10.19080/IJESNR.2017.04.555644
5. Kazuva E, Zhang J. Analyzing municipal solid waste treatment scenarios in rapidly urbanizing cities in developing countries: The case of Dar es Salaam, Tanzania. *International Journal of Environmental Research and Public Health*. 2019; 16(11): 2035. doi:10.3390/ijerph16112035
6. Siddiqua A, Hahladakis JN, Al-Attia WAK. An overview of the environmental pollution and health effects associated with waste landfilling and open dumping. *Environmental Science Pollution Research*. 2022; 29(39): 58514–58536. doi: 10.1007/s11356-022-21578-z
7. Ajibade FO, Adelodun B, Lasisi KH, et al. Environmental pollution and their socioeconomic impacts. *Woodhead Publishing Series in Food Science, Technology and Nutrition*. 2021; E 321–354. doi: 10.1016/B978-0-12-821199-1.00025-0
8. Boadi KO, Kuitunen M. Environmental and health impacts of household solid waste handling and disposal practices in third world cities: The case of the Accra Metropolitan Area, Ghana. *Journal of Environmental Health*. 2005; 68(4): 32–37.

9. Andeobu L, Wibowo S, Grandhi S. Artificial intelligence applications for sustainable solid waste management practices in Australia: A systematic review. *Science of The Total Environment*. 2022; 834: 155389. doi: 10.1016/j.scitotenv.2022.155389
10. Shams SR, Kalantary S, Jahani A, et al. Assessing the effectiveness of artificial neural networks (ANN) and multiple linear regressions (MLR) in forecasting AQI and PM10 and evaluating health impacts through AirQ+ (case study: Tehran). *Environmental Pollution*. 2023; 338: 122623. doi: 10.1016/j.envpol.2023.122623
11. Sahour S, Khanbeyki M, Gholami V, et al. Particle swarm and grey wolf optimization: Enhancing groundwater quality models through artificial neural networks. *Stochastic Environmental Research and Risk Assessment*. 2024; 38(3): 993–1007.
12. Azid A, Juahir H, Toriman ME, et al. Prediction of the level of air pollution using principal component analysis and artificial neural network techniques: A case study in Malaysia. *Water, Air, Soil Pollution*. 2014; 225: 1–14.
13. Moradi MH, Sohani A, Zabihigivi M, et al. Machine learning and artificial intelligence application in land pollution research. *Current Trends Advances in Computer-Aided Intelligent Environmental Data Engineering*. 2022; 273–296. doi: 10.1016/B978-0-323-85597-6.00008-2
14. Olawoyin R. Application of backpropagation artificial neural network prediction model for the PAH bioremediation of polluted soil. *Chemosphere*. 2016; 161: 145–150. doi: 10.1016/j.chemosphere.2016.07.003
15. Kristian S, Hansen SF. Environmental risk assessment of chemicals and nanomaterials—the best foundation for regulatory decision-making? *Science of the Total Environment*. 2016; 541: 784–794. doi: 10.1016/j.scitotenv.2015.09.112
16. Quinlan RJ. Human parental effort and environmental risk. *Proceedings of the Royal Society B: Biological Sciences*. 2007; 274(1606): 121–125. doi: 10.1098/rspb.2006.3690
17. Lindell MK, Perry RW. *Communicating environmental risk in multiethnic communities*. Sage Publications; 2023. Volume 7.
18. Rand GM. *Fundamentals of aquatic toxicology: Effects, environmental fate and risk assessment*. CRC press; 1995.
19. Van der Oost R, Beyer J, Vermeulen NP. Fish bioaccumulation and biomarkers in environmental risk assessment: A review. *Environmental Toxicology and Pharmacology*. 2003; 13(2): 57–149. doi: 10.1016/S1382-6689(02)00126-6
20. McRoberts DB, Quiring SM, Guikema SD. Improving hurricane power outage prediction models through the inclusion of local environmental factors. *Risk Analysis*. 2018; 38(12): 2722–2737. doi:10.1111/risa.12728
21. Abbasi M, El Hanandeh A. Forecasting municipal solid waste generation using artificial intelligence modelling approaches. *Waste Management*. 2016; 56: 13–22. doi: 10.1016/j.wasman.2016.05.018
22. Al-Khatib IA, Fkhidah IA, Khatib JI, et al. Implementation of a multi-variable regression analysis in the assessment of the generation rate and composition of hospital solid waste for the design of a sustainable management system in developing countries. *Waste Management & Research*. 2016; 34(3): 225–234. doi: 10.1177/0734242X15622813
23. Thanh NP, Matsui Y, Fujiwara T. Household solid waste generation and characteristics in a Mekong Delta city, Vietnam. *Journal of Environmental Management*. 2010; 91(11): 2307–2321. doi: 10.1016/j.jenvman.2010.06.016
24. Abbasi M, Rastgoo MN, Nakisa B. Monthly and seasonal modeling of municipal waste generation using radial basis function neural network. *Environmental Progress & Sustainable Energy*. 2019; 38(3): e13033. doi: 10.1002/ep.13033
25. UN. *World population review: Urbanization prospect, United Nations-Population Division*. Available online: <http://worldpopulationreview.com/world-cities/dar-es-salaam-population/#sources> (accessed on 27 February 2024).
26. Anande DM, Luhunga PM. Assessment of socio-economic impacts of the December 2011 flood event in Dar es Salaam, Tanzania. *Atmospheric Climate Sciences*. 2019; 9(03): 421. doi: 10.4236/acs.2019.93029
27. Linster M, Fletcher L. *Using the Pressure-State-Response Model to Develop Indicators of Sustainability*. Available online: https://destinet.eu/resources/...-various-target-groups/individual-puplications/OECD_P-S-R_indicator_model.pdf/download/de/1/OECD_P-S-R_indicator_model.pdf (accessed on 2 February 2024).
28. USEPA. *Solid Waste and Emergency Response—Full Cost Accounting for Municipal Solid Waste Management: A Handbook*. USEPA; 1997.
29. Xie H, Zhang X. The measure and Countermeasure of ecological security in the suburban area in Beijing city of Haidian district as an example. *China Population Resources and Environment*. 2004; 14(3): 23–26.
30. Zebardast L, Salehi E, Afrasiabi H. Application of DPSIR Framework for Integrated Environmental Assessment of Urban Areas: A Case Study of Tehran. *International Journal of Environmental Research*. 2015; 9(2): 445–456.
31. Liu Y, Sun Y, Hao M, et al. Establishment of Assessment Indicator System for Environment-friendly Society Based on DPSIR Model. *Environmental Protection Science*. 2014.

32. Bottero M, Ferretti V. Integrating the analytic network process (ANP) and the driving force-pressure-state-impact-responses (DPSIR) model for the sustainability assessment of territorial transformations. *Management of Environmental Quality*. 2010; 21(5): 618–644. doi: 10.1108/14777831011067926
33. Elliott M. The role of the DPSIR approach and conceptual models in marine environmental management: An example for offshore wind power. *Marine Pollution Bulletin*. 2002; 44(6): iii–vii. doi: 10.1016/S0025-326X(02)00146-7
34. Shao C, Ju M, Zhang Y, Li Z. Eco-environment security assessment study for the Binhai New Area, Tianjin, based on DPSIR model. *Journal of Safety & Environment*. 2008.
35. Chen M, Xu L, Liu T, Huang H. The Ecosystem Health Assessment of Ganjiang River upstream frame model based on PSR. *Journal of Jiangxi Agricultural University*. 2012. 34(4): 839–845.
36. Popescu M-C, Balas VE, Perescu-Popescu L, et al. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*. 2009; 8(7): 579–588.
37. Chen YC. Effects of urbanization on municipal solid waste composition. *Waste Management*. 2018; 79: 828–836. doi: 10.1016/j.wasman.2018.04.017
38. Patel AK, Bundela VS. Quantification and prediction of solid waste generation based on socio-economical parameters. *Journal of Material Cycles and Waste Management*. 2024; 1–19.
39. Teshome YM, Habtu NG, Molla MB, et al. Municipal solid wastes quantification and model forecasting. *Global Journal of Environmental Science and Management*. 2023; 9(2): 227. doi:10.22034/GJESM.2023.02.04
40. Beigl P, Lebersorger S, Salhofer S. Modelling municipal solid waste generation: A review. *Waste management*. 2008; 28(1): 200–214. doi: 10.1016/j.wasman.2006.12.011
41. Wei J, Chen H. Determining the number of factors in approximate factor models by twice K-fold cross-validation. *Economics Letters*. 2020; 191: 109149. doi: 10.1016/j.econlet.2020.109149
42. Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *Peerj computer science*. 2021; 7: e623. doi: 10.7717/peerj-cs.623
43. Karunasingha DSK. Root mean square error or Mean absolute error? Use their ratio as well. *Information Sciences*. 2022; 585: 609–629. doi: 10.1016/j.ins.2021.11.036
44. Dursun B, Fatih A, Metin Z, et al. Modeling and estimating of load demand of electricity generated from hydroelectric power plants in Turkey using machine learning methods. *Advances in Electrical and Computer Engineering*. 2014; 14(1): 121–132.
45. Wolpert DH, Macready WG. Coevolutionary free lunches. *IEEE Transactions on evolutionary computation*. 2005; 9(6): 721–735. doi: 10.1109/TEVC.2005.856205
46. Sentosa I, Rashid AZA, Hizam SM, et al. An empirical study on the internet usage among young creative entrepreneurs in Malaysia: A structural equation modeling approach. *Business, Computer Science*. 2017; 7: 447–456.
47. URT. 2013–2015 National Population Projections in National Bureau of Statistics (NBS)—Ministry of Finance and Planning. Available online: <https://www.nbs.go.tz/nbs/takwimu/census2012/Projection-Report-20132035.pdf> (accessed on 25 June 2023).
48. Flott LW. Using the scatter diagram tool to compare data, show relationship. *Metal Finishing*. 2012; 8(110): 33–35. doi: 10.1016/S0026-0576(13)70148-X
49. Malav LC, Yadav KK, Gupta N, et al. A review on municipal solid waste as a renewable source for waste-to-energy project in India: Current practices, challenges, and future opportunities. *Journal of Cleaner Production*. 2020; 277: 123227. doi: 10.1016/j.jclepro.2020.123227
50. Hossain MS, Santhanam A, Norulaini NAN, et al. Clinical solid waste management practices and its impact on human health and environment—A review. *Waste Management*. 2011; 31(4): 754–766. doi: 10.1016/j.wasman.2010.11.008