Environmental Science and Pollution Research
https://doi.org/10.1007/s11356-024-35388-y

ECOSYSTEM POLLUTION AND EFFECTIVE NATURAL RESOURCES MANAGEMENT

Prediction of municipal waste generation using multi-expression programming for circular economy: a data-driven approach

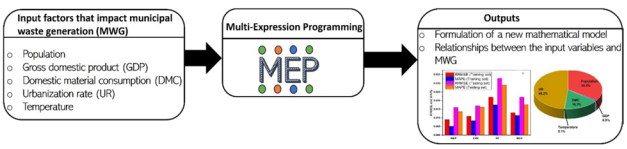
Ayodeji Sulaiman Olawore[1,2] · **Kuan Yew Wong**[1] · **Kamoru Olufemi Oladosu**[2]

Received: 27 May 2024 / Accepted: 19 October 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

The existing surge in municipal waste generation (MWG), characterized by swiftly changing and uncontrollable factors, poses a significant challenge to sustainable development. This prompted the need for improved predictive models to guide strategic waste management within the circular economy framework. This study aims to develop a predictive model using multi-expression programming (MEP) to assess MWG. The model was developed using historical data on socioeconomic and environmental factors and validated via comparative analyses with artificial neural network (ANN), random forest (RF), and multiple linear regression (MLR) using various evaluation metrics. The parametric and sensitivity analyses of the MEP model were also conducted. The MEP, ANN, RF, and MLR models have a coefficient of determination (R^2) (for testing datasets) of 0.977, 0.974, 0.957, and 0.964, respectively. The MEP model is superior in terms of accuracy and performance for the prediction of MWG when compared to the other three models. The sensitivity analysis revealed the relative importance of each input variable in the established MEP model. The novelty of this research lies in the application of MEP to predict MWG and the formulation of a new mathematical model that links socioeconomic and environmental factors with MWG. The model can be used by waste management authorities to optimize waste collection, transportation, and disposal infrastructure for an effective circular economy and sustainable development. This model also aids in the development of effective waste management policies.

Graphical abstract



Keywords Circular economy · Prediction · MEP · Municipal waste · Socioeconomic · Sustainable · Japan

Introduction

Municipalities are increasingly focused on sustainable waste management practices to alleviate the burden on landfills and promote resource efficiency as global environmental concerns escalate (Dobre-Baron et al. 2022). The release of toxins from landfills negatively affects the well-being of nearby residents and living organisms. Certain waste requires more than 500 years to decompose, and the increasing human population, rapid urbanization, and flourishing economy contribute to increasing waste generation that outpaces natural decomposition (He et al. 2022). The transition

Responsible Editor: Kelvin Tsun Wai Ng

✉ Kuan Yew Wong
m-wongky@utm.my

[1] Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Malaysia

[2] Department of Mechanical Engineering, Kwara State University, Malete, Nigeria

Published online: 26 October 2024

toward a circular economy has emerged as a pivotal strategy to address the challenges associated with municipal waste generation (MWG). In the pursuit of sustainable development, the circular economy has emerged as a promising paradigm that emphasizes resource efficiency, waste reduction, and the promotion of closed-loop systems (Petković et al. 2021).

According to the Organization for Economic Cooperation and Development (OECD), municipal waste refers to waste collected by municipal authorities or on their behalf. This category encompasses various waste types, including household waste, bulky waste, and waste from commercial establishments, office buildings, institutions, and small businesses (OECD 2024). A circular economy is well known to have the potential to enhance resource productivity and eco-efficiency, restructure environmental management practices, and promote sustainable development. In a circular economy, the emphasis is on sustainable waste management practices, which include reducing, reusing, recycling, and recovering resources from waste (Kubule et al. 2019).

Accurate estimation of future waste generation quantities is crucial for planning and implementing effective waste management strategies that align with circular economy principles. It empowers authorities and organizations to devise systems fostering resource efficiency, reducing environmental impact, and actively contributing to the overall objective of establishing a more sustainable and circular approach to the management of waste (Jäämaa and Kaipia 2022). A surge in MWG poses a significant challenge to sustainable development, prompting the need for improved predictive models to guide strategic waste management within the circular economy framework. Nevertheless, predicting MWG is often challenging, characterized by swiftly changing and uncontrollable factors (Abbasi and El Hanandeh 2016). Despite the growing recognition of the circular economy's potential to mitigate waste-related environmental impacts, there is a critical gap in the ability to accurately predict MWG using a mathematical model. This problem statement highlights the need to develop a mathematical model using multi-expression programming (MEP) to reveal the intricate relationship and importance of socioeconomic and environmental factors in MWG patterns within the circular economy.

Addressing this gap is imperative for optimizing resource use, minimizing environmental degradation, and propelling the circular economy toward its objectives of sustainable and efficient waste management. The model aids decision-making and allows policymakers to address waste generation at its roots with targeted initiatives (Oliveira et al. 2019). Predicting waste generation quantities aids in determining optimal waste management strategies and supports decision-makers in updating legal acts and regulations (Liu et al. 2021). The application of machine learning is crucial in addressing problems related to waste management due to its ability to solve complex, nonlinear, and multifaceted problems. It offers advantages over conventional statistical models, such as the ability to learn from past data, identify patterns, and make accurate predictions (Bijos et al. 2022). The application of MEP in MWG prediction is justified by its capability to generate interpretable mathematical expressions that directly depict the relationship between input and output variables, in contrast to black-box models (ANN, SVM, RF, etc.) (Jaf et al. 2024). It is also capable of addressing nonlinear correlation, unlike linear regression models. MEP efficiently explores near-optimal solutions by generating several expressions (Iqbal et al. 2021).

This study aims to develop a mathematical model using MEP to predict the municipal waste generated using socioeconomic and environmental factors as input variables. The objectives of this study include (i) to develop a predictive model using MEP and compare its performance with artificial neural network (ANN), random forest (RF), and multiple linear regression (MLR) based on different evaluation metrics; (ii) to determine the relationship between the input variables (socioeconomic and environmental factors) and MWG using parametric analysis; and (iii) to determine the relative importance and contribution of each input variable in the established MEP model using sensitivity analysis. The comprehensive analysis of this study contributed to existing research by developing a novel mathematical expression for MWG predictions and holistically examining the importance and complex relationship of the factors that have not been addressed before using MEP.

Literature review

In the 1990s, municipal waste in the OECD area increased by 19%, closely trending with the growth in private consumption expenditure (33%) and gross domestic product (GDP) (31%). In 2006, the total quantity of municipal waste generated exceeded an estimated 650 million tonnes (OECD 2015). However, this upward trend of MWG is attributed to rapidly fluctuating and unpredictable factors (Fan et al. 2021). This has culminated in various researches in controlling the dynamic nature of MWG at different granularities or time scales such as daily, monthly, and yearly intervals using different predictive models.

ANN was applied by Puntarić et al. (2022) to develop a predictive model for household and similar waste generation. The model was built using socio-demographic characteristics, economic data, and industrial information gathered from Croatia and EU countries. Multivariate linear regression and additive models were developed by He et al. (2022) to establish connections between MSW generation, composition, recovery rates, and factors such as demographics,

economic development, and climate patterns across various cities and regions. Kulisz and Kujawska (2020) took into account various social and economic indicators, including population size, per capita income, the ratio of employment to population, and types of business activities as input variables for the ANN model they developed to predict MWG in Poland. The study by Ribic et al. (2019) involved the application of socioeconomic indicators, including total tourist numbers, total households, and salaries, as independent variables in the development of a prediction model for waste management in Croatia using ANN. In the study of Paulauskaite-Taraseviciene et al. (2022), a k-nearest neighbors (KNN) model was created to forecast municipal solid waste in Lithuania, using socioeconomic and geographical factors like population density, GDP per capita, private property, foreign investment per capita, and tourism as input variables.

Li and Ma (2019) used an Elman neural network (ENN) with input variables including population, passenger volume, GDP, total investment in fixed assets, urban residents' consumption level, and the electricity consumption of the entire society to predict MWG. Also, factors such as GDP, urban population growth, and energy consumption in China were employed by Chhay et al. (2018) to predict MWG. They applied linear regression, gray model, and ANN in developing the MWG prediction models. ANN was also employed by Antanasijević et al. (2013) in their study to forecast the MWG for Bulgaria and Serbia. They used GDP per capita, domestic material consumption (DMC), and resource productivity as the input parameters.

Rosecký et al. (2021) considered socioeconomic and demographic variables, including age, education level, expenditures, and population density, as factors influencing MWG. In Fan et al.'s (2021) study, input variables such as the amount of rainfall, maximum temperature, average temperature, population, household size, educated individuals, GDP, and income were used for predicting MWG. The summary of factors influencing MWG from previous studies is provided in Table 1.

Contemporary waste management models mentioned in the literature include ANN (Oliveira et al. 2019), KNN (Paulauskaite-Taraseviciene et al. 2022), backpropagation ANN (BP-ANN) (Antanasijević et al. 2013), random forest (RF) (Rosecký et al. 2021), support vector regression (SVR) (Liu et al. 2021), generalized regression neural network (GRNN) (Antanasijević et al. 2013), ENN (Li and Ma 2019), and MLR (Rosecký et al. 2021).

MEP has been used in different fields such as supplier selection (Fallahpour et al. 2017) and sustainable utilization of foundry waste (Iqbal et al. 2021). MEP is an evolutionary method that evolves multiple solutions in each chromosome, and the best solution is chosen. This is known as intrinsic parallelism and is a unique feature of MEP (Jaf et al. 2024). This feature does not make MEP more complex than genetic programming or gene expression programming. This method can identify underlying patterns from historical data, allowing for accurate predictions of future waste (Abdalla and Salih 2022). Nonetheless, there is a scarcity of studies utilizing MEP to forecast the MWG based on socioeconomic and environmental factors. To bridge this research gap, this study suggests employing MEP to develop a mathematical model capturing the complex relationship between input and output variables. Furthermore, it is imperative to evaluate the performance and accuracy of the proposed model.

Table 1 Summary of factors influencing MWG

Authors	POP	GDP	RP	DMC	T	EC	EL	R	Age	I	TBA	To	UE	UR
Rosecký et al. (2021)	✓	✓							✓					
Fan et al. (2021)	✓	✓			✓		✓	✓						
He et al. (2022)	✓	✓			✓	✓	✓							
Kulisz and Kujawska (2020)	✓	✓									✓			
Ribic et al. (2019)	✓	✓								✓	✓	✓		
Paulauskaite-Taraseviciene et al. (2022)	✓	✓		✓								✓		
Li and Ma (2019)	✓	✓		✓	✓	✓								✓
Chhay et al. (2018)	✓	✓				✓								✓
Antanasijević et al. (2013)		✓	✓	✓										
Liu et al. (2021)	✓	✓		✓						✓		✓		✓
Keser et al. (2012)	✓	✓			✓			✓		✓			✓	✓

POP, population; *GDP*, gross domestic product; *RP*, resource productivity; *DMC*, domestic material consumption; *T*, temperature; *EC*, energy consumption; *EL*, educational level; *R*, rainfall; *TBA*, types of business activities; *I*, income; *To*, tourist; *UE*, unemployment rate; *UR*, urbanization rate

Methodology

Case study: Japan

Japan is an East Asian nation with 1719 municipalities and a population of about 127 million. It covers a total land area of 377,974 km^2 and experiences a climate spectrum ranging from humid continental in Hokkaido to humid subtropical in parts of Honshu and Okinawa Prefectures. The nation boasts a well-established and diverse economy (Inoue 2022). Japan serves as a case study in this research owing to its rich availability of reliable and detailed data, crucial for building precise predictive models. Japan implements diverse waste management methods, such as recycling, composting, and waste-to-energy incineration, facilitating an exploration of their effects on waste generation rates (Yano and Sakai 2016). The insights gained from this case study have the potential to enhance the development of more efficient waste management strategies worldwide.

Data collection

For a thorough understanding and prediction of MWG within the framework of the circular economy, it is crucial to collect data and information on factors that impact MWG. The literature provides relevant factors, such as population, GDP, DMC, urbanization rate (UR), and average air temperature, as input variables for the development of a predictive model for MWG. These factors are considered in this study because they are the most commonly studied and highly influential factors in MWG, as supported by the analysis of previous research (as shown in Table 1). The definitions of these factors are provided in Table 2.

The accessibility and availability of data for other potential factors constrained their inclusion. This study focused on using socioeconomic factors (such as population, GDP, DMC, and UR) and environmental factors (average air temperature) as the input variables. The data for the input variables, except DMC, were collected from the World Bank's data bank, while the annual municipal waste generated (output variable) and DMC were collected from the OECD's data bank. Table 3 presents the annual collection of these data from 1985 to 2019. Data after 2019 were not available yet at the World Bank's data bank and OECD's data bank. This is not a problem in this study because it aimed to develop an associative prediction model where a set of input variables is used to predict an output variable regardless of time factors. It is not about time series prediction that requires the usage of data from the latest years.

The mean, range, median, standard deviation (SD), coefficient of variation (COV), kurtosis, and skewness are presented as the statistical analyses of the variables in Table 4. The collected data of population, DMC, temperature, and UR between 1985 and 2019 are illustrated in Fig. 1a. The graphical representations of the collected data for GDP and MWG are presented in Fig. 1b. These data were divided into training and testing datasets at 80% and 20% respectively to prevent overfitting of the model (Hoy et al. 2022). This helps to ensure that the model generalizes well to unseen data, as it is tested on data not used during training. The performance of both the training and testing datasets was examined to ensure better generalization (Hoy et al. 2022).

Machine learning algorithms

Multi-expression programming (MEP)

Linear chromosomes are employed by MEP to code computer programs (CPs). Several CPs can be coded into single chromosomes (Oltean and Dumitrescu 2002). The optimum CPs are selected to represent the chromosome by assessing the fitness values of these solutions. The population of CPs is randomly generated at the commencement of an MEP algorithm (Wang and Yin 2020). A binary tournament process is then used to choose two parents, who are then subjected to recombination with a consistent crossover probability. This recombination yields two offspring. The offspring undergo mutation, and the best offspring in the current population replace the worst offspring. These steps are iteratively repeated to develop the best CPs until the termination criterion is fulfilled (Fallahpour et al. 2021).

Table 2 Definitions of the input variables used in this study

Factors	Definition
Population	The total number of residents in a specific area, regardless of their legal status or citizenship (WorldBank 2024)
GDP	The sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products (WorldBank 2024)
DMC	The amount of materials (in terms of weight) used in an economy, i.e., resources extracted or harvested in the country, plus materials and goods imported, excluding materials and products exported (OECD 2024)
Average air temperature	The temperature of the air near the surface of the earth at 2 m above the ground or ocean surface (WorldBank 2024)
UR	Refers to the percentage of people living in urban areas (WorldBank 2024)

Table 3 Japan's annual data collected for MWG and its factors (obtained from WorldBank (2024) and OECD (2024))

Year	Population (million)	GDP (billion in USD)	DMC (tonnes/capita)	Temperature (°C)	UR	MWG (thousand tonnes/year)
1985	120.837	1427.020	11.466	15.18	0.767	42,093.99
1986	121.482	2120.084	11.764	14.67	0.768	43,395.95
1987	122.069	2580.748	11.907	15.60	0.770	45,374.61
1988	122.578	3125.724	12.654	14.80	0.771	47,495.99
1989	123.069	3109.455	12.925	15.78	0.772	49,349.10
1990	123.478	3185.905	14.042	16.49	0.773	50,256.63
1991	123.964	3648.066	14.144	15.70	0.775	51,132.00
1992	124.425	3980.703	13.733	15.49	0.776	50,902.94
1993	124.829	4536.940	13.466	14.93	0.777	51,267.90
1994	125.178	4998.798	13.750	16.38	0.779	51,800.00
1995	125.472	5545.564	13.683	15.55	0.780	52,224.00
1996	125.757	4923.392	13.908	15.13	0.781	52,908.00
1997	126.057	4492.449	13.811	15.86	0.783	53,098.00
1998	126.400	4098.363	13.006	16.23	0.784	53,605.00
1999	126.631	4635.982	12.931	16.23	0.785	53,699.00
2000	126.843	4968.359	13.274	15.93	0.786	54,833.00
2001	127.149	4374.712	12.798	15.65	0.800	54,681.00
2002	127.445	4182.846	12.440	15.96	0.816	54,199.00
2003	127.718	4519.562	12.351	15.52	0.832	54,271.00
2004	127.761	4893.116	12.440	16.66	0.846	53,376.00
2005	127.773	4831.467	12.377	15.60	0.860	52,720.00
2006	127.854	4601.663	12.470	15.72	0.871	52,024.00
2007	128.001	4579.751	12.465	16.31	0.881	50,816.00
2008	128.063	5106.679	11.932	15.99	0.891	48,106.00
2009	128.047	5289.493	10.599	16.03	0.900	46,252.00
2010	128.070	5759.072	10.925	16.25	0.908	45,359.00
2011	127.833	6233.147	10.798	15.85	0.911	45,430.00
2012	127.629	6272.363	11.581	15.57	0.911	45,234.00
2013	127.445	5212.328	11.881	16.13	0.912	44,874.00
2014	127.276	4896.994	11.560	15.92	0.913	44,317.00
2015	127.141	4444.931	11.259	16.33	0.914	43,981.00
2016	127.076	5003.678	11.141	16.44	0.915	43,170.00
2017	126.972	4930.837	11.083	15.88	0.915	42,894.00
2018	126.811	5040.881	10.943	16.48	0.916	42,727.00
2019	126.633	5117.994	11.008	16.68	0.917	42,737.00

Figure 2 depicts the steps involved in MEP. The MEP genes are depicted as sequences of expressions, and the number of these genes is determined by the chromosome length, which remains constant throughout the computation. For an elaborate explanation of the MEP algorithm, readers can refer to the research of Oltean and Dumitrescu (2002).

In this study, the MEP model was developed using an open-source modeling software called MEPX with a 2023.4.3.0-beta version. The parameters used to develop the MEP model in this study are indicated in Table 5. Function sets such as addition (+), subtraction (−), multiplication (×), division (÷), power (^), square root ($\sqrt{}$), and natural logarithm (*Ln*) were used and selected via a trial-and-error approach. The random numbers between the range of − 10 and + 10 were also used through a trial-and-error approach. This range was chosen to promote diversity and prevent numerical instability in some of the mathematical expressions that might result in overflows or underflows during computation. The best CPs were evaluated using mean squared error (MSE) as the fitness function. MSE is commonly used as a fitness function in MEP because it emphasizes larger errors, making it suitable for most regression tasks (Montgomery and Runger 2018). The parameters used to develop the MEP model, as

Table 4 The statistical analyses of the variables

Variables	Population (million)	GDP (billion in USD)	DMC (tonnes/capita)	Temperature (°C)	UR	MWG (thousand tonnes/year)
Minimum	120.837	1427.020	10.599	14.67	0.767	42,093.99
Maximum	128.070	6272.363	14.144	16.68	0.917	54,833.00
Range	7.233	4845.343	3.545	2.01	0.150	12,739.01
Mean	126.051	4476.259	12.358	15.85	0.836	48,874.40
Median	126.843	4635.982	12.440	15.88	0.816	50,256.63
SD	2.075	1059.116	1.061	0.51	0.062	43,10.11
COV	0.016	0.230	0.083	0.03	0.072	0.09
Kurtosis	0.227	1.340	−1.169	−0.08	−1.823	−1.57
Skewness	−1.091	−0.985	0.074	−0.50	0.221	−0.17

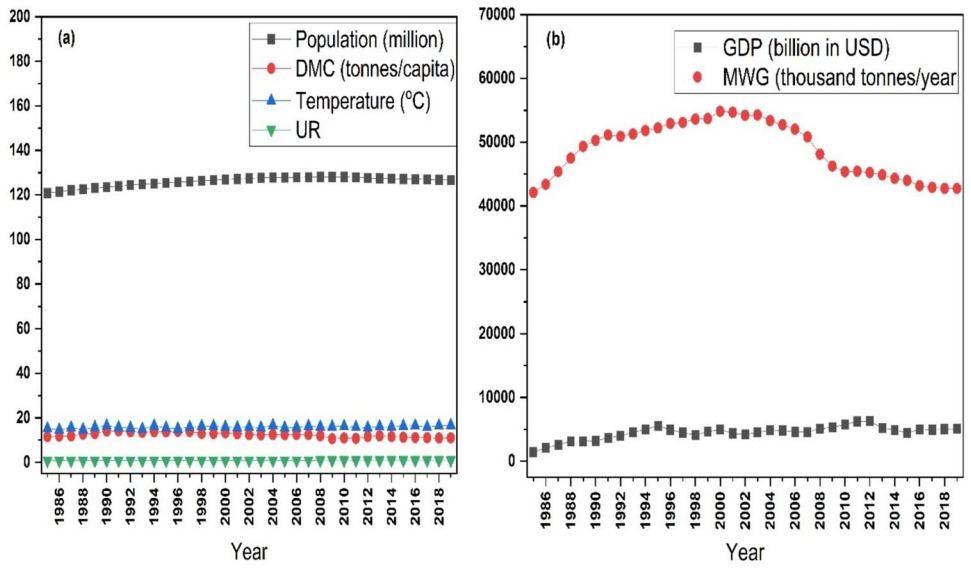


Fig. 1 Graphical representation of the collected data between 1985 and 2019: **a** population, DMC, temperature, and UR; **b** GDP and MWG (obtained from WorldBank (2024) and OECD (2024))

indicated in Table 5, were fine-tuned through a combination of literature guidance and a trial-and-error approach. Specifically, the base values for the number of subpopulations, subpopulation size, code length, crossover probability, mutation probability, tournament size, number of generations, and number of runs were initially sourced from the literature. These base values together with a respective lower value and higher value were tested using a trial-and-error approach for each of the parameters. While the value of one parameter was tested, the values of the other parameters were held constant at their base levels. This resulted in 24 trial-and-error runs. The final parametric values are shown in Table 5, and they corroborate well with the settings suggested by Asghar et al. (2024) to expedite the convergence process in MEP and to obtain better CPs. These parametric values resulted in high performance in terms of accuracy and generalization during preliminary trials.

Artificial neural network (ANN)

ANNs are computational algorithms that mimic and process information in a manner analogous to the functioning of the nervous system. They possess the capability to model one or more outputs by uncovering intricate relationships between input and output variables (Öztürk and Başar 2022). Among the prevalent types of feedforward neural networks is the multilayer perceptron (MLP), characterized by three sequential layers of nodes: an input layer comprising input nodes, one or more hidden layers housing computational nodes, and an output layer (Shao et al. 2019).

In this study, the development of the ANN model was achieved by using MATLAB (R2023b version). A feedforward back propagation neural network system with a hidden layer and an output layer was used for this study due to the regression nature of the research problem (Nguyen et al. 2020). Levenberg–Marquardt function ("*trainlm*"), a

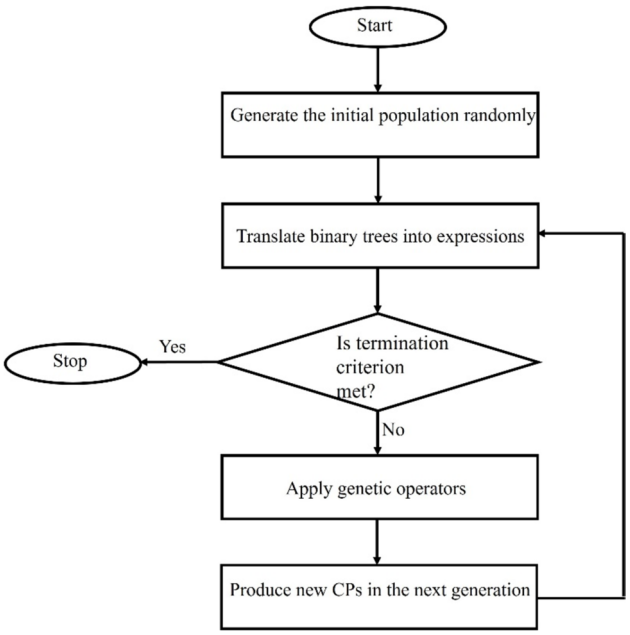


Fig. 2 Steps involved in MEP

Table 5 Parameters used to develop the MEP model

Parameters	Values
Function sets	$+, -, \times, \div, \wedge, \sqrt{}, Ln$
Terminal sets	x_0, x_1, x_2, x_3, x_4, rand
Fitness function	MSE
Number of subpopulations	100
Subpopulation size	2000
Code length	50
Crossover probability	0.9
Mutation probability	0.01
Tournament size	2
Number of generations	2000
Number of runs	10

common and dependable learning technique for ANNs, was used to learn the training dataset and expedite convergence. As highlighted by Caudill (1988), a network containing i inputs necessitates a maximum of $2i+1$ hidden neurons to adequately map any continuous problem. A hidden layer with a number of hidden neurons ranging from 1 to $2i+1$ and an output layer make up the possible ANN models. Since this study used five input variables ($i = 5$), ANN models with 1 to 11 hidden neurons were taken into account.

The transfer functions selected for the hidden and output layers were a hyperbolic tangent function ("*tansig*") and a linear function ("*purelin*"), respectively. The choice of these transfer functions was for effective training and convergence of ANN models for regression tasks (Shao et al. 2019). The ANN models were trained with 200 epochs (Nguyen et al. 2020). The performances of the ANN models with 1 to 11 hidden neurons were examined using the coefficient of determination (R^2) and relative root mean square error (RRMSE) as presented in Table 6 and ranked accordingly (Rao et al. 2018). The RRMSE values of the ANN models were ranked in ascending order from the lowest to the highest, while the R^2 values of the ANN models were ranked in descending order from the highest to the lowest. Table 6 shows that the ANN model with two hidden neurons and a rank value of 4 performs better than the other networks. The optimal network is thus determined to be the ANN model with two hidden neurons.

Random forest (RF)

Random forest is a versatile and robust ensemble learning technique widely used in machine learning and data mining. It belongs to the family of decision tree-based algorithms and is known for its ability to handle complex datasets and produce highly accurate predictions (Ma et al. 2023). During its construction, the model introduces randomness via two mechanisms. Initially, each tree is created using a randomly chosen subset of the dataset, a technique referred to as bagging. Furthermore, at each node of the tree, a random subset of features (input variables) is considered for splitting, further enhancing the diversity among individual trees (Langsetmo et al. 2023). This strategy bolsters the model's robustness and accuracy by using the bagging meta-algorithm, which enhances stability and mitigates the risk of overfitting (Rosecký et al. 2021). In this study, the "*TreeBagger*" function in MATLAB (R2023b version) was used to train or learn the training dataset of the model. The bagging method was selected to improve the performance and robustness of the model.

Multiple linear regression (MLR)

MLR is a fundamental statistical method used for forecasting an output variable based on its linear relationship with more than one input variable. It extends the concept of simple linear regression to accommodate situations where more than one input variable influences the output of the model (Öztürk and Başar 2022). The relationship between the multiple input variables and the output variable is presented in Eq. (1).

$$y = \beta_0 + \beta_1 x_0 + \beta_2 x_1 + \cdots + \beta_n x_{n-1} \quad (1)$$

where y is the output variable; x_0, x_1, and x_{n-1} are the input variables; and β_0, β_1, β_2, and β_n are the constants. However, it assumes a linear relationship between variables, and the presence of multicollinearity or outliers can affect the

Table 6 Performance and ranking of the ANN models with 1 to 11 hidden neurons

Number of hidden neurons	Model performance				Ranking of the model				
	Training dataset		Testing dataset		Training dataset		Testing dataset		Overall ranking
	R^2	RRMSE	R^2	RRMSE	R^2	RRMSE	R^2	RRMSE	
1	0.7701	0.0453	0.6469	0.0608	11	11	11	11	44
2	0.9849	0.0108	0.9735	0.0174	1	1	1	1	4
3	0.8664	0.0321	0.9376	0.0286	10	10	6	6	32
4	0.9509	0.0197	0.9060	0.0298	6	6	8	8	28
5	0.9485	0.0235	0.8987	0.0299	7	7	9	9	32
6	0.9048	0.0299	0.9586	0.0297	9	9	4	7	29
7	0.9608	0.0170	0.9705	0.0193	5	5	2	2	14
8	0.9760	0.0134	0.9321	0.0253	3	3	7	5	18
9	0.9725	0.0164	0.9579	0.0212	4	4	5	3	16
10	0.9839	0.0122	0.9592	0.0234	2	2	3	4	11
11	0.9396	0.0253	0.7866	0.0576	8	8	10	10	36

model's performance. The programming language used in this research to write codes for developing the MLR model was MATLAB (R2023b version). The training datasets for the MLR model were trained or learned using the "*fitlm*" function. The trained model was used to make predictions on the testing datasets using the "*predict*" function.

Performance of models

The performance of the models developed in this study was assessed to determine their degree of accuracy and effectiveness (Chu et al. 2021). The models were assessed using different evaluation metrics, such as root mean square error (RMSE), relative root mean square error (RRMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2). The mathematical expressions for RMSE, RRMSE, MAE, MAPE, and R^2 are presented in Eqs. (2), (3), (4), (5), and (6), respectively. The lower the values of RMSE, RRMSE, MAE, and MAPE, the better the accuracy and effectiveness of a model. The R^2 value ranges from 0 to 1: a model possesses superior accuracy as its R^2 value approaches unity (Liu et al. 2022).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widetilde{y}_i\right)^2} \tag{2}$$

$$RRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widetilde{y}_i\right)^2}}{\overline{y}} \tag{3}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \widetilde{y}_i\right| \tag{4}$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - \widetilde{y}_i}{y_i}\right| \tag{5}$$

$$R^2 = 1 - \left(\frac{\sum_{i=1}^{N}\left(y_i - \widetilde{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i - \overline{y}\right)^2}\right) \tag{6}$$

where y_i and $\widetilde{y}_i$ are the actual and predicted values of MWG, respectively. $\overline{y}$ and N are the average of the actual values and number of datasets, respectively.

A non-parametric test called the Mann–Whitney U test was also conducted on the actual and predicted values of MWG for the testing datasets at a 95% confidence level. It is necessary to examine if there are significant differences between the actual and predicted values of the 4 models. This test was employed in this study since it does not need the data to have a normal distribution (MacFarland and Yates 2016).

Results and discussion

Developed MEP model

The x_o, x_1, x_2, x_3, x_4, and MWG_{MEP} are represented as population, GDP, DMC, temperature, UR, and MWG for the MEP model, respectively. The outcome of the formulated MEP model for the prediction of MWG encompassing the five input variables considered in this study is presented in the form of a sequence of expressions (Table 7). A chromosome's initial expression must be a terminal set. The expressions are arranged hierarchically (prg[1] to prg[48]), demonstrating how simpler expressions are combined to

Table 7 MEP model in the form of expressions

prg[0] = x[0];	prg[17] = prg[11] + prg[11];	prg[34] = prg[32] / prg[26];
prg[1] = x[1];	prg[18] = prg[15] * prg[3];	prg[35] = prg[31] / prg[22];
prg[2] = x[2];	prg[19] = x[3];	prg[36] = prg[30] / prg[17];
prg[3] = x[4];	prg[20] = prg[18]—prg[11];	prg[37] = prg[28] / prg[36];
prg[4] = prg[2] / prg[3];	prg[21] = prg[12] * prg[20];	prg[38] = prg[34] + prg[35];
prg[5] = prg[0] * prg[0];	prg[22] = -2.613;	prg[39] = prg[16] + prg[28];
prg[6] = Ln(prg[3]);	prg[23] = prg[19] * prg[7];	prg[40] = prg[26]—prg[30];
prg[7] = 7.74591;	prg[24] = prg[0] + prg[14];	prg[41] = prg[5]—prg[34];
prg[8] = prg[4] + prg[4];	prg[25] = prg[0] * prg[7];	prg[42] = prg[38] + prg[40];
prg[9] = prg[1] / prg[7];	prg[26] = prg[16] + prg[23];	prg[43] = prg[39]—prg[41];
prg[10] = log(prg[5]);	prg[27] = prg[24] * prg[19];	prg[44] = prg[43] / prg[40];
prg[11] = sqrt(prg[9]);	prg[28] = prg[27]—prg[18];	prg[45] = prg[42] + prg[44];
prg[12] = pow(prg[0], prg[3]);	prg[29] = prg[24] + prg[25];	prg[46] = prg[6] + prg[33];
prg[13] = prg[2] * prg[2];	prg[30] = prg[8] * prg[24];	prg[47] = prg[37] + prg[45];
prg[14] = prg[10]—prg[13];	prg[31] = prg[29] + prg[21];	prg[48] = prg[46] + prg[47];
prg[15] = prg[6] * prg[5];	prg[32] = prg[14] * prg[7];	outputs [0] = prg[48];
prg[16] = prg[15] / prg[2];	prg[33] = prg[15] / prg[9];	

generate more complex ones. Table 7 shows how the solution has evolved across the algorithm's generations. These expressions are decoded into mathematical equations as presented in Eqs. (7–12). The developed MEP model reveals the interrelationship between the input variables (socioeconomic and environmental factors) and MWG. This model serves as a tool for effective planning in transitioning to an environmentally sustainable and economically efficient circular economy.

$$MWG_{MEP} = \frac{x_1 Lnx_4 + \left(7.74591{x_0}^2\right)Lnx_4}{x_1} + \frac{Ax_4\left(x_0x_3 + x_3{Lnx_0}^2 - {x_2}^2x_3 - {x_0}^2x_4Lnx_4\right)}{x_0x_2 + x_2{Lnx_0}^2 - {x_2}^3} + \frac{E}{D} + B + C + D \tag{7}$$

where,

$$A = \sqrt{\frac{x_1}{7.74591}} \tag{8}$$

$$B = \frac{7.74591x_2\left({Lnx_0}^2 - {x_2}^2\right)}{{x_0}^2Lnx_4 + 7.74591x_2x_3} \tag{9}$$

$$C = \frac{8.74591x_0 + {Lnx_0}^2 - {x_2}^2 + {x_0}^{x_4}\left({x_0}^2x_4Lnx_4 - A\right)}{-2.613} \tag{10}$$

$$D = \frac{{x_0}^2x_4Lnx_4 + 7.74591x_2x_3x_4 - 2{x_2}^2x_0 - 2{x_2}^2{Lnx_0}^2 + 2{x_2}^4}{x_2x_4} \tag{11}$$

$$E = \frac{{x_0}^2Lnx_4 + x_0x_2x_3 + x_2x_3{Lnx_0}^2 - {x_2}^3x_3 - {x_0}^2x_2x_4Lnx_4 - x_2\left({x_0}^2 - B\right)}{x_2} \tag{12}$$

Validation of MEP model

The MEP model was validated via comparative analysis with ANN, RF, and MLR. ANN and RF were considered for this comparative analysis due to their ability to solve nonlinear and complex problems (Abdalla and Salih 2022). However, MLR is applied for linear problems, but it was considered for comparative analysis due to its simplicity and ability to formulate mathematical models (Sahoo and Jha 2013). All of these models are used for regression tasks. The models were trained and tested using the same datasets to ensure consistency in the performance of the models, data input, and repeatability. The mean performance and error percentage were determined to account for variability. The performances of these models were analyzed and compared using R^2, RMSE, RRMSE, MAE,

and MAPE. The R^2 values of the training datasets for the MEP, ANN, RF, and MLR models (Fig. 3a) are 0.991, 0.985, 0.960, and 0.976, respectively, while the R^2 values of the testing datasets for the MEP, ANN, RF, and MLR models (Fig. 3b) are 0.977, 0.974, 0.957, and 0.964, respectively. The MEP, ANN, RF, and MLR models have 0.5–1.5%, 0.5–2.0%, 1.5–3.0%, and 1.0–2.0% error percentages respectively, for the training and testing datasets.

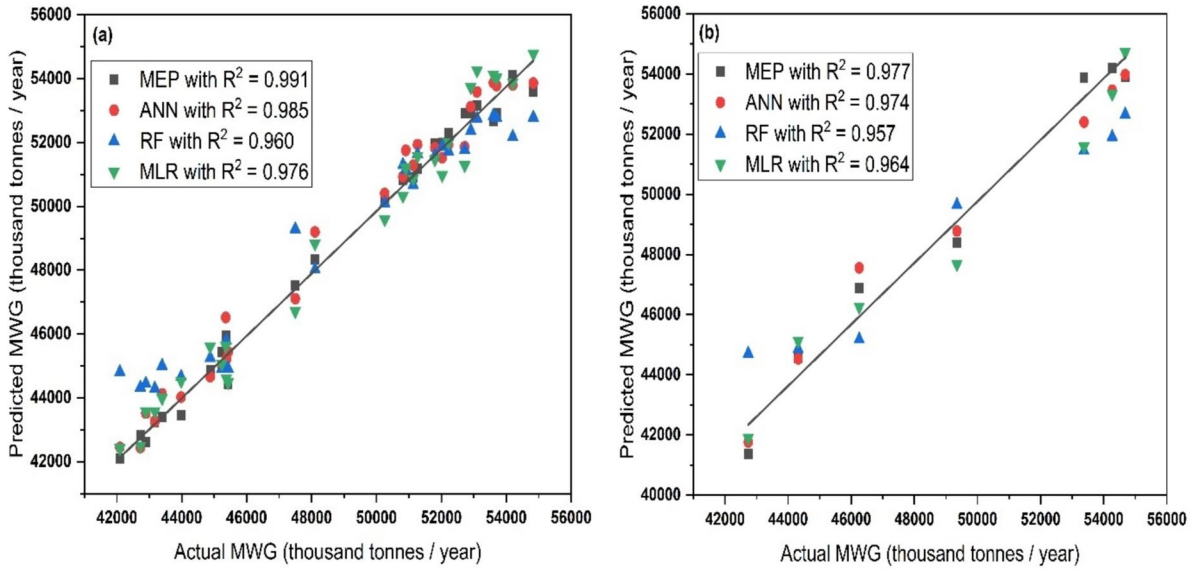


Fig. 3 The R^2 plot of the MEP, ANN, RF, and MLR models for the **a** training datasets and **b** testing datasets

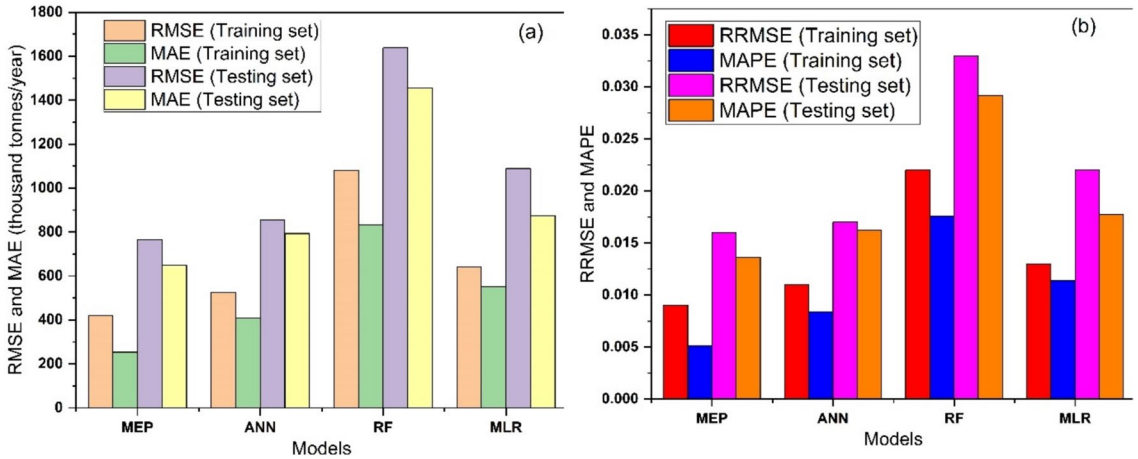


Fig. 4 Performance of the models using **a** RMSE and MAE and **b** RRMSE and MAPE

Table 8 Performance of the models using different evaluation metrics

Evaluation metrics	Training dataset				Testing dataset			
	MEP	ANN	RF	MLR	MEP	ANN	RF	MLR
RMSE*	420.194	524.968	1081.410	639.715	764.887	855.569	1637.835	1089.161
RRMSE	0.009	0.011	0.022	0.013	0.016	0.017	0.033	0.022
MAE*	253.208	408.682	833.668	552.050	649.560	793.235	1455.496	874.520
MAPE (%)	0.512	0.839	1.756	1.136	1.363	1.621	2.915	1.772
R^2	0.991	0.985	0.960	0.976	0.977	0.974	0.957	0.964

*The unit of RMSE and MAE is thousand tonnes/year

The RMSE, RRMSE, MAE, and MAPE values for the training and testing datasets for the models are shown in Fig. 4 and Table 8. The MLR model was developed and presented in the form of a linear mathematical expression as shown in Eq. (13).

$$MWG_{MLR} = -212630.2844 + 2499.3538x_0 - 0.92652x_1 + 1306.336x_2 - 571.9131x_3 - 67698.1224x_4 \quad (13)$$

The comparison of the models (MEP, ANN, RF, and MLR) in terms of R^2 values shows that MEP has the highest R^2 values for the training and testing datasets. This suggests that the MEP model has a higher goodness of fit and predictive power than the other models. The comparison in terms of RMSE, RRMSE, MAE, and MAPE values for the training and testing datasets among the models revealed that the MEP model has the lowest values. This indicates that the MEP model has superior accuracy, effectiveness, and performance than ANN, RF, and MLR.

The MEP model outperforms ANN and RF in forecasting the annual MWG due to its inherent advantages. The MEP model is formulated in the form of mathematical expressions that can be interpreted to reveal the correlation between input (socioeconomic and environmental factors) and output variables (MWG) (Fallahpour et al. 2017). On the other hand, ANN and RF are characterized as black-box models since they cannot reveal the interrelationships between variables (Debnath and Mourshed 2018). This, in turn, limits the ability of policymakers and other waste management stakeholders to effectively engage in a circular economy. Despite the ability of MLR to develop a mathematical model akin to MEP, the MEP model outperforms MLR in terms of accuracy and effectiveness due to the inability of the MLR model to solve nonlinear and complex problems (Chu et al. 2021). However, the limitation of MEP as a metaheuristic method is that it may not guarantee obtaining the precise global optimum but aims to discover a near-optimal solution. Also, the performance of MEP is strongly reliant on various parameters, including population size, mutation rates, crossover rates, and the maximum number of generations. Finding the optimal combination of these parameters is often difficult and necessitates systematic experimentation.

A Mann–Whitney *U* test was carried out in MATLAB (R2023b version) using the "*ranksum*" function. Table 9 displays the *p*-value results of Mann–Whitney *U* tests for the four models, confirming the acceptance of the null hypothesis regarding the comparison of actual and predicted MWG values. The computed *p*-values are higher than the predefined significance level of 0.05. Thus, it can be concluded that there is no statistically significant difference between the actual and predicted values for the four models.

Table 9 *p*-values of Mann–Whitney *U* tests at a 95% confidence level

Model	*p*-value	H_0
MEP	0.9015	Accept
ANN	0.8048	Accept
RF	0.8048	Accept
MLR	0.9015	Accept

Null hypothesis (H_0): There is no significant difference between the actual and predicted values

Comparison with previous studies

Several researchers have applied different predictive models to predict MWG due to the dynamic availability of artificial intelligence (AI) algorithms to solve complex problems. These AI algorithms are applied to improve the performance and accuracy of predictive models. Thus, it is necessary to examine how effective the MEP model is when compared with previous studies. The comparison of MWG predictions across studies faces impediments due to differences in evaluation metrics, granularity of data (daily, monthly, and annually),

Table 10 Comparison with prior studies in terms of applied approach and evaluation metrics on the testing dataset

Authors	Approach used	Quantity of data	Evaluation metrics		
			R^2	RRMSE	MAPE (%)
Antanasijević et al. (2013)	BP-ANN	28	0.647		
Antanasijević et al. (2013)	GRNN	28	0.843		
Ribic et al. (2019)	ANN	36	0.915		
Rosecký et al. (2021)	RF	34	0.370		53.920
Liu et al. (2021)	SVR	42			45.300
Liu et al. (2021)	Logistic regression	42			27.400
Oliveira et al. (2019)	ANN	42	0.980		
Oliveira et al. (2019)	Nonlinear regression	42	0.730		
Paulauskaite-Taraseviciene et al. (2022)	Hybrid KNN	Not specified			11.050
This study	*MEP*	*34*	*0.977*	*0.016*	*1.363*

case studies, and parameter configurations of models. Nonetheless, this study aligns with others by using annual data collection, enhancing comparability among the literature.

Table 10 presents the evaluation metrics for prior and present studies that used different approaches. Antanasijević et al. (2013) applied BP-ANN and GRNN to predict the quantity of municipal solid waste, with R^2 outcomes of 0.647 and 0.843, respectively. However, this study provides a more accurate and effective MEP model with an R^2 of 0.977. Similarly, the MEP model is better in terms of accuracy and performance when compared with the study of Ribic et al. (2019), which used ANN to predict MWG with an R^2 of 0.915. The low R^2 value from Rosecký et al.'s (2021) work compared to this study might result from missing datasets at different municipal levels due to outlying data or inadequate data availability at the required spatial–temporal resolutions. These constraints can inhibit the ability of the model to truly capture the underlying computation of MWG. Also, the SVR and logistic regression applied by Liu et al. (2021) to predict the volume of municipal waste exhibited lesser accuracy with MAPE values of 45.3% and 27.4%, respectively, when compared with the current study.

Moreover, Oliveira et al. (2019) applied ANN and nonlinear regression to predict MWG, with the outcomes of R^2 being 0.98 and 0.73, respectively. In comparison with the present study, the MEP model has approximately the same R^2 value as the ANN model, but the black-box nature of the ANN model makes it difficult to determine and interpret the relationship between the input variables (socioeconomic and environmental factors) and MWG. The MEP model is more accurate and effective than the nonlinear regression model based on their respective R^2 values. The degree of accuracy and performance of the MEP model is higher than the hybrid KNN model developed by Paulauskaite-Taraseviciene et al. (2022) based on their respective MAPE values. Thus, the superior performance of the MEP model presented in this study demonstrates its robustness to other different predictive approaches for MWG.

Parametric analysis of MEP model

The parametric analysis involved the performance and behavior of the MEP model when one of the input variables (socioeconomic and environmental factors) was varied and the others were held constant at their respective mean values. The input variables were varied within the range of their maximum and minimum values. The change in that input variable (while the others were held constant at their mean values) was plotted against the resulting MWG values (using Eq. 7). The process was applied to each of the input variables individually (Althoey et al. 2023).

The influence of each input variable on MWG is presented in Fig. 5. The MWG is bound to increase as the population in municipalities increases, as illustrated in Fig. 5a. Thus, higher quantities of waste are generated due to an increase in resource consumption resulting from a larger population (Goel et al. 2017). The MWG increases as the GDP goes up, as illustrated in Fig. 5b. A higher GDP indicates enhanced economic activity, which leads to increased production and consumption rates (Goel et al. 2017). This enhanced economic activity may result in a higher generation of waste from the processes of manufacturing and consumption. DMC quantifies the entire amount of resources or materials consumed directly in an economy (Antanasijević et al. 2013). When the degree of DMC increases, the amount of municipal waste produced increases, as illustrated in Fig. 5c. As a result, more waste is produced as resources or materials are harvested, utilized, and discarded.

Higher temperatures tend to increase outdoor activities, resulting in a higher generation of waste, especially in recreational centers. On the contrary, colder temperatures reduce outdoor activities, which subsequently lower waste generation, as indicated in Fig. 5d. The MWG decreases in Fig. 5e as the UR increases. The trend is surprising, probably because as the UR increases, there are better cultural practices and modernization of improved waste management systems in Japan, thus reducing the MWG.

Sensitivity analysis of MEP model

A sensitivity analysis was performed based on the testing dataset to gain a better understanding of the impact of each socioeconomic and environmental factor (input variable) on the predicted MWG. It was used to determine the relative importance and contribution of each input variable in the established MEP model using Eqs. (14) and (15) (Chu et al. 2021).

$$I_i = \frac{N_i}{\sum_{i=1}^{n} N_i} \tag{14}$$

where,

$$N_i = f_{MAX}(X_i) - f_{MIN}(X_i) \tag{15}$$

where $f_{MAX}(X_i)$ and $f_{MIN}(X_i)$ are the maximum and minimum values of the predicted MWG values for input variable i, respectively, and n is the number of input variables. N_i was computed for each input variable separately while the other input variables were held constant at their respective mean values. I_i is the relative importance of input variable i. The result of the relative contribution of the input variables is provided in Fig. 6. The most and least important factors in the prediction of MWG via the developed MEP model are UR and temperature, respectively. The order of their importance from the most important to the least important factors in the prediction of MWG is UR, population, DMC, GDP,

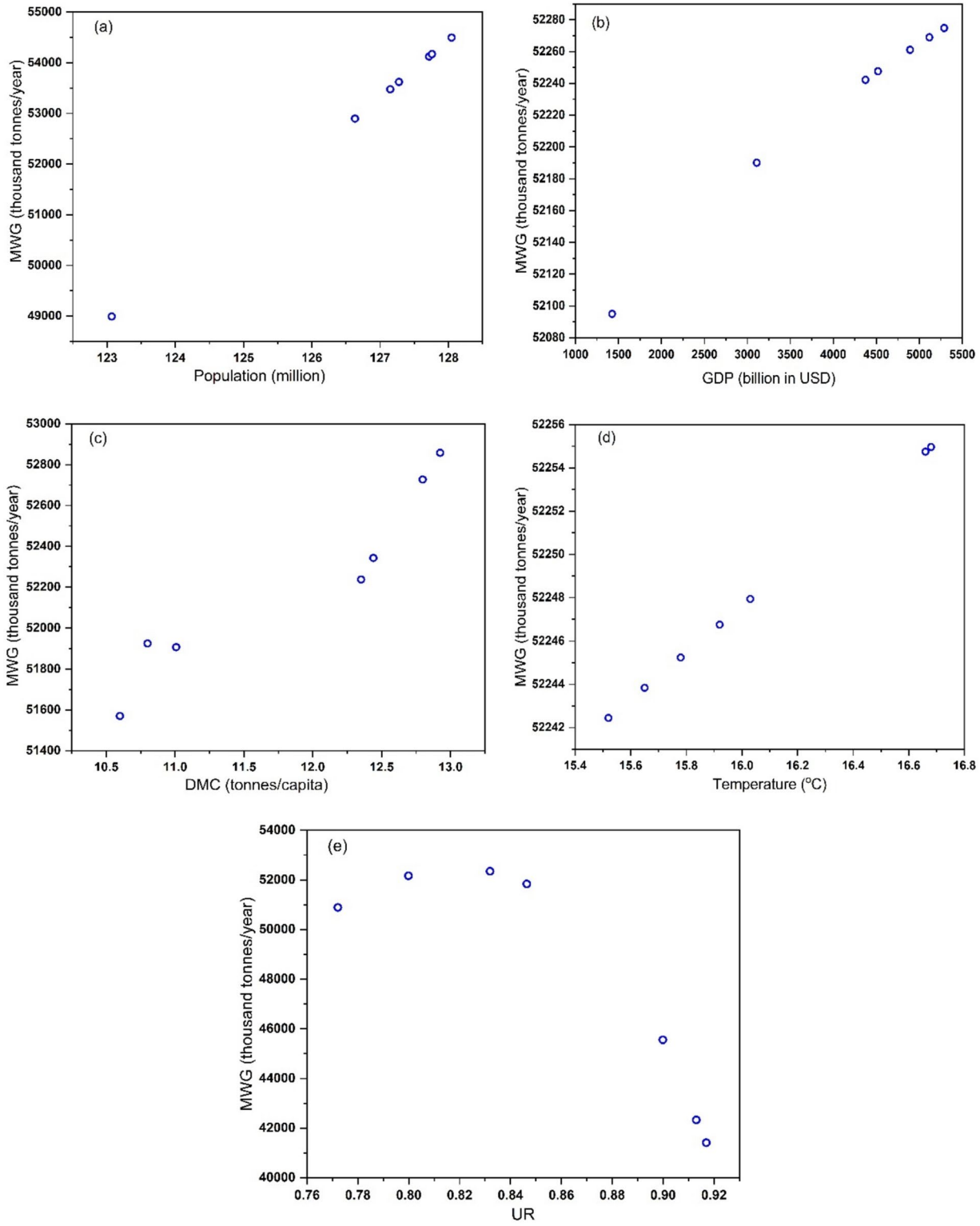


Fig. 5 The parametric analysis of the MEP model based on each input variable of **a** population, **b** GDP, **c** DMC, **d** temperature, and **e** UR

and temperature. Ultimately, recognizing UR and its consequences for waste generation is essential for policymakers, waste management authorities, and other stakeholders to develop strategies for sustainable waste management and environmental protection in municipalities for an effective circular economy to take place.

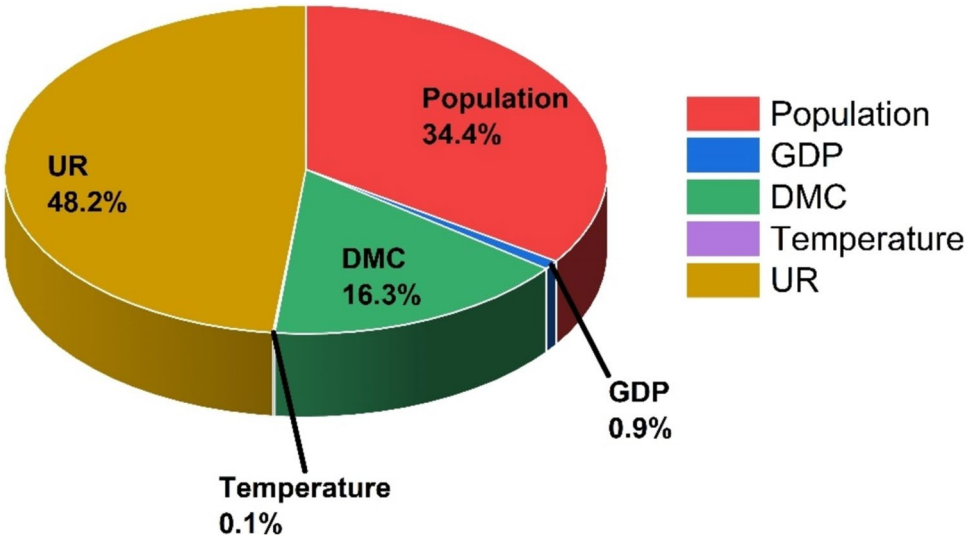


Fig. 6 Relative contribution of the input variables

Implications of research

Theoretically, the implications of this research contribute to the advancement of modeling approaches with improved predictive accuracy in the area of waste management. This study also offers theoretical insights into using the MEP model to build efficient circular economy policies by identifying critical variables and their impact on MWG to reduce waste generation and promote resource efficiency. For instance, UR is identified as the most important factor influencing MWG in this study. Policies like sustainable urban planning, waste reduction strategies, and recycling initiatives in urban areas can effectively manage UR's impact. Population is another important factor that necessitates policies such as family planning and community-based activities (such as waste reduction campaigns, waste separation, recycling programs) to reduce MWG. Regardless of the impact of DMC and GDP on MWG, policies such as incentivizing industries to adopt eco-friendly technologies and regulating resource extraction and consumption can help to reduce resource consumption and promote sustainable production practices. Climate-related initiatives, such as renewable energy adoption and sustainable agriculture, can help to alleviate environmental impacts, despite temperature having a relatively low influence on MWG in this study. The practical implications of this research include the application of the MEP model by waste management authorities to optimize waste collection, transportation, and disposal infrastructure. Waste management approaches can be more efficient and cost-effective by allocating resources for waste management activities such as waste collection routes, sorting facilities, recycling centers, and processing facilities based on predicted waste generation quantities.

Conclusions

In this study, an MEP model was developed to predict MWG using annual historical data on population, GDP, DMC, average air temperature, and UR as the input variables. The model was validated via comparative analyses with ANN, RF, and MLR models using evaluation metrics to determine the degree of accuracy and performance of the models. The Mann–Whitney *U* tests were applied to determine whether there are significant differences between the actual and predicted MWG values. Parametric and sensitivity analyses were conducted on the MEP model. The following are the key conclusions:

1. A new mathematical model for MWG prediction was established using MEP to show the complex interrelationships among the considered socioeconomic and environmental factors. This is a novel mathematical model generated through the maiden application of MEP.
2. The MEP model has the lowest values for the evaluation metrics (RMSE, RRMSE, MAE, and MAPE), which indicates its superior accuracy and performance over the other three prediction models.
3. The MEP, ANN, RF, and MLR models have R^2 values (for testing datasets) of 0.977, 0.974, 0.957, and 0.964 respectively. The comparison showed that MEP has the highest R^2 values. This suggests that the MEP model has a higher goodness of fit and predictive power than the other models.
4. The results obtained from the Mann–Whitney *U* tests revealed that there is no statistically significant difference between the actual and predicted values at a 95% confidence level.

5. The parametric analysis showed that the MWG increases as the population, GDP, DMC, and temperature increase.
6. The sensitivity analysis revealed the order of relative importance and contribution of each input variable to the prediction of MWG via the formulated MEP model as UR, population, DMC, GDP, and temperature.

This study has limitations because MEP might not guarantee the precise global optimum solution based on the difficulties in finding the optimal values of the parameters. In addition, this study did not account for uncertainty and variability in the prediction of MWG. In subsequent research, optimization methods like particle swarm optimization and genetic algorithm can be employed to effectively find the optimal parametric values for the MEP model, in order to enhance the model's performance and convergence rate. Future research can also use probabilistic modeling techniques, such as Monte Carlo simulation, to handle uncertainty and variability in the prediction of MWG.

Author contribution Ayodeji Sulaiman Olawore: conceptualization, methodology, formal analysis, validation, writing original draft, reviewing and editing.

Kuan Yew Wong: conceptualization, methodology, formal analysis, writing original draft, reviewing and editing, supervision, project administration.

Kamoru Olufemi Oladosu: conceptualization, methodology, formal analysis, reviewing and editing.

Data availability The data used in this study are available in this published article.

Declarations

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

Abbasi M, El Hanandeh A (2016) Forecasting municipal solid waste generation using artificial intelligence modelling approaches. Waste Manag 56:13–22. https://doi.org/10.1016/j.wasman.2016.05.018

Abdalla A, Salih A (2022) Implementation of multi-expression programming (MEP), artificial neural network (ANN), and M5P-tree to forecast the compression strength cement-based mortar modified by calcium hydroxide at different mix proportions and curing ages. Innov Infrastruct Solut 7(2):153. https://doi.org/10.1007/s41062-022-00761-8

Althoey F, Sor NH, Hadidi HM, Shah SFA, Alaskar A, Eldin SM, Bashir T, Javed MF (2023) Crack width prediction of self-healing engineered cementitious composite using multi-expression programming. J Mater Res Technol 24:918–927. https://doi.org/10.1016/j.jmrt.2023.03.036

Antanasijević D, Pocajt V, Popović I, Redžić N, Ristić M (2013) The forecasting of municipal waste generation using artificial neural networks and sustainability indicators. Sustain Sci 8(1):37–46. https://doi.org/10.1007/s11625-012-0161-9

Asghar R, Javed MF, Saqib M, Alaskar A, Ali M, Nawaz R (2024) Multi-expression programming based prediction of the seismic capacity of reinforced concrete rectangular columns. Eng Appl Artif Intell 131:107834. https://doi.org/10.1016/j.engappai.2023.107834

Bijos JCBF, Zanta VM, Morató J, Queiroz LM, Oliveira-Esquerre KPSR (2022) Improving circularity in municipal solid waste management through machine learning in Latin America and the Caribbean. Sustain Chem Pharm 28:100740. https://doi.org/10.1016/j.scp.2022.100740

Caudill M (1988) Neural networks primer, Part III. AI Expert 3:53–59

Chhay L, Reyad MAH, Suy R, Islam MR, Mian MM (2018) Municipal solid waste generation in China: influencing factor analysis and multi-model forecasting. J Mater Cycles Waste Manag 20(3):1761–1770. https://doi.org/10.1007/s10163-018-0743-4

Chu H-H, Khan MA, Javed M, Zafar A, Ijaz KM, Alabduljabbar H, Qayyum S (2021) Sustainable use of fly-ash: use of gene-expression programming (GEP) and multi-expression programming (MEP) for forecasting the compressive strength geopolymer concrete. Ain Shams Eng J 12(4):3603–3617. https://doi.org/10.1016/j.asej.2021.03.018

Debnath KB, Mourshed M (2018) forecasting methods in energy planning models. Renew Sustain Energy Rev 88:297–325. https://doi.org/10.1016/j.rser.2018.02.002

Dobre-Baron O, Nițescu A, Niță D, Mitran C (2022) Romania's perspectives on the transition to the circular economy in an EU context. Sustainability 14(9):5324. https://doi.org/10.3390/su14095324

Fallahpour A, Olugu EU, Musa SN (2017) A hybrid model for supplier selection: integration of AHP and multi expression programming (MEP). Neural Comput Appl 28(3):499–504. https://doi.org/10.1007/s00521-015-2078-6

Fallahpour A, Wong KY, Rajoo S, Tian G (2021) An evolutionary-based predictive soft computing model for the prediction of electricity consumption using multi expression programming. J Clean Prod 283:125287. https://doi.org/10.1016/j.jclepro.2020.125287

Fan L, Abbasi M, Salehi K, Band SS, Chau K-W, Mosavi A (2021) Introducing an evolutionary-decomposition model for prediction of municipal solid waste flow: application of intrinsic time-scale decomposition algorithm. Eng Appl Comput Fluid Mech 15(1):1159–1175. https://doi.org/10.1080/19942060.2021.1945496

Goel S, Ranjan VP, Bardhan B, Hazra T (2017) Forecasting solid waste generation rates. In Modelling trends in solid and hazardous waste management, Springer Singapore, pp 35–64. https://doi.org/10.1007/978-981-10-2410-8_3

He R, Sandoval-Reyes M, Scott I, Semeano R, Ferrão P, Matthews S, Small MJ (2022) Global knowledge base for municipal solid waste management: framework development and application in waste generation prediction. J Clean Prod 377:134501. https://doi.org/10.1016/j.jclepro.2022.134501

Hoy ZX, Woon KS, Chin WC, Hashim H, Fan YV (2022) Forecasting heterogeneous municipal solid waste generation via Bayesian-optimised neural network with ensemble learning for improved generalisation. Comput Chem Eng 166:107946. https://doi.org/10.1016/j.compchemeng.2022.107946

Inoue T (2022) Statistical handbook of Japan. https://www.stat.go.jp/english/data/handbook/index.html. Accessed 5 Mar 2024

Iqbal MF, Javed MF, Rauf M, Azim I, Ashraf M, Yang J, Liu Q (2021) Sustainable utilization of foundry waste: forecasting mechanical

properties of foundry sand based concrete using multi-expression programming. Sci Total Environ 780:146524. https://doi.org/10.1016/j.scitotenv.2021.146524

Jäämaa L, Kaipia R (2022) The first mile problem in the circular economy supply chains – collecting recyclable textiles from consumers. Waste Manag 141:173–182. https://doi.org/10.1016/j.wasman.2022.01.012

Jaf IDK, Abdalla A, Mohammed AS, Abdulrahman PI, Rawaz K, Mohammed AA (2024) Hybrid nonlinear regression model versus MARS, MEP, and ANN to evaluate the effect of the size and content of waste tire rubber on the compressive strength of concrete. Heliyon 10(4):e25997. https://doi.org/10.1016/j.heliyon.2024.e25997

Keser S, Duzgun S, Aksoy A (2012) Application of spatial and non-spatial data analysis in determination of the factors that impact municipal solid waste generation rates in Turkey. Waste Manag 32(3):359–371. https://doi.org/10.1016/j.wasman.2011.10.017

Kubule A, Klavenieks K, Vesere R, Blumberga D (2019) Towards efficient waste management in Latvia: an empirical assessment of waste composition. Environ Clim Technol 23(2):114–130. https://doi.org/10.2478/rtuect-2019-0059

Kulisz M, Kujawska J (2020) Prediction of municipal waste generation in Poland using neural network modeling. Sustainability 12(23):10088. https://doi.org/10.3390/su122310088

Langsetmo L, Schousboe JT, Taylor BC, Cauley JA, Fink HA, Cawthon PM, Kado DM, Ensrud KE (2023) Advantages and disadvantages of random forest models for prediction of hip fracture risk versus mortality risk in the oldest old. JBMR Plus 7(8). https://doi.org/10.1002/jbm4.10757

Li K, Ma H (2019) Prediction of municipal solid waste generation with Elman Neural Network-case study: Shanghai City in China. 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) 1174–1177. https://doi.org/10.1109/ITNEC.2019.8729395

Liu B, Zhang L, Wang Q (2021) Demand gap analysis of municipal solid waste landfill in Beijing: based on the municipal solid waste generation. Waste Manag 134:42–51. https://doi.org/10.1016/j.wasman.2021.08.007

Liu W, Hu X, Lin X, Yang X-G, Song Z, Foley AM, Couture J (2022) Toward high-accuracy and high-efficiency battery electrothermal modeling: a general approach to tackling modeling errors. eTransportation 14:100195. https://doi.org/10.1016/j.etran.2022.100195

Ma X, Chen Z, Chen P, Zheng H, Gao X, Xiang J, Chen L, Huang Y (2023) Predicting the utilization factor of blasthole in rock roadways by random forest. Undergr Space 11:232–245. https://doi.org/10.1016/j.undsp.2023.01.006

MacFarland TW, Yates JM (2016) Mann–Whitney U test. In Introduction to nonparametric statistics for the biological sciences using R, Springer International Publishing, pp 103–132. https://doi.org/10.1007/978-3-319-30634-6_4

Montgomery DC, Runger GC (2018) Applied statistics and probability for engineers. Wiley, New York

Nguyen H, Bui X-N, Bui H-B, Mai N-L (2020) A comparative study of artificial neural networks in predicting blast-induced air-blast overpressure at Deo Nai open-pit coal mine Vietnam. Neural Comput Appl 32(8):3939–3955. https://doi.org/10.1007/s00521-018-3717-5

OECD (2015) Municipal waste. OECD Environment Statistics (database). https://doi.org/10.1787/data-00601-en

OECD (2024) Municipal waste (indicator). https://doi.org/10.1787/89d5679a-en

Oliveira V, Sousa V, Dias-Ferreira C (2019) Artificial neural network modelling of the amount of separately-collected household packaging waste. J Clean Prod 210:401–409. https://doi.org/10.1016/j.jclepro.2018.11.063

Oltean M, Dumitrescu D (2002) Multi expression programming. www.mepx.org. Accessed 7 Mar 2024

Öztürk OB, Başar E (2022) Multiple linear regression analysis and artificial neural networks based decision support system for energy efficiency in shipping. Ocean Eng 243:110209. https://doi.org/10.1016/j.oceaneng.2021.110209

Paulauskaite-Taraseviciene A, Raudonis V, Sutiene K (2022) Forecasting municipal solid waste in Lithuania by incorporating socioeconomic and geographical factors. Waste Manag 140:31–39. https://doi.org/10.1016/j.wasman.2022.01.004

Petković B, Agdas AS, Zandi Y, Nikolić I, Denić N, Radenkovic SD, Almojil SF, Roco-Videla A, Kojić N, Zlatković D, Stojanović J (2021) Neuro fuzzy evaluation of circular economy based on waste generation, recycling, renewable energy, biomass and soil pollution. Rhizosphere 19:100418. https://doi.org/10.1016/j.rhisph.2021.100418

Puntarić E, Pezo L, Zgorelec Ž, Gunjača J, Kučić GD, Voća N (2022) Prediction of the production of separated municipal solid waste by artificial neural networks in Croatia and the European Union. Sustainability 14(16):10133. https://doi.org/10.3390/su141610133

Rao KDVSK, Premalatha M, Naveen C (2018) Analysis of different combinations of meteorological parameters in predicting the horizontal global solar radiation with ANN approach: a case study. Renew Sustain Energy Rev 91:248–258. https://doi.org/10.1016/j.rser.2018.03.096

Ribic B, Pezo L, Sincic D, Loncar B, Voca N (2019) Predictive model for municipal waste generation using artificial neural networks—case study City of Zagreb. Croatia Int J Energy Res 43(11):5701–5713. https://doi.org/10.1002/er.4632

Rosecký M, Šomplák R, Slavík J, Kalina J, Bulková G, Bednář J (2021) Predictive modelling as a tool for effective municipal waste management policy at different territorial levels. J Environ Manag 291:112584. https://doi.org/10.1016/j.jenvman.2021.112584

Sahoo S, Jha MK (2013) Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment. Hydrogeol J 21(8):1865–1887. https://doi.org/10.1007/s10040-013-1029-5

Shao Z, Jahed AD, Yazdani BB, Mu'azu MA, Tonnizam ME (2019) Estimating the friction angle of black shale core specimens with hybrid-ANN approaches. Measurement 145:744–755. https://doi.org/10.1016/j.measurement.2019.06.007

Wang HL, Yin ZY (2020) High performance prediction of soil compaction parameters using multi expression programming. Eng Geol 276:105758. https://doi.org/10.1016/j.enggeo.2020.105758

WorldBank (2024) Population. data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS. Accessed on 12 February 2024

Yano J, Sakai S (2016) Energy recovery and greenhouse gas reduction potential from food waste in Japan. J Mater Cycles Waste Manag 18(4):631–645. https://doi.org/10.1007/s10163-016-0531-y

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.