



Machine learning-based prediction of construction and demolition waste generation in developing countries: a case study

Milad Jafari¹ · Ehsan Mousavi²

Received: 3 October 2023 / Accepted: 24 July 2024

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Data is needed for making informed decisions regarding managing waste in the time of construction and demolition phases of buildings. However, data availability is very limited in most developing countries in the area of waste generation. The objective of this study is to employ an artificial intelligence (AI)-based approach to develop a reliable model for forecasting monthly construction and demolition waste (C&DW) generation in the case study of Tehran, Iran. We have trained different prediction models using various AI algorithms, including multilayer perceptron neural network, radial basis function neural network, support vector machines, and adaptive neuro-fuzzy inference system (ANFIS). According to the findings, all employed AI algorithms demonstrated high prediction performance for C&DW forecasting models. The ANFIS model, with $R^2=0.96$ and $RMSE=0.04209$, was identified as the model that better represented the observed values of C&DW generation. The better efficiency of the ANFIS model could be due to its effective enhancement of neural networks to model subjective variables based on fuzzy logic capabilities. The developed prediction model can be employed as an efficient tool for policy and decision-making for C&DW management by predicting waste quantities in the future.

Keywords Construction and demolition waste · Machine learning · ANFIS · ANN · C&DW prediction · Developing Countries · ML · SVM

Introduction

The problem of waste creation, particularly construction and demolition, is a worldwide problem that has been studied extensively. Construction and demolition operations have expanded as a result of rapid urbanization, resulting in massive amounts of waste. C&DW generation is recognized as one of the primary challenges in the construction sector because of its substantial environmental effects as well as the industry's efficacy (Jain 2021). It can impact our health and the environment we live. The building industry utilizes

roughly 40% of all materials produced globally (Jafari et al. 2024). Even though most C&DW created are non-toxic, they are incredibly troublesome because they are created in substantial quantities (Gao et al. 2018). Nevertheless, C&DW can represent a severe danger to the environment, with negative consequences such as increased energy use, greenhouse gas (GHG) emissions, resource depletion, and land degradation (Khoshand et al. 2020; Fakhri et al. 2023).

C&DW includes a diverse set of materials that are created in different phases of the building lifecycle, from construction through demolition. They can be categorized as inert, non-inert, non-hazardous, and hazardous waste (Białko 2018). Furthermore, C&DW includes materials that may be formed unexpectedly due to natural disasters, like earthquakes, hurricanes, and floods (Umar et al. 2017). The amount and nature of C&DW vary by region and are influenced by population growth, law, regional planning, and the country's construction sector (Menegaki and Damigos 2018). C&DW generation is expected to rise in response to the global expansion of construction activities. Given the scarcity of landfill space, it is evident that reducing and managing C&DW is critical (Khoshand et al. 2020). Otherwise,

Responsible Editor: Philippe Garrigues

✉ Milad Jafari
miladj@clemson.edu

Ehsan Mousavi
mousavi@clemson.edu

¹ Department, of Construction Science and Management, Clemson University, 1-171 Lee Hall, Clemson, SC, USA

² Department of Construction Science and Management, Clemson University, 2-132 Lee Hall, Clemson, SC, USA

inefficient waste management may have an influence on the rising amount of demolition waste. This would add to the difficulties faced by solid waste management programs, which are already looking for innovative solutions to deal with increasing solid waste in cities which is mainly due to urban population growth. To address these concerns, we needed more accurate and valuable statistics on C&DW generation (Jain 2021).

Measuring the volume of C&DW produced is widely acknowledged as necessary for designing and executing management systems at the project or regional level (Cai et al. 2020). Due to large volumes of waste and limited landfill capacity, anticipating possible C&DW generation on a regional scale will assist the government in estimating current landfill capacity and enacting legislation to cope with it (Parisi Kern et al. 2015). In-depth research has been done on time series prediction methods in order to produce accurate different types of waste forecasts. Classic approaches based on AI methods and mathematical-statistical models are the two types of techniques that are commonly used (Abbasi and El Hanandeh 2016). Traditional approaches, on the other hand, are unsuitable for time series prediction with more complexity, such as nonlinearity and irregularity, for which AI methods may be more favorable.

Cochran and Townsend (2010) proposed a technique for analyzing material flow from manufacturing through destruction. They calculated C&DW using historical data on construction material use and average lifetimes. The suggested approach was used to anticipate the volume of C&DW in a vast area of the US. Martínez Lage et al. (2010) proposed C&DW quantification by comparing the population with information on new constructions, restorations, and demolitions in the area of Galicia, Spain. The primary goal of the study of Akanbi et al. (2018) was to establish a framework based on BIM for calculating building elements' salvage value over the lifespan of a building. However, the models' utility is limited since nearly all buildings slated for demolition and restoration lack a BIM model.

Data-driven AI approaches have been employed to map various impacting factors to C&DW production quantity. The effectiveness of these approaches is in their capability to model non-linear relationships among the parameters without having prior knowledge about the relationship expression or shape (Hu et al. 2021). Lu et al. (2021) evaluated the advantages and disadvantages of several waste quantification models for forecasting C&DW generation in the Great Bay Area (GBA) in China. From 2005 to 2019, 43 sets of yearly socio-economic, construction-related, and C&DW production statistics were gathered from local government bodies. All examined algorithms, including multiple linear regression, ANN, and decision tree, were demonstrated to achieve satisfactory results. Long short-term memory networks (LSTM) were used by Huang et al. (2020) to predict the

quantity of C&DW created in upcoming years derived from regional historic data from 1980 to 2011. Cha et al. (2020) investigated the Random Forest (RF) algorithm to generate demolition waste prediction models. Categorical and continuous data were included in their databases. They showed that in spite of their small dataset, the RF algorithm was able to demonstrate a consistent forecasting performance. A study was done by Song et al. (2017) with the goal of estimating the yearly quantity of C&DW in China by integrating the gray model and support vector regression. After estimating the yearly total construction area, a transition matrix was utilized to calculate the C&DW amounts. Future C&DW quantity was enumerated and examined using the suggested technique, including prospective components and distribution in different Chinese provinces.

By using deep learning models, Akanbi et al. (2020) tried to estimate the material waste quantity produced during the service life of the buildings. Two thousand two hundred eighty data on building demolition gathered from UK demolition industry practitioners are employed to train a deep neural network model. Their findings reveal that given fundamental building characteristics, it is feasible to anticipate the amount of materials retrieved from a structure following demolition with high accuracy. A study was done by Oliveira et al. (2019) with the goal of estimating the yearly quantity of packaging waste of households developing ANN and genetic algorithms. The R^2 value of the proposed ANN model, applied to a dataset comprising 42 cities in central Portugal, was 0.98. They also compared their ANN and conventional regression approaches, demonstrating the outperformance of ANN.

By integrating principal component analysis with decision trees, k -nearest neighbors (KNN), and linear regression methods, Cha et al. (2023) established a hybrid model for predicting the demolition-waste-generation rate in redevelopment regions in South Korea. The highest accuracy was obtained by employing PCA to the KNN algorithm, with an R^2 of 0.897. To build a machine learning model that can precisely forecast the quantity of CW produced at different phases and from varying sources, Gulghane et al. (2023) attempted to quantify the amount of CW at various stages of the construction process using KNN and Decision Trees (DT). The two models were adequately foreseeing the CW generation at each stage, as evidenced by their combined accuracy of about 90% on average.

In the recent literature on waste management, fuzzy inference algorithms have been utilized extensively (Abbasi and El Hanandeh 2016; Khoshand et al. 2023). Fuzzy rules are generated throughout the inference's data training process, and fuzzy logic handles them. An efficient and natural method for helping people with the justification and decision-making processes is the generation of information from a trained database using fuzzy rules (Khoshand et al.

2023). Using the fuzzy logic technique, Chhay et al. (2018) explored the influencing socio-economic aspects of municipal solid waste (MSW) creation in China, and short-term anticipation of MSW generation was undertaken using a multi-model approach. Their work suggests that the main socio-economic element for MSW generation is urban population expansion, while the impact of GDP on waste creation is less clear. A recent study by Ghanbari et al. (2021) uses the Pearson correlation analysis to find the most crucial factors affecting Solid Waste Generation (SWG) in Tehran, Iran. They have shown that income, GDP, population, and month are the essential variables for the monthly prediction of SWG.

The major goal of the study of Abbasi and El Hanandeh (2016) was to create a model for precise prediction of MSW creation that would assist waste management organizations in better designing and operating successful MSW management systems. According to the findings, the most precise peak predictions came from the ANFIS system. Adeleke et al. (2022) used South Africa as a case study to evaluate how the parameters of three clustering techniques, fuzzy *c*-means (FCM), grid partitioning (GP), and subtractive clustering (SC), affect the efficiency of the ANFIS model to forecast waste generation. According to their findings, the best model is an ANFIS model clustered with GP, which uses a triangular input and linear type output membership function.

Although there are some studies trying to predict different waste types (such as MSW and electronic waste) by application of AI (Abdallah et al. 2020), the number of studies on the prediction of C&DW generation is limited. Previous works (Abbasi and El Hanandeh 2016; Adeleke et al. 2022) have demonstrated that ANFIS has an outstanding performance in forecasting MSW generation. No study was found in the literature to use the ANFIS algorithm to predict C&DW generation. Due to the different nature of C&DW, the primary aim of this research is to evaluate the performance of the ANFIS algorithm in predicting monthly C&DW generation in the region of Tehran, Iran. Because of lacking reliable waste data in most developing countries like Iran, the outcome of this study will help plan a more efficient waste management program. Moreover, the findings will be compared with the results of ANN and SVM algorithms to better evaluate the ANFIS algorithm's efficiency in predicting C&DW generation.

Methodology

A brief explanation of the data and procedures of this research is provided in the subsections that follow. First, a brief explanation of the area of study is presented. Second, the process of collecting the data and selecting different variables is explained. Next, brief information about ANN,

SVM, and ANFIS algorithms is provided. Finally, the accuracy metrics used in this study are presented.

Study area

The study area is Tehran, Iran, with a population of around 8.7 million people living in the city. Also, Tehran has a large metropolitan population of 15 million people. The city is situated on the central Alborz mountain range's slopes and has a total area of 700 km². The city's altitude fluctuates, with 1800 m in the north, 1200 m in the middle, and 1050 m in the south (Ghanbari et al. nd). Tehran is considered a megacity with 22 central districts. The study map with its districts is shown in Fig. 1.

Since 2004, Iran has had a general waste management law in place, which covers C&DW management as well. It should be noted this general rule is inapplicable to C&DW management. Insufficient funding dedicated to C&DW management, absence of an efficient C&DW management plan, lack of rules in C&DW management field, shortage of skillful employees in executing a C&DW management plan, and restricted public engagement are all issues that plague C&DW management in Tehran (Khoshand et al. 2020). Furthermore, there is a scarcity of official C&DW statistics data. According to research done by Asgari et al. (2017), the created C&DW in the city of Tehran is mostly made up of mixed sand and cement (30%), concrete (19%), broken bricks (18%), and soil (11%). Only around a quarter of the C&DW produced gets recycled, with the majority of the rest is dumped in a Tehran's south landfill, Aradkooh landfill, that is situated in Tehran's 18th district.

Data collection, pre-processing, and variable selection

The quantity of C&DW produced is influenced by a variety of factors. When constructing a reliable prediction model in real-world scenarios, selecting the most appropriate input variables becomes a crucial challenge. The factors that impact C&DW can generally be categorized into two types: construction-related factors (e.g., total construction output, floor space of newly started buildings, and floor space completed) (Lu et al. 2021) and socio-economic factors (e.g., population and gross domestic product) (Zhao et al. 2011). Although it is challenging to determine an exact estimate of C&DW generation, considering these factors, sufficient data availability, and proper analytics can lead to a reasonable forecast. The factors affecting C&DW creation were carefully chosen based on relevant literature and data availability, as detailed in Table 1.

The amount of monthly C&DW was collected from the Tehran Waste Management Organization from March 2017 to January 2022. Due to the lack of access to monthly data

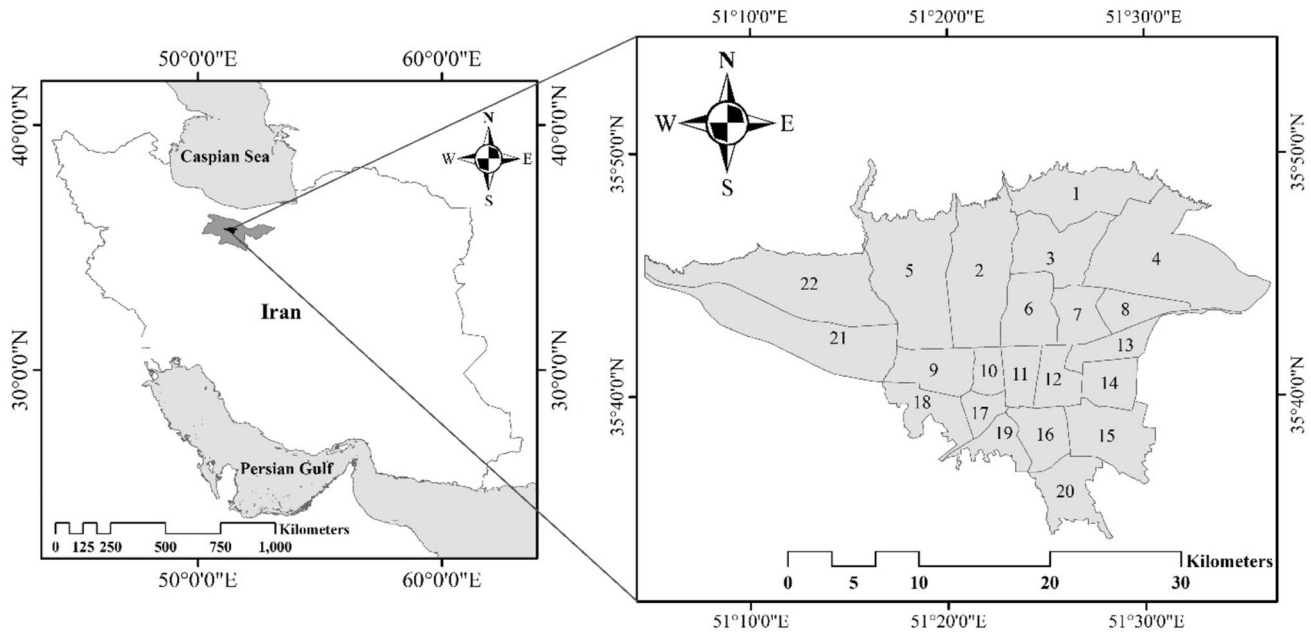


Fig. 1 Location of the study area

Table 1 The utilized parameters

| Parameter | Abbreviation | Unit | Source |
|--------------------------------------|----------------|----------------|--|
| Population | POP | – | Management and Planning Organization of Tehran |
| Gross domestic product per capita | GDP per capita | USD | World Bank |
| Month | – | – | – |
| Floor space of newly begun buildings | FB | m ² | Tehran Municipality Organization |
| Floor space of completed buildings | FC | m ² | Tehran Municipality Organization |
| Number of new construction permits | NCP | – | Tehran Municipality Organization |

on GDP per capita and population for Tehran city, annual data on these two variables are used for model development, while all remaining features are on a monthly basis.

The next section presents the Pearson correlation coefficient computed between each input and the output of the model. The correlation analysis may be used to filter components before they are modeled further (Ghanbari et al. [nd](#)). The correlations might be regarded as inconsequential when the Pearson correlation coefficient's absolute value is below 0.3 (Lu et al. [2021](#)).

Each input sample has a matching output label in supervised learning, and algorithms are trained on labeled data (Boroujeni et al. [2024](#)). Overfitting is an issue that might affect supervised learning algorithms in general (Abbasi et al. [2019](#)). This indicates that these algorithms have the potential to produce models that are excessively tailored to the noise and complexities present in the training data, leading to poor performance when applied to new and

unseen data. In order to prevent overfitting, 80% of data (47 months) were chosen at random to serve as training models, and the rest of the data (11 months) were used for testing developed models.

It is worth noting that using raw data might lead machine learning models to lose accuracy. As a result, all input data of this research were pre-processed. Pre-processing data improves accuracy while decreasing computing performance. Data pre-processing aims to decrease the input dataset size, establish smoother relationships, and eliminate noisy data (Khoshand [2021](#)). Several strategies for data pre-processing have been established in the literature, including z -score normalization, scaling normalization, and min–max normalization (Khoshand [2021](#)). The min–max normalization approach was used in the current investigation. All the datasets underwent normalization to scale them within the range of zero to one using the following formula:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where X is the variable and X_{max} and X_{min} are the variable's maximum and lowest values, respectively.

Artificial neural network (ANN)

The concept of employing ANN in computer model development originates from drawing parallels with the functioning and design of the brain and central nervous system. An ANN resembles a parallel computer since it comprises a multitude of interconnected processing elements (Sunayana et al. 2021). The input and output of a neural network are used to train it. An ANN's primary layer comprises input neurons that provide data to the hidden layer, which then sends the processed data to the third layer's output neurons. The neurons' number in the hidden layer is determined through a process of trial and error.

This study utilizes two well-known types of ANNs, namely the multilayer perceptron (MLP) and the radial basis function (RBF). The MLP is the most often utilized artificial neural network, particularly in environmental investigations (Šajin et al. 2022). This approach may be used to solve problems involving feature matching and pattern recognition. This research employs a single hidden layer feed-forward neural network. The multilayer perceptron neural network (MLPNN) design has been used several times and has been shown to be effective (Lu et al. 2021). This ANN model can solve issues of any complexity if the single hidden layer has sufficient neurons. Within MLPNN, the activation function was the sigmoid transfer function (Kannangara et al. 2018). The Levenberg–Marquardt backpropagation technique was used to train the neural network (Lu et al. 2021).

The radial basis function neural network (RBFNN) is a forward-type network known for its ability to approximate any nonlinearity. It comes with notable benefits, including a straightforward structure, quick training rate, strong ability to locally approximate, and high accuracy in its approximations (Xiaonan et al. 2020). Broomhead and Lowe (1988) introduced the RBF network, which employed the RBF as an activation function in neural network construction. The concept of approximating arbitrary continuous functions served as the inspiration for developing the RBF network, which is a specialized and adaptive type of neural network. In this study, the σ value, representing the spread of function, was set to 100. Through a process of trial and error, the ideal number of neurons in the hidden layer was established. The maximum training cycle was limited to $l=200$, and the learning rate for the weight, center, and width parameters was set to $\eta_1 = \eta_2 = \eta_3 = 0.001$. Additionally, the minimum error accuracy was defined as 0. The ANN model utilized in this investigation is schematically shown in Fig. 2.

Support vector machine (SVM)

SVM is a versatile binary classification method that searches for identifying the optimal hyperplane in a multi-dimensional space. The goal is to optimize the spacing between the nearest data sample and the hyperplane (Guo et al. 2021). Support vectors are a small subset of training observations employed to determine the best position for decision surfaces (Ayeleru et al. 2021). SVM was designed initially to address classification issues, but it was discovered to beat numerous traditional regression techniques; therefore, it was adapted to solve regression problems. Overfitting is less likely to occur with support vector regression techniques. They are capable of reducing both error estimates and model dimensions at the same time (Abdallah et al. 2020).

The foundation of training these algorithms involves locating a hyperplane in the feature space for data modeling. Specimens within the epsilon distance from this plane are considered to exhibit similar behavior, while their distance from this plane influences the behavior of other specimens (ξ) (Golbaz et al. 2019). The location of this plane is determined by points known as support vectors. Different kernels (equations), including linear, polynomial, radial basis, and sigmoid, are used in this research to describe this plane.

Adaptive neuro-fuzzy inference system (ANFIS)

A feed-forward network is employed by the data-oriented model ANFIS to explore a fuzzy membership function that relates inputs and outputs (Younes et al. 2015). The power of fuzzy logic and artificial neural networks are combined in ANFIS. Membership functions, model inputs, and fuzzy rule generation are all used to determine the fuzzy logic structure. In this work, the most basic and common Takagi–Sugeno technique of fuzzy inference structure system, which possesses training data validation capacity, was applied (Abbasi and El Hanandeh 2016). Figure 3 depicts a two-rule ANFIS system having a single output and two inputs. The desired input/output connection is acquired by changing the membership function's shape through the training stage. This process was repeated 100 epochs until sufficient convergence was attained. The fuzzy rules are presented as if–then rules, with input and output membership functions defined.

For the proposed system in Fig. 3, rules are as follows:

$$\text{If } x \text{ is } A_1 \text{ and } y \text{ is } B_1, \text{ then } f_1 = k_1x + l_1y + r_1 \quad (2)$$

$$\text{If } x \text{ is } A_2 \text{ and } y \text{ is } B_2, \text{ then } f_2 = k_2x + l_2y + r_2 \quad (3)$$

A_i and B_i are fuzzy sets, f_i is the output inside the fuzzy area defined by the fuzzy rule, and k_i , l_i , and r_i are design parameters obtained during the training phase. The degree of

Fig. 2 The structure of ANN model

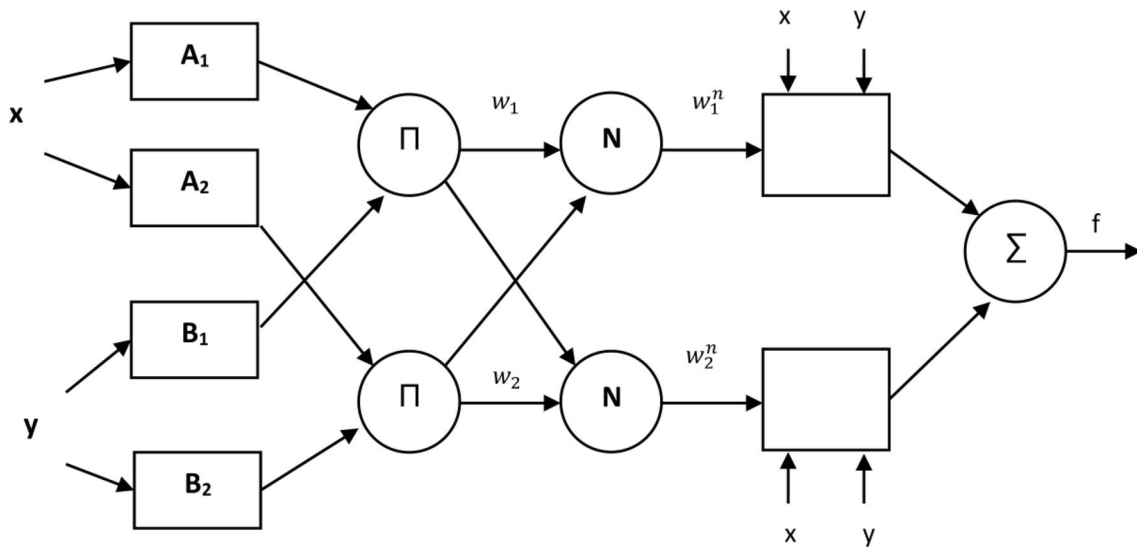
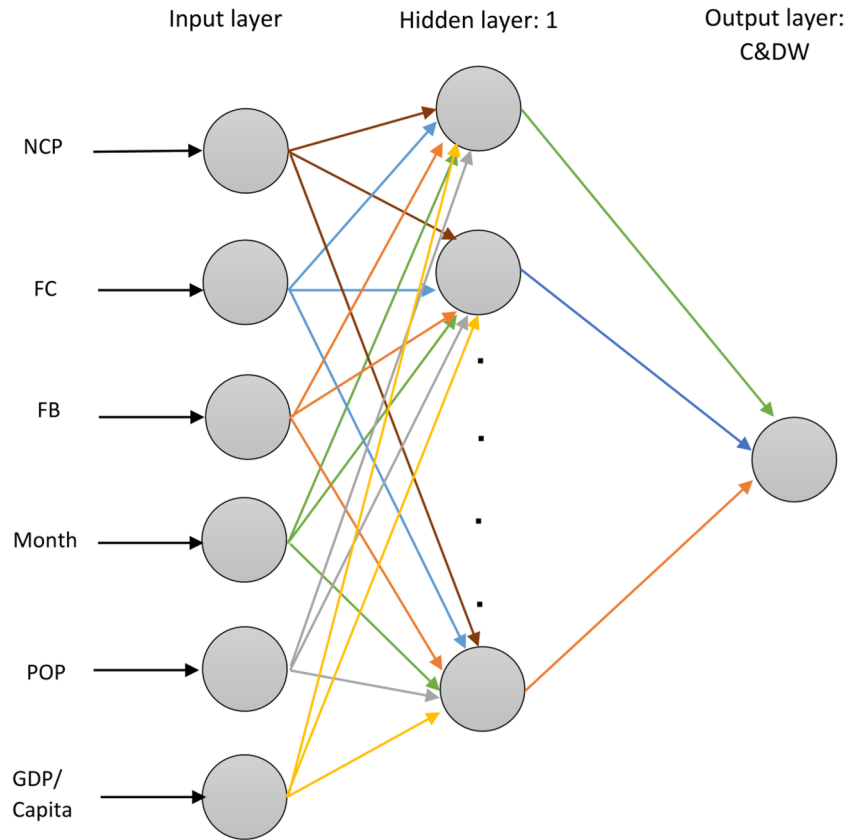


Fig. 3 An ANFIS architecture Takagi–Sugeno system

any rule activation in the second layer is calculated by each node. This layer then multiplies the membership functions:

$$w_i = \mu_{A_i}(x) * \mu_{B_i}(y) \quad i = 1,2 \tag{4}$$

where $\mu_A(x)$ represents membership degree of x in A sets and also $\mu_B(y)$ represents the membership degree of y in B sets.

The average nodes are found in the third layer. The normalized firing strength of each rule is output by the nodes of this layer.

$$\bar{w}_i = \frac{w_i}{w_1 + w_2} \quad i = 1,2 \tag{5}$$

The nodes of the fourth layer calculate the output of the model as it relates to the *i*th rule.

$$\bar{w}_i f = \bar{w}_i (k_i x + l_i y + r_i) \quad i = 1,2 \tag{6}$$

where \bar{w}_i is the outcome of the third layer and *k*, *l*, and *r* are changeable consequent parameters.

The output node, which conducts the summing of all input signals and calculates the final output, makes up the final layer.

$$\text{Final output} = \frac{\sum w_i f_i}{\sum w_i} \tag{7}$$

ANFIS organizes data into related fuzzy clusters, assigns membership functions, and builds the structure of the fuzzy inference system from the data using clustering methods (Adeleke et al. 2022). Grid partitioning, fuzzy *C*-means, and subtractive clustering are some of the most frequent clustering algorithms used in ANFIS. The one-pass subtractive clustering technique is utilized to classify data in this study. It is a rapid approach that uses the cluster radius to determine the number of clusters and their centers in a data set. Using trial and error, the radius of each data cluster was optimized between 0.1 and 0.95 (Abbasi and El Hanandeh 2016). Membership functions come in various shapes, including the Gaussian, the generalized bell, the Z-shape, the sigmoid, the trapezoidal, and the triangle functions. One-pass subtractive clustering uses the Gaussian function to calculate the membership function’s degree related to input variables. Linear and constant output membership function types were tested.

Evaluation of model performance

To evaluate the effectiveness of prediction models, a variety of accuracy metrics can be utilized (Oliveira et al. 2019). These accuracy measurements represent the disparity between the model’s generated values and the actual values. The performance of the proposed models was evaluated

using the root-mean-square error (RMSE) and coefficient of determination (R^2). The R^2 is a statistic that expresses the proportion of the model’s initial uncertainty. $R^2 = 1$ indicates that the anticipated and observed values are perfectly aligned, which is extremely rare to occur. The most precise model is the one with the lowest RMSE, which measures the discrepancy between anticipated and actual values. For the test data, the stated statistical criteria are determined as follows:

$$\text{RMSE} = \sqrt{\sum_{i=1}^n \frac{(X_t - X_0)^2}{n}} \tag{8}$$

$$\text{Coefficient of determination } (R^2) = 1 - \frac{\sum (X_t - X_0)^2}{\sum (X_t - \bar{x}_t)^2} \tag{9}$$

where X_t is the known output and X_0 denotes the anticipated output, *n* denotes the output's number, and \bar{x}_t denotes the average of the known output. All calculations were carried out using Windows 7 operating system and MATLAB software (Version R2020a).

Result and discussion

Selecting the optimal input variables is a critical task when creating a dependable prediction model in real-world scenarios. With a small input variable number, the properties of C&DW generation cannot be reliably identified. Overfitting issues might also result from having many input variables. Prior to designing the predictive model, the Pearson correlation analysis was carried out to identify the influence of each input variable on C&DW generation. As shown in Table 2, all of these values are more than 0.4 in our investigation, indicating that these parameters are capable of being used for modeling. Month, POP, GDP per capita, FC, NCP, and FB are the strongest correlations in order.

To assess performance of ANFIS algorithm in prediction of C&DW, this study simulated monthly C&DW generation using four AI models: ANFIS, RBFNN, MLPNN, and SVM. We needed to find the optimum model structure for each

Table 2 Variable correlations using the Pearson correlation coefficient

| | <i>FC</i> | <i>NCP</i> | <i>FB</i> | <i>Month</i> | <i>GDP/ Capita</i> | <i>POP</i> | <i>C&W waste</i> |
|----------------------|-----------|------------|-----------|--------------|--------------------|------------|----------------------|
| <i>FC</i> | 1 | | | | | | |
| <i>NCP</i> | 0.319439 | 1 | | | | | |
| <i>FB</i> | 0.509584 | 0.5 | 1 | | | | |
| <i>month</i> | 0.422629 | -0.6265 | -0.50096 | 1 | | | |
| <i>GDP/ Capita</i> | 0.136924 | 0.26494 | 0.185574 | -0.35829 | 1 | | |
| <i>POP</i> | -0.17628 | -0.36791 | -0.15662 | 0.268967 | -0.657135648 | 1 | |
| <i>C&W waste</i> | 0.493968 | 0.463693 | 0.443704 | -0.63769 | 0.58320019 | -0.63453 | 1 |

method before beginning the modeling process by choosing model parameters. As previously stated, these parameters vary depending on the model hypothesis.

An ANN is made up of the connections between groups of neurons. Neurons in the hidden layer process the information acquired by the input layer using non-linear transfer functions. The network complexity is influenced by the number of hidden neurons. It is worth noting that the network performance cannot be anticipated with certainty when the number of learning nodes is increased or decreased. Each node learns fewer samples when there are more learning nodes. The efficiency of training sample anticipation improves in this situation. However, the method's predictive performance on test data significantly decreases when the test data has a wide range of variation. Reducing the number of nodes, on the other hand, results in each node being optimized by lots of samples, which might lead to over-fitting.

Consequently, trial and error is the most effective way to establish the nodes' proper numbers in the hidden layer. It should be noted that 10% of the training dataset was used for

validating the neural networks during the training phase. As illustrated in Figs. 4 and 5, the ideal neurons' number in the hidden layer is 7 for MLPNN and 8 for RBFNN. In general, the models' performance improved as the number of neurons grew to 7 for MLPNN and 8 for RBFNN, but then dropped owing to an over-fitting problem.

The data was imported into the MATLAB machine-learning regression learner, which contains built-in SVM algorithms for SVM forecasts. Six different SVM algorithms were tested (linear, quadratic, cubic, fine Gaussian, medium Gaussian, and coarse Gaussian). The results of each SVM algorithm are shown in Table 3. The quadratic SVM model and medium Gaussian SVM model provided the highest forecasting efficiency based on the findings. Due to the slightly better results of RMSE, the medium Gaussian model was considered the best SVM model.

In ANFIS, the data classification was done using a one-pass subtractive clustering approach, with the radius chosen by trial and error. The cluster radius is a measurement of a cluster's effect range, and as a result, the number of rules

Fig. 4 Determining ideal neuron number in MLPNN method

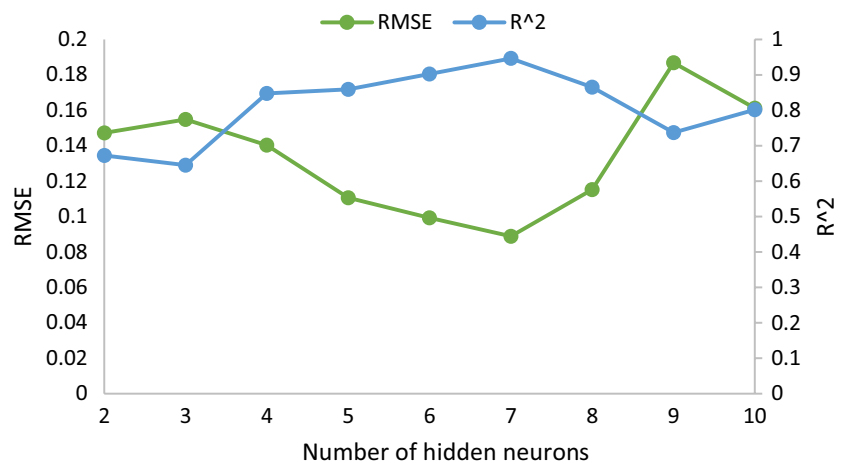


Fig. 5 Determining ideal neuron number in RBFNN method

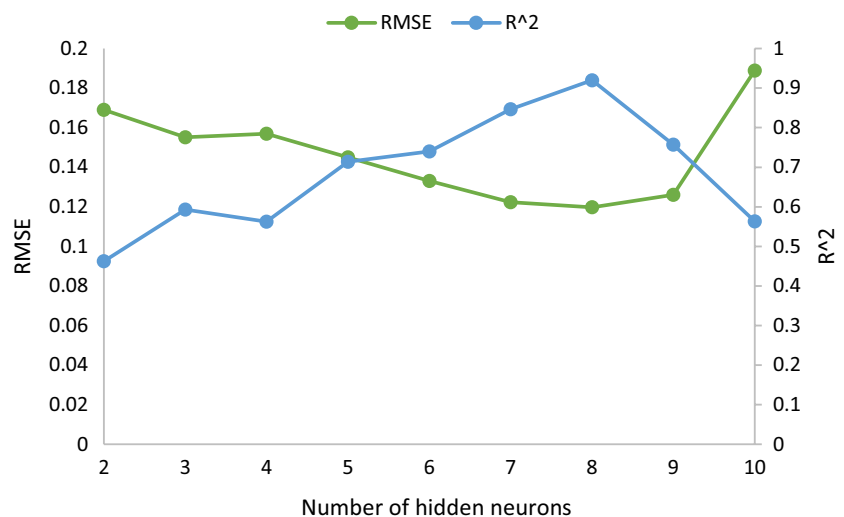


Table 3 Evaluation criteria for SVM models

| SVM type | RMSE | R ² |
|---------------------|---------|----------------|
| Linear SVM | 0.10455 | 0.80 |
| Quadratic SVM | 0.10191 | 0.81 |
| Cubic SVM | 0.16011 | 0.53 |
| Fine Gaussian SVM | 0.17748 | 0.42 |
| Medium Gaussian SVM | 0.10126 | 0.81 |
| Coarse Gaussian SVM | 0.12448 | 0.72 |

will change. Generally, the discrepancy between the forecasted and actual values diminishes as the number of rules increases, allowing for more complicated relationships to be simulated with a larger number of rules. In this research, the ANFIS model with various cluster radiuses was built to calculate the number of clusters based on the model’s findings. The optimum fuzzy structure for forecasting C&DW generation in a neural-fuzzy network was at radius 0.25, with R² and RMSE values of 0.96491 and 0.042099 correspondingly for test data. The model’s performance with various r values is shown in Fig. 6. As the cluster radius was extended from 0.1 to 0.9, an uneven trend in the ANFIS model’s performance was detected.

The hybrid training technique was utilized, and the number of epochs was kept constant at 100 when determining the best cluster radius. Variations in the accept and reject ratio values were demonstrated not to affect the model’s efficiency. Furthermore, the model’s performance was examined for up to 1000 epochs using the optimal cluster radius, and the improvement in results for training and testing data was negligible. Figure 7 provides scatter plots of observed versus anticipated C&DW quantities for the test data using the best model developed. The error histogram for test data is shown in Fig. 8. It can be seen that the errors are normally

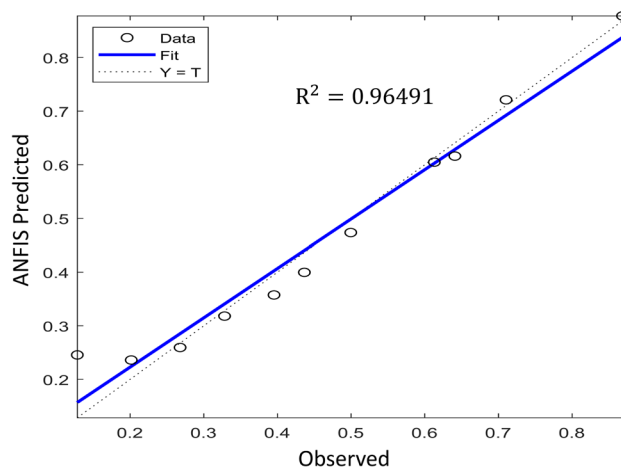
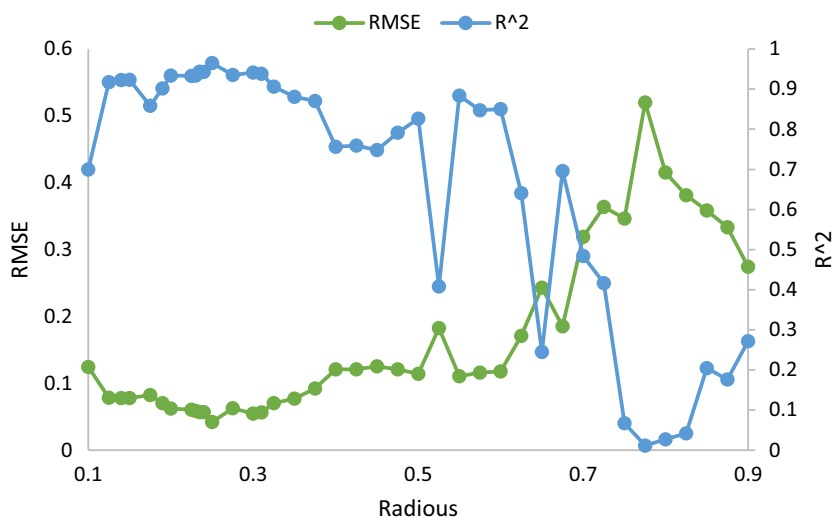


Fig. 7 Observed against optimal ANFIS predicted C&DW quantity

distributed, which shows that the model is well-fit. Additionally, positive and negative values show that the projected values are higher or lower than the actual values, indicating the model’s accuracy.

As shown in Table 4, all models functioned satisfactorily throughout the training and testing stages, with R² values greater than 0.8. However, it was found that the ANFIS model performed tangibly better when comparing the models. The ANFIS model’s RMSE value for the training and testing datasets, respectively, was 0.00484 and 0.04209, which was significantly better than the other methods. This result demonstrates that the ANFIS forecasted C&DW generation rate matches the real data procedure; also this model anticipated the C&DW generation rate more accurately and carefully. It also should be highlighted that using the ANFIS network provides benefits over using the ANN. ANFIS will no longer be a black-box system, and it will have additional

Fig. 6 Performance of ANFIS model with different r values



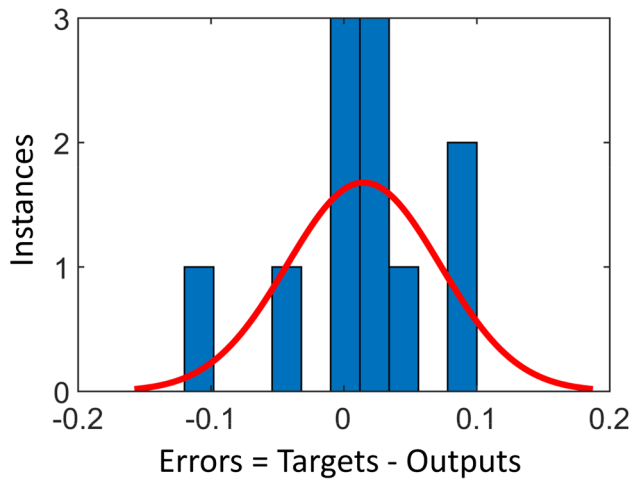


Fig. 8 Error histogram of the ANFIS model

Table 4 Training and testing results of the models

| Models | Phase | R^2 | RMSE |
|--------|----------|-------|---------|
| MLPNN | Training | 0.93 | 0.05443 |
| | Testing | 0.94 | 0.08888 |
| RBFNN | Training | 0.87 | 0.08717 |
| | Testing | 0.92 | 0.11974 |
| SVM | Training | 0.95 | 0.5315 |
| | Testing | 0.81 | 0.10126 |
| ANFIS | Training | 0.99 | 0.00484 |
| | Testing | 0.96 | 0.04209 |

benefits when it comes to the interpretation of fuzzy systems, with the end result being stated in the form of linguistic rules.

ANFIS is an adaptive network with nonlinear capabilities and quick learning ability (Abbasi and El Hanandeh 2016). ANFIS' capacity to anticipate MSW production has been established in past research (Younes et al. 2015; Adeleke et al. 2022). This study's outcomes corroborate these findings and demonstrate the efficacy of the ANFIS algorithm

for estimating monthly C&DW generation. Table 5 provides a comparison between the ANFIS models' performance results in the current study and those of other research studies that established models for waste prediction. It is inferred that the ANFIS algorithm is also a proper method for predicting C&DW generation as well as other types of waste. Moreover, our proposed ANFIS model is capable of improving the accuracy of C&DW prediction in terms of RMSE and R^2 compared with the literature.

This study's results have a variety of practical consequences for researchers, legislators, and environmental protection organizations. First and foremost, the data may be utilized to evaluate urban metabolism to establish a circular economy. Second, it might be utilized to inform a variety of evidence-based policy decisions. It might be employed, for instance, to design a region's waste management capacity, such as landfill space. When implementing this activity, planners frequently encounter a data shortage. Governments can also create appropriate measures for recycling incentives and penalties for polluters. Interregional cooperation can also benefit from the data. The globalization of construction resources, for instance, has expanded the boundaries of an urban metabolism system to multiple locations. Policy-makers are looking for expanded producer accountability or cross-jurisdictional waste material sharing in this situation (Lu et al. 2021). The credible prediction of C&DW generation from this study will provide crucial information for policy-making activities.

Conclusion

Iran's rising population in recent decades has necessitated the expansion of the country's present housing and infrastructural stock. Consequently, there has been a build-up of C&DW that has the potential to promote environmentally appropriate waste management methods. Iran's rulers have issued waste management and disposal legislation and policies. They are, nonetheless, insufficient to resolve the current problem fundamentally. Effective C&DW

Table 5 Accuracy comparison of the current study with the literature

| Reference | Location | Best reported ML method | Waste type | RMSE | R^2 |
|-------------------------------|------------------|-------------------------|----------------------|-------------|-------|
| Younes et al. (2015) | Malaysia | ANFIS | Solid waste | 3.988 | 0.98 |
| Abbasi and El Hanandeh (2016) | Logan, Australia | ANFIS | MSW | 0.002 | 0.99 |
| Golbaz et al. (2019) | Karaj, Iran | ANFIS | Hospital solid waste | N.A | 0.66 |
| Adeleke et al. (2022) | South Africa | ANFIS | MSW | 1.237 | 0.939 |
| Cha et al. (2023) | South Korea | KNN | C&DW | N.A | 0.897 |
| Gulghane et al. (2023) | Nagpur, India | KNN, DT | C&DW | 0.17 to 0.9 | N.A |
| Current study | Tehran, Iran | ANFIS | C&DW | 0.0421 | 0.96 |

management requires knowing the quantity of present C&DW at the regional level and precisely forecasting the amount of it in the future.

This article offers the development of four AI models, including MLPNN, RNFNN, SVM, and ANFIS, for estimating the quantity of C&DW in the Tehran megacity. By March 2017 to January 2022, monthly amount of C&DW was collected from the Tehran Waste Management Organization. To simulate a reasonably decent model, suitable input variables must be chosen. Moreover, given the limited data records in most developing countries, selecting the most representative input variables becomes crucial to enhance the modeling process's efficiency in such contexts. Using the Pearson correlation analysis, it was shown that all inputs have the potential to be utilized in modeling. The Tehran Municipality Organization provided the input data, including month, POP, GDP per capita, FC, NCP, and FB. Data pre-processing was done using the min–max normalization technique to enhance the models' stability and accuracy.

Based on the findings, it can be stated that all of the aforementioned models' errors are satisfactory; thus, these models can be employed to predict monthly C&DW generation estimations. However, the ANFIS model has the highest R^2 value and lowest error for both the training and testing stages, with $R^2 = 0.99$ and $RMSE = 0.00484$ in the training stage and $R^2 = 0.96$ and $RMSE = 0.04209$ in the testing phase.

In future research on C&DW quantity prediction in developing countries, this study recommends using hybrid models, such as the ANFIS model with Genetic Algorithm, to see if the models' accuracy improves. Also, in order to develop a better C&DW estimation model, future studies should expand the range of the data and conduct comparison analysis utilizing other machine learning methods.

Author contribution Conceptualization, M.J. (Milad Jafari); methodology, M.J. and E.M. (Ehsan Mousavi); software, M.J.; validation, M.J. and E.M.; formal analysis, M.J. and E.M.; investigation, M.J.; data curation, M.J. and E.M.; writing—original draft preparation, M.J.; writing—review and editing, M.J. and E.M.; visualization, M.J.; supervision, E.M.; project administration, M.J.; funding acquisition, n/a. All authors read and approved the final manuscript.

Data availability Data will be shared on reasonable request to the corresponding author.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

- Abbasi M, El Hanandeh A (2016) Forecasting municipal solid waste generation using artificial intelligence modelling approaches. *Waste Manage* 56:13–22. <https://doi.org/10.1016/j.wasman.2016.05.018>
- Abbasi M, Rastgoo MN, Nakisa B (2019) Monthly and seasonal modeling of municipal waste generation using radial basis function neural network. *Environ Prog Sustain Energy* 38:. <https://doi.org/10.1002/ep.13033>
- Abdallah M, Abu Talib M, Feroz S et al (2020) Artificial intelligence applications in solid waste management: a systematic research review. *Waste Manage* 109:231–246
- Adeleke O, Akinlabi SA, Jen TC, Dunmade I (2022) Prediction of municipal solid waste generation: an investigation of the effect of clustering techniques and parameters on ANFIS model performance. *Environmental Technology (United Kingdom)* 43:1634–1647. <https://doi.org/10.1080/09593330.2020.1845819>
- Akanbi LA, Oyedele LO, Akinade OO et al (2018) Salvaging building materials in a circular economy: a BIM-based whole-life performance estimator. *Resour Conserv Recycl* 129:175–186. <https://doi.org/10.1016/j.resconrec.2017.10.026>
- Akanbi LA, Oyedele AO, Oyedele LO, Salami RO (2020) Deep learning model for demolition waste prediction in a circular economy. *J Clean Prod* 274:. <https://doi.org/10.1016/j.jclepro.2020.122843>
- Asgari A, Ghorbanian T, Yousefi N, et al (2017) Quality and quantity of construction and demolition waste in Tehran. *J Environ Health Sci Eng* 15:. <https://doi.org/10.1186/s40201-017-0276-0>
- Ayeleru OO, Fajimi LI, Oboirien BO, Olubambi PA (2021) Forecasting municipal solid waste quantity using artificial neural network and supported vector machine techniques: a case study of Johannesburg, South Africa. *J Clean Prod* 289:. <https://doi.org/10.1016/j.jclepro.2020.125671>
- Białko M (2018) Open Access (CC BY-NC 4.0). 21:419–436. <https://doi.org/10.17512/ios.2018.4.8>
- Boroujeni SPH, Razi A, Khoshdel S et al (2024) A comprehensive survey of research towards AI enabled unmanned aerial systems in pre-, active-, and post-wildfire management. *Information Fusion* 108:102369. <https://doi.org/10.1016/j.inffus.2024.102369>
- Broomhead, Lowe (1988) Radial basis functions, multi-variable functional interpolation and adaptive networks
- Cai T, Wang G, Guo Z (2020) Construction and demolition waste generation forecasting using a hybrid intelligent method. In: 2020 9th International Conference on Industrial Technology and Management (ICITM). IEEE, pp 312–316
- Cha GW, Moon HJ, Kim YM et al (2020) Development of a prediction model for demolition waste generation using a random forest algorithm based on small datasets. *Int J Environ Res Public Health* 17:1–15. <https://doi.org/10.3390/ijerph17196997>
- Cha GW, Choi SH, Hong WH, Park CW (2023) Developing a prediction model of demolition-waste generation-rate via principal component analysis. *Int J Environ Res Public Health* 20:. <https://doi.org/10.3390/ijerph20043159>
- Chhay L, Reyad MAH, Suy R et al (2018) Municipal solid waste generation in China: influencing factor analysis and multi-model forecasting. *J Mater Cycles Waste Manag* 20:1761–1770. <https://doi.org/10.1007/s10163-018-0743-4>
- Cochran KM, Townsend TG (2010) Estimating construction and demolition debris generation using a materials flow analysis approach. *Waste Manage* 30:2247–2254. <https://doi.org/10.1016/j.wasman.2010.04.008>
- Fakhri M, Ahmadi T, Shahryari E, Jafari M (2023) Evaluation of fracture behavior of stone mastic asphalt (SMA) containing

- recycled materials under different loading modes at low temperatures. *Constr Build Mater* 386. <https://doi.org/10.1016/j.conbuildmat.2023.131566>
- Gao Y, Gong Z, Yang N (2018) Estimation methods of construction and demolition waste generation: a review. In: IOP Conference Series: Earth and Environmental Science. Institute of Physics Publishing
- Ghanbari F, Kamalan H, Sarraf A. ND An evolutionary machine learning approach for municipal solid waste generation estimation utilizing socioeconomic components. <https://doi.org/10.1007/s12517-020-06348-w/Published>
- Golbaz S, Nabizadeh R, Sajadi HS (2019) Comparative study of predicting hospital solid waste generation using multiple linear regression and artificial intelligence. *J Environ Health Sci Eng* 17:41–51. <https://doi.org/10.1007/s40201-018-00324-z>
- Gulghane A, Sharma RL, Borkar P (2023) Quantification analysis and prediction model for residential building construction waste using machine learning technique. *Asian Journal of Civil Engineering* 24:1459–1473. <https://doi.org/10.1007/s42107-023-00580-x>
- Guo H nan, Wu S biao, Tian Y jie, et al (2021) Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: A review. *Bioresour Technol* 319
- Hu R, Chen K, Chen W et al (2021) Estimation of construction waste generation based on an improved on-site measurement and SVM-based prediction model: a case of commercial buildings in China. *Waste Manage* 126:791–799. <https://doi.org/10.1016/j.wasman.2021.04.012>
- Huang L, Cai T, Zhu Y et al (2020) Lstm-based forecasting for urban construction waste generation. *Sustainability (switzerland)* 12:1–12. <https://doi.org/10.3390/su12208555>
- Jafari M, Khoshand A, Sadeghi N, Mirzanagh PA (2024) A comparative LCA of external wall assemblies in context of Iranian market: considering embodied and operational energy through BIM application. *Environ Sci Pollut Res* 31:7364–7379. <https://doi.org/10.1007/s11356-023-31451-2>
- Jain MS (2021) A mini review on generation, handling, and initiatives to tackle construction and demolition waste in India. *Environ Technol Innov* 22
- Kannangara M, Dua R, Ahmadi L, Bensebaa F (2018) Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. *Waste Manage* 74:3–15. <https://doi.org/10.1016/j.wasman.2017.11.057>
- Khoshand A (2021) Application of artificial intelligence in groundwater ecosystem protection: a case study of Semnan/Sorkheh plain. *Iran Environ Dev Sustain* 23:16617–16631. <https://doi.org/10.1007/s10668-021-01361-9>
- Khoshand A, Khanlari K, Abbasianjahromi H, Zoghi M (2020) Construction and demolition waste management: fuzzy analytic hierarchy process approach. *Waste Manage Res* 38:773–782. <https://doi.org/10.1177/0734242X20910468>
- Khoshand A, Karami A, Rostami G, Emaminejad N (2023) Prediction of e-waste generation: application of modified adaptive neuro-fuzzy inference system (MANFIS). *Waste Management & Research: the Journal for a Sustainable Circular Economy* 41:389–400. <https://doi.org/10.1177/0734242X221122598>
- Lu W, Lou J, Webster C et al (2021) Estimating construction waste generation in the Greater Bay Area, China using machine learning. *Waste Manage* 134:78–88. <https://doi.org/10.1016/j.wasman.2021.08.012>
- Martínez Lage I, Martínez Abella F, Herrero CV, Ordóñez JLP (2010) Estimation of the annual production and composition of C&D Debris in Galicia (Spain). *Waste Manage* 30:636–645. <https://doi.org/10.1016/j.wasman.2009.11.016>
- Menegaki M, Damigos D (2018) A review on current situation and challenges of construction and demolition waste management. *Curr Opin Green Sustain Chem* 13:8–15
- 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>
- Parisi Kern A, Ferreira Dias M, Piva Kulakowski M, Paulo Gomes L (2015) Waste generated in high-rise buildings construction: A quantification model based on statistical multiple regression. *Waste Manage* 39:35–44. <https://doi.org/10.1016/j.wasman.2015.01.043>
- Šajin R, Stafilov T, Balabanova B, Alijagić JA (2022) minerals Multi-scale application of advanced ANN-MLP model for increasing the large-scale improvement of digital data visualisation due to anomalous lithogenic and anthropogenic elements distribution. <https://doi.org/10.3390/min>
- Song Y, Wang Y, Liu F, Zhang Y (2017) Development of a hybrid model to predict construction and demolition waste: China as a case study. *Waste Manage* 59:350–361. <https://doi.org/10.1016/j.wasman.2016.10.009>
- Sunayana KS, Kumar R (2021) Forecasting of municipal solid waste generation using non-linear autoregressive (NAR) neural models. *Waste Manage* 121:206–214. <https://doi.org/10.1016/j.wasman.2020.12.011>
- Umar UA, Shafiq N, Malakahmad A et al (2017) A review on adoption of novel techniques in construction waste management and policy. *J Mater Cycles Waste Manag* 19:1361–1373. <https://doi.org/10.1007/s10163-016-0534-8>
- Xiaonan W, Wei W, Ting C, et al (2020) Using RBF neural network in forecasting urban construction and demolition waste generation. In: *Proceedings - 2020 International Conference on Big Data and Social Sciences, ICBDS 2020*. Institute of Electrical and Electronics Engineers Inc., pp 198–201
- Younes MK, Nopiah ZM, Basri NEA et al (2015) Solid waste forecasting using modified ANFIS modeling. *J Air Waste Manage Assoc* 65:1229–1238. <https://doi.org/10.1080/10962247.2015.1075919>
- Zhao W, Ren H, Rotter VS (2011) A system dynamics model for evaluating the alternative of type in construction and demolition waste recycling center – the case of Chongqing, China. *Resour Conserv Recycl* 55:933–944. <https://doi.org/10.1016/j.resconrec.2011.04.011>

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.