RESEARCH ARTICLE

Comparative study of predicting hospital solid waste generation using multiple linear regression and artificial intelligence

Somayeh Golbaz[1] · **Ramin Nabizadeh**[1] · **Haniye Sadat Sajadi**[2]

Abstract

Purpose A successful hospital solid waste (HSW) management needs an accurate estimation of waste generation rates. The conventional regression methods upon increasing the number of input variables hardly can predict the HSW generation rate and require more complex modeling. In return, application of machine learning methods seems to be able to increase the power of predicting the produced wastes.

Methods To predict the HSW, Multiple Linear Regression(MLR) and several Neuron- and Kernel-based machine learning methods were employed to analyze data from hospitals of Karaj metropolis. The number of wards, active and occupied beds, staffs and inpatients, and ownership type and activity years of hospital were defined as the model inputs. In addition, proposed models performance was evaluated based on coefficient of determination (R^2) and Mean-Square Error (MSE).

Results The performance of Neuron- and Kernel-based machine learning methods indicated that both models were satisfactory in predicting HSW. However, the better results of 0.82–0.86 for average R^2 value and 0.003–0.008 for average MSE value, indicated relative superiority of Kernel-based models compared to Neuron based (average $R^2 = 0.68$–0.74, average MSE = 0.009–0.023) and MLR models. Number of staffs and hospital ownership type were the most influential model variables in predicting the HSW generation rate.

Conclusions The machine learning methods could interpret the relationship between waste generation rate and model inputs, appropriately. Thus, they may play an effective role in developing cost-effective methods for suitable HSW management.

Keywords Multiple linear regression · Machine learning method · Hospital · Solid waste

Introduction

Hospital solid waste (HSW) is a critical public health issue in all societies due to the presence of different pathogens, hazardous and chemical anticancer agents and radioactive wastes therein which include all kinds of perilous wastes. Moreover, cutting and sharp materials are available in those places which

are extremely dangerous for the people who are in contact with them. Poor management system for such materials causes environmental pollution and endangers the human health [1, 2]. Thus, a well-established management system is required. Establishing such management system is very difficult because of the complexity and heterogeneous nature of hospital waste productions. One of the crucial factors in the start-up of such a complex system is the extent of accurate estimation of waste generation rates that may be either short-term or long-term. The short-term estimation of HSW generation rates is necessary for better design and management of storage, collection, and transfer systems [3, 4] and long-term estimation is required for selecting appropriate waste treatment technologies or selecting landfill sites or understanding the impacts of new policies and initiatives [4]. HSW generation rates can be measured by direct sampling; but in many cases, the hospitals do not have enough resources to create a complete database of HSW quantity [4, 5]. Several methods including data mining, sample surveys, and models based on knowledge of

✉ Ramin Nabizadeh
rnabizadeh@tums.ac.ir

1 Department of Environmental Health Engineering, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

2 Health Services Management, National Institute for Health Research, Tehran University of Medical Sciences, Tehran, Iran

effective factors have been used for prediction of waste generation rate. These models include statistical models or conventional method that mostly focus on deterministic methods or trend analysis regardless of the dynamic properties of municipal solid waste (MSW) generation [4].

Therefore, data modeling of complex systems requires advanced methods, which have acceptable performance in the prediction of the behavior of the dynamic systems, to establish a nonlinear relationship between inputs and outputs. In recent years, machine learning methods such as Artificial Neural Networks (ANN), Fuzzy Logic - Artificial Neural Networks (ANFIS), and Support Vector Regression (SVM) have emerged and are becoming popular because of their high flexibility and proven prediction abilities. The ANN model was shown to be able to predict industrial solid waste generation by Tiwari et al. [6]. Wieland et al., (2002) pointed to algorithms for deriving qualitative rules from ANN models and reported this could be developed [7]. ANFIS is one of these developed algorithms. The review of studies shows that only a small number of 106 published articles were associated with advanced and non-conventional methods [4] and unfortunately the use of artificial intelligence models as a tool for the planning, operation and optimization of healthcare waste management system is not widespread as in other fields of environmental engineering. In addition, most of the articles focused on the municipal solid waste issue [8–10] and there is limited evidence about the prediction of hospital solid waste generation as well as an optimal model for this purpose.

Accordingly, this study was aimed to determine the variables that affect the HSW generation by using different methods such as feature selection. Then various data mining methods was examined in order to achieve a more accurate prediction. Therefore, Multiple Linear Regression (MLR) along with several machine learning methods including Artificial Neural Networks (ANN), Fuzzy Logic - Artificial Neural Networks (ANFIS), Support Vector Regression (SVM), Least Squares Support Vector Regression (LSSVM), Fuzzy Logic - Support Vector Regression (FSVM) have been employed in this study to introduce an appropriate model in prediction of HSW generation rate.

Materials and methods

Dataset

The data of eight single-specialty hospitals in Karaj metropolis (35°48′45″N, 51°0′30″E, Iran) in 2016 was used in this study. The hospitals included four university hospitals (H_1 to H_4), three private hospitals (H_5 to H_7), and one social security hospital (H_8).

Model variables

In this study, hospital waste was divided into three groups of infectious (IHSW), general (GHSW), and total (THSW) waste. Their values were obtained by sampling and weighing of waste for four months according to the procedures described by Farzadkia et al. [11]. These dependent variables were considered as model outputs.

According to an overview of effective parameters in HSW generation rate [11–16], interviews with academics and hospital administrators, medical waste management checklist of Ministry of Health and Medical Education of Iran, and feature selection method (Relief-F for regression (RRelief-F)), seven independent variables were selected as input features in HSW generation rate prediction, including: number of active beds (NAB): total hospital beds which are regularly maintained and staffed and immediately available for the care of admitted patients; number of the hospital's wards (NHW): total wards within the hospital for the care of numerous patients having the same condition, e.g., a maternity ward; number of hospital's staff (NHS); hospital ownership type (HOT) that was encoded governmental = 1, private = 2, and social hospital = 3; number of occupied beds (NOB): total beds that are licensed, physically available, staffed, and occupied by a patient; number of inpatients (NIP): total patients who come to the hospital for diagnosis or treatment that requires an overnight stay; number of hospital's activity years (NAY). Table S1 shows the ranks and weights of mentioned predictors for each of response vector (IHSW, GHSW, and THSW) using Relief-F for regression (RRelief-F). The weights (a range from −1 to 1) and ranks were the indexes of the most important predictors. It should be noted that the multicollinearity test showed that there is no similarity between the independent variables in the model.

In this study, inputs data was a 105×7 matrix, representing static data: 105 samples of 7 elements. Also, target was a 105×1 matrix, representing static data: 105 samples of 1 element.

Table 1 presents the average values of model input and output variables in HSW generation rate prediction.

Note that three targets i.e., IHSW, GHSW, and THSW were modeled separately.

Models

MLR model

Since HSW generation forecasting depends on several factors, Multiple Linear Regression (MLR) method is commonly used. In fact, a MLR model states

Table 1 Mean and standard deviation of the model variables

Hospital	Input variables							Output variables		
	NAB	NHW	NHS	HOT *	NOB	NIP	NAY	IHSW (Kg/day)	GHSW (Kg/day)	THSW (Kg/day)
H_1	117	9	406	1	103	68 ± 1.6	33	193.2 ± 45.2	128.4 ± 22.1	321.6 ± 66.3
H_2	231	10	474	1	199	195 ± 30	36	195.4 ± 54.8	116.9 ± 52.8	312.3 ± 100
H_3	133	6	492	1	120	96 ± 19	64	217.5 ± 57.6	29.1 ± 4	246.6 ± 58.1
H_4	188	12	508	1	145	39 ± 2	52	200 ± 39.6	121 ± 10.4	321 ± 29.6
H_5	115	15	543	2	61	161 ± 18	24	187.4 ± 5.6	109.7 ± 13.7	297.1 ± 11.8
H_6	181	15	500	2	110	46 ± 10	15	155.8 ± 24.6	141.4 ± 20.3	297.2 ± 32.5
H_7	70	7	320	2	47	47 ± 10	44	71.03 ± 26.2	114.5 ± 34.3	185.5 ± 34.4
H_8	353	17	825	3	259	101 ± 9	14	395.9 ± 87.2	185.8 ± 68.1	581.7 ± 123.42
Mean	174	11	509	–	131	94	35	202.0	118.4	320.4
Total	1388	91	4068	–	1044	753	282	1616.2	946.8	2563

*Hospital ownership type (HOT) that was encoded governmental = 1, private = 2, and social hospital = 3

relationship between the independent variables and the dependent variable according to the following equation:

$$HSW = \beta_0 + \beta_1 NAB + \beta_2 NHW + \beta_3 NHS + \beta_4 HOT$$
$$+ \beta_5 NOB + \beta_6 NIP + \beta_7 NAY + e, \tag{1}$$

where the dependent variable HSW represents the response variables (IHSW, GHSW, and THSW); and NAB, NHW, NHS, HOT, NOB, NIP and NAY are input variables with the coefficients β_0 to β_7 to be estimated from the data.

Multiple linear regression model was calculated using Entry method (using SPSS software version 16). The standard method is simultaneous; all independent variables were entered into the equation at the same time. It is an appropriate analysis when dealing with a small set of predictors and when the researcher does not know which independent variables will create the best prediction equation. In addition, the MLR model was used with the forward and backward stepwise method as a selection method. Also, the Normal probability plot was used for testing normality of the dependent variables. Since the variable HOT is categorical we used dummy variables HOT_1 (Governmental), and HOT_2 (Private) represents the binary independent variables (Table 2).

Machine learning methods

Different types of machine learning methods exist, but they are typically classified in two major groups: a) Neuron based

Table 2 Dummy coding for a type of hospital variable

Type	Governmental	Private	Social
Governmental	1	0	0
Private	0	1	0
Social	0	0	1

methods i.e., ANN and ANFIS, and b) Kernel-based methods i.e., SVM, LSSVM, and FSVM [17].

In order to compare the performance of the models, feature scaling was used to standardize the range of input and output variable between 0 and 1 [18].

ANN model One of the machine learning methods that have acceptable performance in the prediction of nonlinear and time series regression problems, is ANN method. This method is formed based on the nodes derived from a simplified model of nervous system Neurons [19]. This method usually has three layers including input, learner (hidden) and output layers. The nodes in the learning layer learn the relationships between inputs and outputs as some of the optimized sigmoid functions [20]. These sigmoid functions are introduced with bias (b) and width (w) parameters. During the training process, the parameters of sigmoid functions change to the extent that results in the lowest prediction error. After optimizing the nodes functions, the output variables are obtained based on a linear composition of optimizing sigmoid [21]. The ANN architecture used in this study is shown in Fig. 1.

In this study, Levenberg–Marquardt back-propagation algorithm has been used for optimization of nodes' learning functions. Since seven input parameters have been used for prediction of HSW in this study, the network architecture is $7 \times n \times 1$. The number of Neurons in the hidden layer changed to determine the most appropriate number of nodes in the hidden layer for prediction of output parameters of the system.

ANFIS model One of the machine learning methods is ANFIS, therein nodes learning is based on the fuzzy rules. In this method, before learning the training samples by learning layer nodes, the input data is fuzzified using fuzzy membership functions [22]. The membership functions are designed based on the linguistic variables that can map the values of features from mathematical space to the human logic [23].

Fig. 1 The structure of ANN model
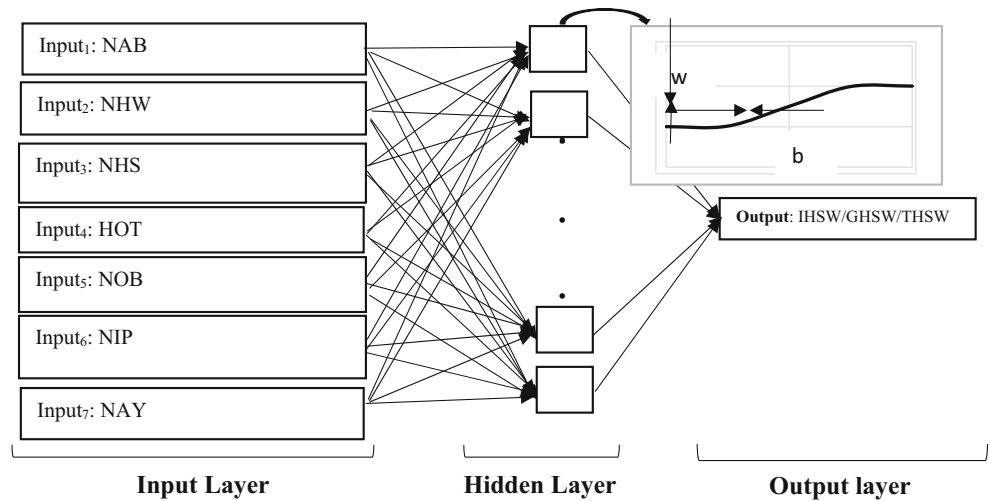
The fuzzy rules are propounded as {if-then} rules, defining the relationships between input and output membership functions. For instance, a fuzzy rule may be explained as {if the number of active beds is increased, then the waste produced in the wards is increased}. Using the membership functions of input and output features, the target is calculated as fuzzy values. The last step in this method is the conversion of the fuzzy value of the target to the mathematical value which is called Defuzzification. In this research, two membership functions {low, high} have been considered for each of problem variables, and HSW generation rate was predicted using this network for different levels of input parameters (Fig. 2).

SVM model In recent years, SVM method has been used widely for prediction of the nonlinear behavior of dynamic and complex systems. This method has been used in the classification and regression problems as well [24]. Learning these machines is based on finding a hyperplane in the features space for data modeling. The specimens that are located within the epsilon distance to this plane, are assumed to have similar behaviors, and the behavior of other specimens are determined based on the distance from this plane (ξ)

(Fig. 3). The position of this plane is specified based on points which are called support vector.

This plane may be explained by different equations (Kernels) such as linear ($f = \gamma x x_0$), polynomial ($f = (\gamma x x_0)^n$), radial basis ($f = e^{(-\gamma(x-x_0)^2)}$), and sigmoid ($f = \tanh(\gamma x x_0)$) [25]. In these relations, γ denotes the Kernel parameter.

LSSVM model In SVM method, hyperplane position is optimized based on the Kernel margin, but in LSSVM method, hyperplane position is optimized based on minimizing the total square of the prediction error of training data [26]. In this method, radial basis function and sigmoid Kernels are usually used. One of the advantages of this method is a better prediction of the behavior of data closer to the hyperplane in comparison to the SVM method.

FSVM model One of the ideas in designing machine learning is the use of fuzzy logic in learning the training data using support vector machine. In this method, the behavior of samples is not calculated linearly based on their distance to the hyperplane [27]. Figure 4 shows a triangular membership function that its center lies on the data having similar behavior near the
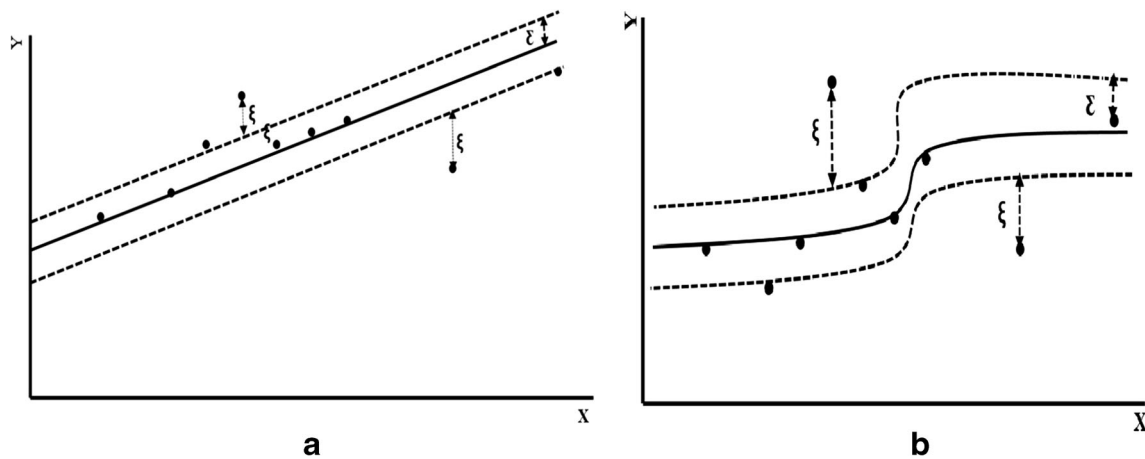

Fig. 2 ANFIS model structure
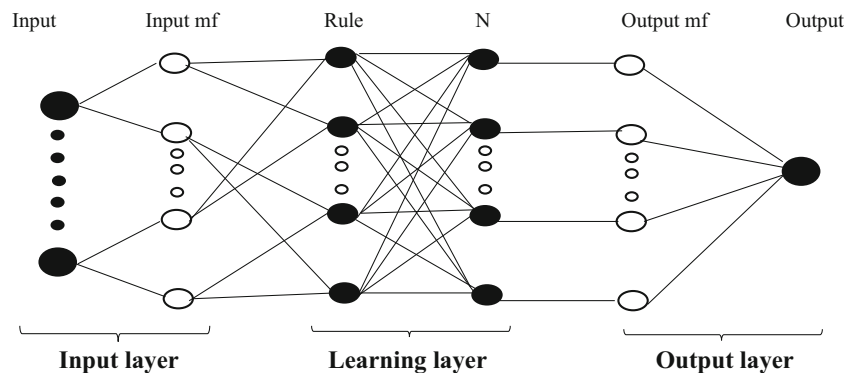
Fig. 3 Conventional kernels: **a** linear kernel, **b** sigmoid kernel

hyperplane. Whatever the sample's distance from hyperplane is increased, its membership degree to the values more different than the values lying on the plane is increased and vice versa. In Fig. 4, the sample M that has a distance from the plane equal to b, its membership degree to the K is lower than the samples near the center.

Training and test procedure of machine learning models

In this study, a code written in MATLAB programming environment was used for implementation of machine learning methods. In the experiments, five-fold cross-validation was employed which four folds were used for training and the last fold was used for the test. Since a validation process is required in ANN method, 70, 15, and 15% of data was considered for training, validation, and test, respectively. LibSVM 3.1 and LSSVM v 1.8 with default Kernels and parameter values were used for Kernel-based methods.

Performance criteria

The performance of the methods was evaluated by comparing their predicted outputs with observed data. In this study, Mean-
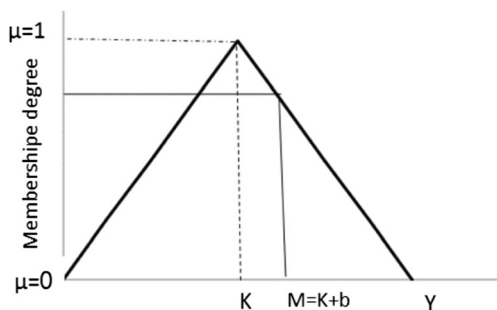

Fig. 4 Membership value of FSVM model

Square Error (MSE), and coefficient of determination (R^2) were considered for performance evaluation (eqs. 2 and 3).

$$MSE(t) = \left(\frac{1}{n}\right) \sum_{i=1}^{n} \left(HSW_p(t) - HSW_A(t)\right)^2 \qquad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(HSW_p(t) - HSW_A(t)\right)^2}{\sum_{i=1}^{n} \left(HSW_A(t) - \overline{HSW}_A(t)\right)^2} \qquad (3)$$

where $HSW_p(t)$, $HSW_A(t)$, $\overline{HSW}_A(t)$ and n are predicted, actual, and average value of HSW and the number of samples, respectively.

Results and discussion

MLR model

Results of normal probability plot are shown in **Fig. S1.** The plots display that the infectious and total waste values had a p value less than 0.05 which showed the rejection of normal distribution for data, whilst the general waste data was normally distributed as p value was greater than 0.05. To deal with the problem of non-normality distributed data of IHSW and THSW, transformation method was used. Since the box and cox method was not applicable, the Johnson transformation formula was used as follows:

$$HSW^* = a + b \text{ Arc sin h}((HSW - c)/d), \qquad (4)$$

where b and d > 0 and a, c arereal numbers.

Fig. S2 shows the transformation results for infectious and total waste data.

Table 3 Statistical characteristics of the MLR model (after transformation)

Output variables	Input variables		Unstandardized coefficients		Standardized coefficients	T-value	a-Level
			B	Standard error	Beta		
IHSW [*]	Intercept		−2.516	2.557		−0.984	0.328
	NAB		−0.009	0.044	−0.798	−0.210	0.834
	NHW		−0.033	0.182	−0.125	−0.183	0.855
	NHS		0.008	0.003	1.129	2.544	0.013
	HOT	HOT_1	0.481	0.601	0.237	0.800	0.426
		HOT_2	−0.278	1.328	−0.127	−0.209	0.835
	NOB		0.007	0.053	0.460	0.130	0.897
	NIP		−0.001	0.003	−0.028	−0.157	0.876
	NAY		−0.014	0.014	−0.230	−1.005	0.317
GHSW	Intercept		444.681	190.666		2.332	0.022
	NAB		1.049	3.265	1.599	0.321	0.749
	NHW		13.177	13.569	0.863	0.971	0.334
	NHS		−0.686	0.229	−1.738	−2.995	0.003
	HOT	HOT_1	−166.774	44.800	−1.441	−3.723	0.000
		HOT_2	−225.898	99.010	−1.811	−2.282	0.025
	NOB		−1.145	3.925	−1.351	−0.292	0.771
	NIP		0.098	0.238	0.096	0.410	0.683
	NAY		−0.028	1.076	−0.008	−0.026	0.979
THSW [*]	Intercept		0.283	3.105		0.091	0.927
	NAB		0.009	0.053	0.792	0.173	0.863
	NHW		0.077	0.221	0.285	0.349	0.728
	NHS		0.001	0.004	0.110	0.205	0.838
	HOT	HOT_1	−0.200	0.729	−0.098	−0.275	0.784
		HOT_2	−1.658	1.612	−0.752	−1.029	0.306
	NOB		−0.014	0.064	−0.947	−0.222	0.825
	NIP		0.000	0.004	0.017	0.081	0.936
	NAY		−0.015	0.018	−0.242	−0.880	0.381

Table 3 indicated statistical characteristics of the developed MLR model after transformation on IHSW and GHSW. Accordingly, in term of infectious waste, the number of hospital's staff was significant ($p = 0.013$), while in general waste, the constant term/intercept ($p = 0.022$), the number of hospital's staff ($p = 0.003$), and type of hospitals ($p < 0.05$) were significant. Furthermore, for total waste, none of the independent variables were significant. Furthermore, MSE values of model were 0.274, 1525.7, and 1748.5, respectively, in the prediction of IHSW, GHSW, and THSW. Whilst the R^2 values were obtained 0.75, 0.57 and 0.48, respectively.

Also, the results of statistical characteristics of the developed MLR model (stepwise regression) are shown in Table 4. As shown in the Table, independent variables of number of staff and type 2 of ownership had an impact on infectious waste generation rate; whilst in the prediction of general waste, number of staff and hospital's active years were statistically significant ($p < 0.05$).

Table 4 Statistical characteristics of the developed MLR model after transformation (Stepwise)

Output variables	Input variables	Unstandardized Coefficients		Standardized Coefficients	T-value	a-Level
		B	Standard error	Beta		
IHSW [*]	Intercept	−2.020	0.210		−9.642	0.000
	NHS	0.004	0.000	0.646	11.794	0.000
	HOT_2	−0.814	0.120	−0.372	−6.793	0.000
GHSW	Intercept	206.208	10.030		20.559	0.000
	NAY	−2.491	0.256	−0.693	−9.747	0.000
THSW [*]	Intercept	−1.447	0.397		−3.642	0.000
	NHS	0.004	0.001	0.579	7.395	0.000
	NAY	−0.015	0.005	−0.242	−3.090	0.003

Ultimately, the following equations are obtained for predicting HSW generation rate, i.e., infectious waste (IHSW*), and total waste (THSW*):

$$IHSW^* = -2.02 + 0.004 NHS - 0.814\, HOT_2, \qquad (5)$$

where

$$IHSW^* = -0.365 + 1.29 \sinh^{-1}\left(\frac{IHSW - 168.995}{95.93}\right)$$

$$GHSW = 206.208 - 2.491\, NAY \qquad (6)$$

$$THSW^* = -1.445 + 0.004\, NHS - 0.015\, NAY, \qquad (7)$$

where

$$THSW^* = -1.11 + 1.35 \sinh^{-1}\left(\frac{THSW - 199}{98.8}\right)$$

Finally, the summary of MLR modeling performance is given in Table 5. MSE 0.30, 1748 and 0.50 and R^2 0.73, 0.48 and 0.53 show a weak performance of the model in predicting the generation rate of IHSW, GHSW, and THSW, respectively.

Given our results, it seems that hospital ownership type was significantly effective on waste production. In private hospitals economic status and customer orientation were improved by equipment and multi services, but lack of skills in staffs and workers caused inappropriate separation of infectious waste from general waste. Hence, these factors can be affected on the generation rate of hospital solid waste. Moreover, RRelief-F analysis showed that the variables number of occupied beds (NOB), staff (NHS) and inpatient (NIP), respectively were the most effective factors on the production rate of infectious, general and total waste. Obviously, variations of number of inpatient and occupied bed affect the production rate of infectious waste, but excessive variations of this value, as demonstrated by results of this section, seem to be connected to the other variables such as the number of hospital's staff (NHS). Since, the producers of hospital waste (such as doctors, nurses, and paramedics) are responsible for initial separation thereof, therefore their role in waste production is very effective. If they do not pay enough attention in this relation, it will be make intensive changes in the rate of each one of waste types. It should be mentioned that as long as wastes are mixed together, not only the health consequences will be more acute but also the process of separation will be so difficult and costly.

Table 5 Performance criteria of MLR model in prediction of the HSW generation rate

Output variables Performance criteria	IHSW	GHSW	THSW
R^2	0.73	0.48	0.53
MSE	0.30	1748	0.50

literature review indicated that one of the causes of lack of full separation of waste in hospitals was lack of initial training of new nurses and personnel of hospitals [2, 28, 29]. Whereas personnel of service department were always changing or replacing, it is required to repeat these training continuously. The importance of continuous training was also reported in the study of Sawalem et al., (2009) [30]. Accordingly, it is recommended to engage the trained and experienced persons in the management of HSW, and consider the least job rotation and displacement, as possible.

Machine learning methods

ANN model

The summary of ANN modeling performance in the training, validation and test phases is listed in Tables 6 and 7. Values of 0.012, 0.025, and 0.033 for MSE and 0.73, 0.65, and 0.70 for R^2 (Fig. 5) show a high performance in prediction of the IHSW, GHSW, and THSW generation rate, respectively. It is clear that the ANN shows absolutely lower error measure comparing to the MLR method. Accordingly, MSE 1748 in MLR model was reduced to 0.025 in ANN upon predicting the general waste. Based on the results, ANN indicates superiority in assessing the quality performance. These results can be attributed to the nonlinear behavior of ANN in problem-solving, which provides the opportunity for relating independent variables to dependent ones non-linearly [31]. The neural network method has other advantages such as high-speed generalization subsequently ANN models, and fault tolerance.

The most appropriate number of learning nodes in the hidden layer was obtained 9, 14 and 10 for prediction of IHSW, GHSW, and THSW, respectively (Table 6). Note that upon increasing or decreasing the number of learning nodes, the network performance cannot be predicted certainly. Increasing the number of learning nodes causes each node learns a few number of samples. In this condition, the accuracy of the prediction of training samples is improved, but in a case where the variation range of test data features is high, the accuracy of the method in the prediction of test data is reduced significantly. On the other side, reduction of a number of nodes causes each node to be optimized by lots of samples resulting in over-fitting. In this condition, the parameters of sigmoid functions for prediction of test data will have a long distance from its desired value and the network error will be increased. Therefore, the best method for determination of an appropriate number of nodes in the hidden layer is trial and error.

ANFIS model

The optimum ANFIS architecture found to be $7 \times 14 \times 128 \times 128 \times 1$. The training, validation and test results are shown in Table 7. The test results showed that R^2 values of the model

Table 6 Backpropagation network architecture in prediction of the HSW generation rate according to R^2

Network architecture	IHSW			GHSW			THSW		
	Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
$7 \times 5 \times 1$	0.70	0.45	0.69	0.62	0.42	0.47	0.65	0.63	0.59
$7 \times 6 \times 1$	0.77	0.69	0.62	0.51	0.67	0.41	0.72	0.69	0.63
$7 \times 7 \times 1$	0.75	0.48	0.69	0.61	0.37	0.51	0.72	0.07	0.55
$7 \times 8 \times 1$	0.65	0.59	0.65	0.58	0.53	0.56	0.72	0.39	0.67
$7 \times 9 \times 1$	0.79	0.76	0.64	0.38	0.64	0.38	0.69	0.70	0.69
$7 \times 10 \times 1$	0.71	0.71	0.60	0.60	0.36	0.50	0.74	0.78	0.59
$7 \times 11 \times 1$	0.62	0.69	0.62	0.50	0.66	0.50	0.67	0.56	0.67
$7 \times 12 \times 1$	0.78	0.56	0.64	0.61	0.67	0.51	0.75	0.70	0.57
$7 \times 13 \times 1$	0.71	0.68	0.69	0.54	0.64	0.54	0.57	0.66	0.61
$7 \times 14 \times 1$	0.65	0.55	0.69	0.70	0.66	0.60	0.66	0.67	0.69
$7 \times 15 \times 1$	0.78	0.70	0.61	0.55	0.47	0.55	0.50	0.71	0.49

for the IHSW, GHSW, and THSW were 0.66, 0.71, and 0.59, respectively. Accordingly, it was observed that ANFIS model had higher performance compared to ANN. It should also be noted that the use of ANFIS network has advantages over the ANN, e.g. it does not remain as a black-box system and with respect to the interpretation of Fuzzy systems, it will have more advantages and the final result can be expressed in the form of linguistic rules. Moreover, the MSE values of the ANFIS model for IHSW, GHSW and THSW were 0.005, 0.012, and 0.004, respectively, which is reduced comparing to the neural network. Tiwari et al., (2012) showed that R^2 value was 0.33 for ANN and 0.41 for ANFIS, predicting the industrial waste generation rate; whilst the MSE in both models was very high (2798 and 3096, respectively) [6]. Finally, according to MSE value of the model (approximately 0.000), as observed ANFIS model predicted the HSW generation rate more successfully.

SVM model

Table 7 shows R^2 and MSE performance criteria for SVM model. The test results indicated that R^2 value of the model was 0.79, 0.98 and 0.90 for estimating the IHSW, GHSW, and THSW, respectively, which were significantly improved in comparison with the aforesaid methods. This result shows that the waste generation rate predicted by the SVM follows the actual data procedure and this model predicted the waste generation rate, more successfully and carefully. In the meantime, Table 7 also lists the regression performance of SVM models surpassing the others models with their smaller MSE performance criteria and greater prediction accuracy. In addition, SVM in comparison to the ANN and ANFIS models was a faster method with higher accuracy and reproducibility.
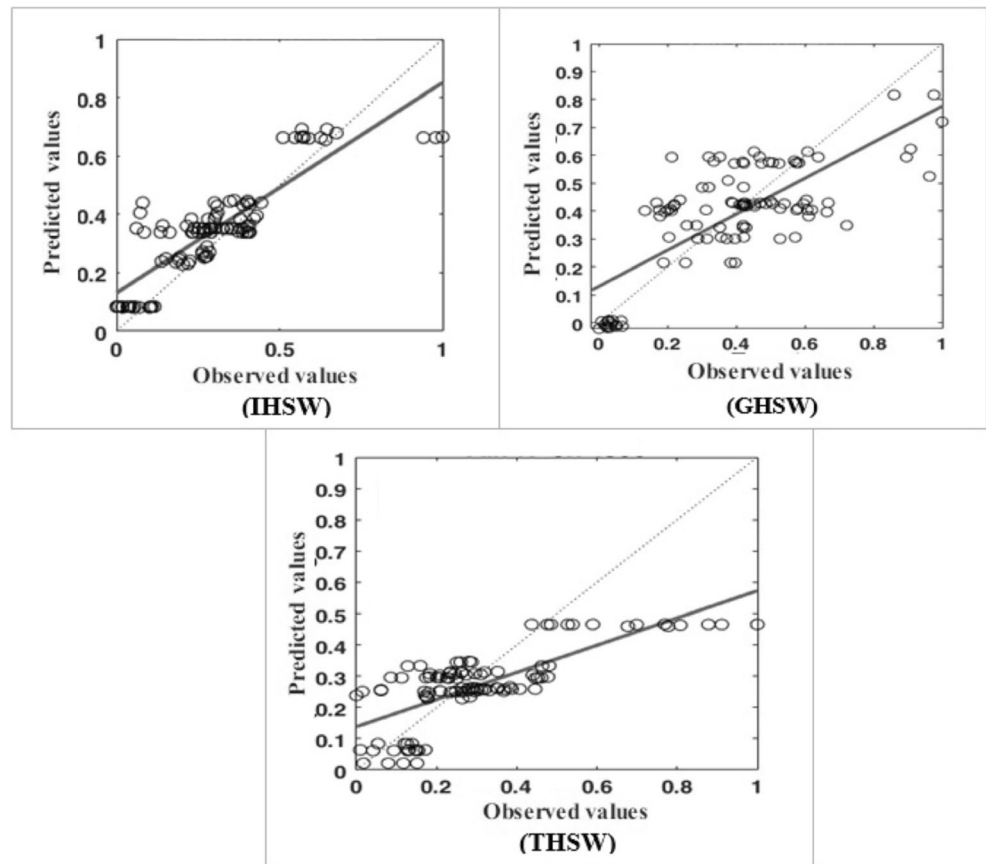
LSSVM model

The test results showed that the MSE of the LSSVM model was obtained 0.004, 0.007 and 0.002 for each of the IHSW, GHSW, and THSW, respectively, which comparing to the ANFIS and neural network models were lower (Table 7). It reveals that this method comparing to the neural network and ANFIS models could appropriately interpret the complex non-linear relationships between input and output variables, and reduce the error rate close to zero. Though, this error in comparison to the MSE of SVM model has not changed substantially. Abbasi et al., (2012) showed that the combination of SVM with the least partial squares for estimation of weekly production of municipal solid waste was successful and due to high-speed computing, also saved time which was consistent with this study [5].

Table 7 The training, validation, and testing results for various models

Models		IHSW		GHSW		THSW	
		R^2	MSE	R^2	MSE	R^2	MSE
ANN	Training	0.79	0.012	0.70	0.023	0.74	0.031
	Validation	0.76	0.011	0.66	0.025	0.78	0.034
	Testing	0.64	0.012	0.60	0.027	0.58	0.035
ANFIS	Training	0.85	0.006	0.74	0.016	0.78	0.009
	Testing	0.66	0.005	0.71	0.012	0.59	0.004
SVM	Training	0.83	0.008	0.77	0.016	0.89	0.009
	Testing	0.79	0.003	0.98	0.002	0.90	0.001
LSSVM	Training	0.92	0.003	0.63	0.021	0.82	0.007
	Testing	0.77	0.004	0.76	0.007	0.70	0.002
FSVM	Training	0.90	0.003	0.88	0.006	0.84	0.004
	Testing	0.92	0.002	0.87	0.001	0.79	0.001

Fig. 5 Plots of predicted values versus observed values for ANN

FSVM model

The ranges of 0.79–0.92 for prediction R^2 showed the strong prediction capability of the FSVM models (Table 7). Also, the better results of 0.001–0.002 for MSE, indicated relative superiority of FSVM compared to aforesaid models (Fig. 6). The prediction accuracy and speed of FSVM model was greater than other models.

Finally, the performance of machine learning methods indicated that both Neuron- and Kernel-based methods were satisfactory in predicting HSW. However, the better results of 0.82–0.86 for average R^2 value and 0.004–0.009 for average MSE value, indicated relative superiority of Kernel-based models compared to Neuron based (average $R^2 = 0.68$–0.74, average MSE = 0.009–0.023).

Conclusion

Overall, our study demonstrated that:

a) The average HSW generation rates, i.e., IHSW, GHSW, and THSW were found 202, 118.4 and 320.4 kg/day.

Fig. 6 Comparison of models' performance by MSE in the test phase
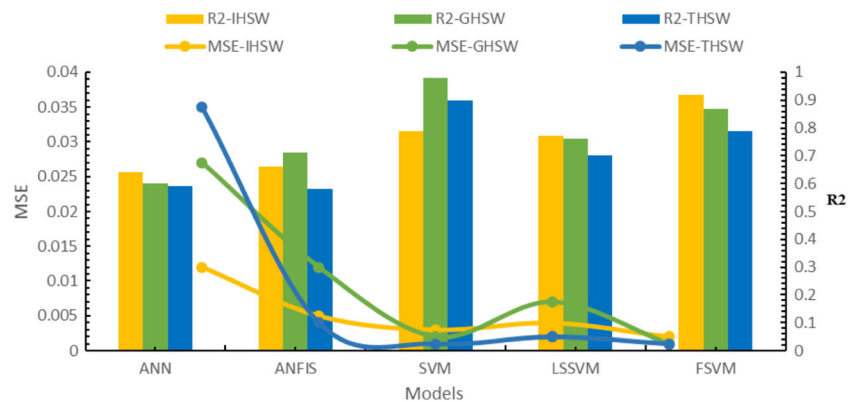
b) The results of MLR analysis showed that the number of hospital's staff and hospital ownership type may be effective on HSW production rate. Also, the result of this study showed that the conventional MLR methods upon increasing the number of input variables hardly can predict the HSW generation rate and require more complex modeling.

c) There was a nonlinear relationship between the seven independent variables and each of dependent variables i.e. IHSW, GHSW, and THSW and application of Kernel-based models such as SVM, FSVM, and subsequently, ANN-based models can provide high predictability with the lowest errors in consistency to the observed data.

d) Also, contrary to some hybrid models such as ANFIS and LSSVM, combining some models abilities such as Fuzzy Logic and SVM, and using such abilities in a hybrid form, FSVM, may result in the development of powerful models that could interpret the relationship between produced HSW rate and seven input variables of the model, appropriately.

Ultimately, accurate forecasts of HSW generation rate are crucial and fundamental for the planning, operation and optimization of hospital waste management system, so, to reach this purpose artificial intelligence techniques, especially kernel bass methods, are more successful than conventional regression models.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

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

References

1. Sabour MR, Mohamedifard A, Kamalan H. A mathematical model to predict the composition and generation of hospital wastes in Iran. Waste Manag. 2007;27(4):584–7.

2. Dastpak H, Golbaz S, Farzadkia M, editors. Hospital waste minimisation, separation, treatment and disposal in Iran: a mini review study. Proceedings of the Institution of Civil Engineers-Waste and Resource Management; 2017: Thomas Telford Ltd.

3. Matsuto T, Tanaka N. Data analysis of daily collection tonnage of residential solid waste in Japan. Waste Manag Res. 1993;11(4):333–43.

4. Sengupta D, Agrahari S. Modelling trends in solid and hazardous waste management: Springer; 2017.

5. Abbasi M, Abduli M, Omidvar B, Baghvand A. Forecasting municipal solid waste generation by hybrid support vector machine and partial least square model. International Journal of Environmental Research. 2012;7(1):27–38.

6. Tiwari MK, Bajpai S, Dewangan U. Prediction of industrial solid waste with ANFIS model and its comparison with ANN model-a case study of Durg-Bhilai twin city India. International Journal of Engineering and Innovative Technology (IJEIT). 2012;6(2):192–201.

7. Wieland D, Wotawa F, Wotawa G. From neural networks to qualitative models in environmental engineering. Computer-Aided Civil and Infrastructure Engineering. 2002;17(2):104–18.

8. Abbasi M, El Hanandeh A. Forecasting municipal solid waste generation using artificial intelligence modelling approaches. Waste management (New York, NY). 2016;56:13.

9. Ali Abdoli M, Falah Nezhad M, Salehi Sede R, Behboudian S. Longterm forecasting of solid waste generation by the artificial neural networks. Environ Prog Sustain Energy. 2012;31(4):628–36.

10. Noori R, Abdoli M, Ghasrodashti AA, Jalili Ghazizade M. Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis: a case study of Mashhad. Environ Prog Sustain Energy. 2009;28(2):249–58.

11. Farzadkia M, Emamjomeh M, Golbaz S, Sajadi H. An investigation on hospital solid waste management in Iran. Glob NEST J. 2015;17:771–83.

12. Da Silva C, Hoppe A, Ravanello M, Mello N. Medical wastes management in the south of Brazil. Waste Manag. 2005;25(6):600–5.

13. Bdour A, Altrabsheh B, Hadadin N, Al-Shareif M. Assessment of medical wastes management practice: a case study of the northern part of Jordan. Waste Manag. 2007;27(6):746–59.

14. Komilis D, Katsafaros N. Statistical predictors of hazardous medical waste generation rates in a 40-bed general hospital. Global NEST Journal. 2011;13(2):170–5.

15. Awad A, Obeidat M, Al-Shareef M. Mathematical-statistical models of generated hazardous hospital solid waste. J Environ Sci Health A. 2004;39(2):315–27.

16. Tesfahun E, Kumie A, Beyene A. Developing models for the prediction of hospital healthcare waste generation rate. Waste Manag Res. 2016;34(1):75–80.

17. Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods: Cambridge university press; 2000.

18. Carbonell JG, Michalski RS, Mitchell TM. An overview of machine learning. In: Machine learning: Springer; 1983. p. 3–23.

19. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform. 2002;35(5):352–9.

20. Wang S-C. Artificial neural network. In: Interdisciplinary computing in java programming: Springer; 2003. p. 81–100.

21. Zhang G, Patuwo BE, Hu MY. Forecasting with artificial neural networks: the state of the art. Int J Forecast. 1998;14(1):35–62.

22. Jang J-S. ANFIS: adaptive-network-based fuzzy inference system. IEEE transactions on systems, man, and cybernetics. 1993;23(3):665–85.

23. Jang J-S, editor. Input selection for ANFIS learning. Fuzzy Systems, 1996., Proceedings of the Fifth IEEE International Conference on; 1996: IEEE.

24. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–97.

25. Schölkopf B, Smola AJ. Learning with kernels: support vector machines, regularization, optimization. and beyond MIT press. 2002.

26. Suykens JA, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett. 1999;9(3):293–300.

27. Hong DH, Hwang C. Support vector fuzzy regression machines. Fuzzy Sets Syst. 2003;138(2):271–81.

28. Akter N, Hussain Z, Trankler J, Parkpian P. Hospital waste management and it's probable health effect: a lesson learned from Bangladesh. Indian journal of environmental health. 2002;44(2): 124–37.

29. Raygan Shirazi A, Mari Oryad H, Malekzadeh J. Solid wastes management of Yasuj hospitals, Iran 2006. Armaghane danesh. 2008;13(1):105–13.

30. Sawalem M, Selic E, Herbell J-D. Hospital waste management in Libya: a case study. Waste Manag. 2009;29(4):1370–5.

31. Jahandideh S, Jahandideh S, Asadabadi EB, Askarian M, Movahedi MM, Hosseini S, et al. The use of artificial neural networks and multiple linear regression to predict rate of medical waste generation. Waste Manag. 2009;29(11):2874–9.