


Extreme learning machine model for water network management

Ahmed M. A. Sattar¹ · Ömer Faruk Ertuğrul² · B. Gharabaghi³  · E. A. McBean³ · J. Cao⁴

Received: 5 May 2016 / Accepted: 27 March 2017
© The Natural Computing Applications Forum 2017

Abstract A novel failure rate prediction model is developed by the extreme learning machine (ELM) to provide key information needed for optimum ongoing maintenance/rehabilitation of a water network, meaning the estimated times for the next failures of individual pipes within the network. The developed ELM model is trained using more than 9500 instances of pipe failure in the Greater Toronto Area, Canada from 1920 to 2005 with pipe attributes as inputs, including pipe length, diameter, material, and previously recorded failures. The models show recent, extensive usage of pipe coating with cement mortar and cathodic protection has significantly increased their lifespan. The predictive model includes the pipe protection method as pipe attributes and can reflect in

its predictions, the effect of different pipe protection methods on the expected time to the next pipe failure. The developed ELM has a superior prediction accuracy relative to other available machine learning algorithms such as feed-forward artificial neural network that is trained by backpropagation, support vector regression, and non-linear regression. The utility of the models provides useful inputs when planning and budgeting for watermain inspection, maintenance, and rehabilitation.

Keywords Water pipe network · Pipe failure · Extreme machine learning · Management tool

1 Introduction

Many pipes within water distribution networks in large cities around the world are in their final stages of design life. These aged infrastructures are prone to frequent major failures and/or leaks that may lead to water losses, interruption in delivery of essential service, and allow contaminated water ingress resulting in hazardous exposure to water consumers. The ongoing cost of repair of the aging water pipe network has reached billions of dollars per year in the North American cities alone [1, 2].

In response, many researchers have developed watermain failure models to predict potential failures and help municipalities forecast the cost of maintenance of water networks. Nishiyama and Filion [3] reviewed several existing watermain failure models and reported that all models have low coefficient of determination. Datasets of pipe failures used in these studies were not large and ranged from 50 to 2000 break instances.

Kleiner et al. [4] was one of the first studies to use machine-learning techniques for prediction of the next pipe break failure. They used the feed-forward backpropagation artificial

✉ B. Gharabaghi
bgharaba@uoguelph.ca

Ahmed M. A. Sattar
ahmoudy77@yahoo.com

Ömer Faruk Ertuğrul
omerfarukertugrul@gmail.com

E. A. McBean
emcbean@uoguelph.ca

J. Cao
jwcao@hdu.edu.cn

¹ Department of Irrigation & Hydraulics, Faculty of Engineering, Cairo University, Giza, Egypt

² Department of Electrical and Electronics Engineering, Batman University, Batman, Turkey

³ School of Engineering, University of Guelph, N1G 2W1, Guelph, Ontario, Canada

⁴ Institute of Information and Control, Hangzhou Dianzi University, Zhejiang 310018, China

neural networks that trained by backpropagation (FFBP-ANN) for deriving complex relations between variables. However, the main disadvantage of traditional ANN methods is that often the solution is caught in a local minimum not reaching the optimum solution. As an alternative, the extreme learning machine (ELM) calculates optimum weights in a single hidden layer feed-forward artificial neural network [5]. Hence, ELM-ANN differs from the traditional FFBP-ANN method, as the optimum weights in the network are calculated analytically, resulting in high performance capacity and fast training for large data sets [6–18]. However, although having many desirable features, the authors have not identified any application of ELM-ANN to water pipe networks.

It has been a challenge for many municipalities to gain knowledge about the frequency and expected timing of future pipe failures. While general guidelines on expected service life gathered from the literature play a major role in developing asset management plans, decision makers need more accurate tools to help provide specific information on the expected cost to maintain/rehabilitate water pipe networks. On this basis, the extreme learning machine (ELM) is described in a novel application to predict time to failure of distribution pipes, including important attributes such as the pipe protective coatings effect, material of pipes, length, and diameter. It is demonstrated how the new evolutionary model can serve as an alternative to ANN and other machine learning models for application on prioritizing rehabilitation of water pipe networks.

2 Materials and methods

2.1 Extreme learning machine

The ELM is a training method for a single hidden layer neural network. It has many advantages over a traditional backpropagation (BP) algorithm. In a BP algorithm, a gradient descent-based learning method, each network weight or bias, is determined by tuning. Due to the nature of tuning, the learning speed in BP is slow and has a tendency to converge to local minima. Determining the network parameters such as the hidden neuron, transfer function, training method, and performance criteria are other disadvantages of BP [12]. The ELM has three layers, one input layer, one output layer, and a hidden layer (Fig. 1). These layers form a single hidden layer forward network where linear algebra is utilized to solve the equations for achieving optimal weights in the output layer. In the ELM, weights of the input layer are randomly assigned. The output weights are calculated analytically using a pre-defined training procedure. Based on the calculation scheme

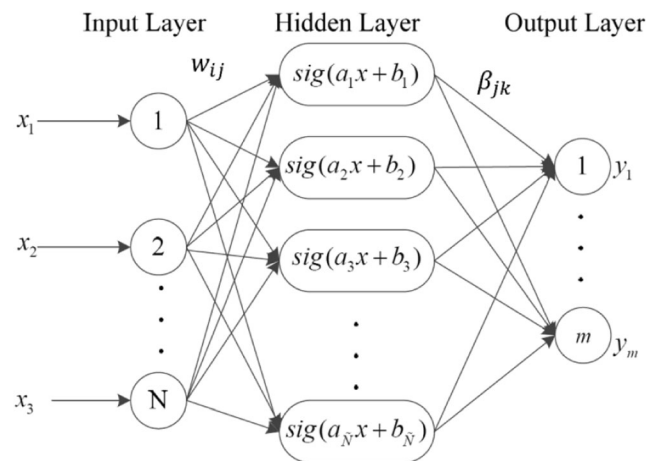


Fig. 1 Basic structure of a single-layer ELM network

of weights and biases in ELM, its training stage is extremely fast and its generalization capacity is high [5, 19].

The output of a single hidden layer feed-forward neural network can be calculated by:

$$y = \sum_{j=1}^m \beta_j g \left(\sum_{i=1}^n w_{i,j} x_i + b_j \right) \quad (1)$$

where y is output of the network, x shows inputs of network, n represents the features and equal to the number of input variables, m is the number of the hidden layer neurons and equal to the output variables of the problem considered, $w_{i,j}$ denotes input weights that connect the i th neuron of the input layer of the neural networks model to the j th neuron of the hidden layer, β_j is a coefficient that connects the j th neuron of the network hidden layer to the related neuron in the output layer, b_j symbolizes biases of the neurons in the hidden layer, and $g(\cdot)$ indicates the activation function. The output of a single hidden layer feed-forward neural network is calculated in two stages. Initially, a single hidden layer network is formed based on user-defined network parameters. These parameters are equal to the number of neurons in the hidden layer linked by the transfer function. The number of neurons that exist in the hidden layer (m) is chosen such that it is less than or equal to the number of data observations. Moreover, the activation function ($g(\cdot)$) can be any piece of infinitely differentiable function [5]. Following the formation of the neurons in the hidden layer, the weights in the output layer are calculated. This is achieved by arbitrary assignments of the weights $w_{i,j}$ in the input layer and the biases b_j .

Thus, Eq. (1) can be written as follows:

$$\mathbf{H}\beta = \mathbf{y} \quad (2)$$

where \mathbf{H} defines the ELM feature mapping matrix [5]:

$$\mathbf{H}(w_{i,j}, b_j, x_i) = \begin{bmatrix} H_{1,1} & \cdots & H_{1,m} \\ \vdots & \ddots & \vdots \\ H_{n,1} & \cdots & H_{n,m} \end{bmatrix} = \begin{bmatrix} g(w_{1,1}x_1 + b_1) & \cdots & g(w_{1,m}x_m + b_m) \\ \vdots & \ddots & \vdots \\ g(w_{n,1}x_n + b_1) & \cdots & g(w_{n,m}x_m + b_m) \end{bmatrix} \quad (3)$$

here, \mathbf{y} and β can be defined as:

$$\mathbf{y} = [y_1 \ y_2 \ \cdots \ y_m] \text{ and } \beta = [\beta_1 \ \beta_2 \ \cdots \ \beta_n] \quad (4)$$

The weights β_j are found by minimizing error in the approximation by the Moore–Penrose generalized inverse method [20] such that:

$$\hat{\beta} = \mathbf{H}^+ \mathbf{y} \quad (5)$$

where \mathbf{H}^+ symbolizes the generalized Moore–Penrose inverse matrix of \mathbf{H} . Huang et al. [5] showed that relying only on determining optimal output weights is sufficient to achieve high accuracy and calculating the output weights instead of tuning, is the fundamental rationale behind the speed and the generalization capacity.

Three more versions of the ELM are employed in this study; backpropagation ELM (tELM), linear regression ELM (ELMr), and self-adaptive ELM (SaELM). In tELM, the output weights are calculated by tuning, while the weights in the input layer and related biases are assigned randomly [12]. The weights in the output layer are optimized by backpropagating the mean square error ($\text{MSE} = \frac{1}{2} \sum_{i=1}^N (d_i - y_i)^2$) and changed by:

$$\Delta \beta_{j,k} = \eta \sum_{i=1}^N (d_i - y_i) H_{ji} \quad (6)$$

where N indicates the dataset length, η shows the learning rate parameter, and d_i and y_i are the desired and actual outputs, respectively. In ELMr, the weights in the output layer are calculated by linear regression and an error term is added such that Eq. (2) is written as follows:

$$\mathbf{y} = \mathbf{H}\beta + \epsilon \quad (7)$$

where ϵ is an error matrix. On the other hand, the SaELM employs the differential evolution (DE) method for optimizing network parameters [19]. In the SaELM method, the self-

adaptive DE is utilized to determine the input weights and hidden node biases with the ELM method being used to develop the output weights. Initially, the self-adaptive DE algorithm is used to generate random N_P vectors $\theta_{k,G}$ as populations in the first generation. In the G th generation, the i th parameter vector can be written as:

$$\theta_{kG} = [\theta_{kG}^1, \theta_{kG}^2, \dots, \theta_{kG}^D] \quad (8)$$

where $i = 1, 2, \dots, N_P$ and vectors are generated randomly through the following:

$$\theta_{k,G} = \theta_{\min} + \text{rand}(0, 1) \cdot (\theta_{\max} - \theta_{\min}) \quad (9)$$

where

$$\begin{cases} \theta_{\min} = [\theta_{\min}^1, \theta_{\min}^2, \dots, \theta_{\min}^D] \\ \theta_{\max} = [\theta_{\max}^1, \theta_{\max}^2, \dots, \theta_{\max}^D] \end{cases} \quad (10)$$

In this equation, θ_{\min} and θ_{\max} are the bounds of the considered parameters.

The weight matrix for the output is determined by the following equation:

$$\beta_{k,G} = \mathbf{H}_{k,G}^{-1} \mathbf{T} \quad (11)$$

where $\mathbf{H}_{k,G}^{-1}$ is the generalized inverse of $\mathbf{H}_{k,G}$ and can be written as:

$$\mathbf{H}_{k,G} = \begin{bmatrix} g(a_{1,(k,G)}, b_{1,(k,G)}, x_1) & \cdots & g(a_{L,(k,G)}, b_{L,(k,G)}, x_1) \\ \vdots & \ddots & \vdots \\ g(a_{1,(k,G)}, b_{1,(k,G)}, x_N) & \cdots & g(a_{L,(k,G)}, b_{L,(k,G)}, x_N) \end{bmatrix} \quad (12)$$

In addition, the root mean squared error (RMSE) of each individual is calculated as:

$$\text{RMSE}_{k,G} = \sqrt{\frac{\sum_{i=1}^N \left| \sum_{j=1}^L \beta_j g(a_{j,(k,G)}, b_{j,(k,G)}, x_i) - t_i \right|^2}{m \times N}} \quad (13)$$

The population vector with the best RMSE is stored in the first generation. In subsequent generations, the parameter vectors are evaluated using the following equation

$$\theta_{k,G+1} = \begin{cases} u_{k,G+1} & \text{if } \text{RMSE}_{\theta_{k,G}} - \text{RMSE}_{\theta_{k,G+1}} > \epsilon \cdot \text{RMSE}_{\theta_{k,G}} \\ u_{k,G+1} & \text{if } |\text{RMSE}_{\theta_{k,G}} - \text{RMSE}_{\theta_{k,G+1}}| < \epsilon \cdot \text{RMSE}_{\theta_{k,G}} \\ \theta_{k,G} & \text{else} \end{cases} \quad \text{and} \quad |\beta_{u_{k,G+1}}| < |\beta_{\theta_{k,G}}| \quad (14)$$

In the self-adaptive DE algorithm utilized herein, the trial vectors are generated by using one of the following four mutation strategies [21]:

Strategy 1:

$$\nu_{i,G} = \theta_{r_1^i,G} + F \cdot (\theta_{r_2^i,G} - \theta_{r_3^i,G}) \quad (15)$$

Strategy 2:

$$\begin{aligned} \nu_{i,G} = & \theta_{r_1^i,G} + F \cdot (\theta_{\text{best},G} - \theta_{r_1^i,G}) + F \cdot (\theta_{r_2^i,G} - \theta_{r_3^i,G}) \\ & + F \cdot (\theta_{r_4^i,G} - \theta_{r_5^i,G}) \end{aligned} \quad (16)$$

Strategy 3:

$$\nu_{i,G} = \theta_{r_1^i,G} + F \cdot (\theta_{r_2^i,G} - \theta_{r_3^i,G}) + F \cdot (\theta_{r_4^i,G} - \theta_{r_5^i,G}) \quad (17)$$

Strategy 4:

$$\nu_{i,G} = \theta_{i,G} + F \cdot (\theta_{r_1^i,G} - \theta_{i,G}) + F \cdot (\theta_{r_2^i,G} - \theta_{r_3^i,G}) \quad (18)$$

where r_k^i are integers obtained randomly within the range $[1, 2, \dots, N_P]$ interval. The strategy choice at each generation is accomplished according to a probability procedure $P_{l,G}$. The $P_{l,G}$ is the probability that the l th strategy is selected in the G th generation. In the developed model, l can be 1, 2, 3, or 4. The $P_{l,G}$ is updated such that if G is less than or equal P (number of generated vectors in each population), the four considered strategies have equal probabilities and $P_{l,G} = 0.25$. Else, if G is bigger than P , then $P_{l,G}$ is obtained from the following equation:

$$P_{l,G} = \frac{S_{l,G}}{\sum_{l=1}^4 S_{l,G}} \quad (19)$$

Where

$$S_{l,G} = \frac{\sum_{g=G-P}^{G-1} n_{S_{l,g}}}{\sum_{g=G-P}^{G-1} n_{S_{l,g}} + \sum_{g=G-P}^{G-1} n_{f_{l,g}}} + \varepsilon \quad (20)$$

where $n_{f_{l,g}}$ is the trial vectors that are entered in the coming generations, $n_{S_{l,g}}$ is the number of trial vectors that are discarded from the coming generations, and ε is a positive constant to prevent the zero improvement rate. The F and CR parameters are chosen for each target vector by selection from the normal distribution function. The generation of the trial vectors for the next generation is accomplished by using the $\theta_{k,G+1}$ equation that is presented before. In the SaELM, the evolution continues until the specified fitness is achieved.

The initialization step is similar in ELM, tELM, and ELMr, but in ELM, tELM, and ELMr, the output weights are calculated by the Moore–Penrose

generalized inverse method, backpropagation, and linear regression, respectively (Fig. 2).

2.2 Procedure for predictive model development

The following steps are followed for the predictive model development using ELM:

1. Non-dimensionalize the input and output variables
2. Specify the number of network features (input variables)
3. Specify the number of neurons in hidden layer (output variables)
4. Define the network parameters as follows: population size, weights, mutation rates, crossover constant, hidden layer neurons biases, and the termination criteria
5. Choose an activation function
6. Initialize the problem by randomly generating the parameters in the hidden node \mathbf{a}_j and \mathbf{b}_j for $j = 1, \dots, J$
7. Construct the $H(\mathbf{x})$
8. Train the network by calculating the output weights using either the Moore–Penrose generalized inverse method (ELM), or backpropagation (tELM) and linear regression (ELMr), or by DE (SaELM)
9. The trained network weights and biases are utilized to generate the ELM model
10. The developed ELM model is scored against selected error indicators. These indicators are the square of the Pearson product moment correlation coefficient (R^2), the root mean square error (RMSE), coefficient of efficiency (E_{sn}), and index of agreement (D). The indicators are calculated by the following equations [22]:

$$R_i^2 = \left(\frac{\frac{1}{n} \sum_{j=1}^n (T_j - \bar{T})(P_{(ij)} - \bar{P})}{\sqrt{\sum_{j=1}^n (T_j - \bar{T})^2 / n} \sqrt{\sum_{j=1}^n (P_{(ij)} - \bar{P})^2 / n}} \right)^2 \quad (21)$$

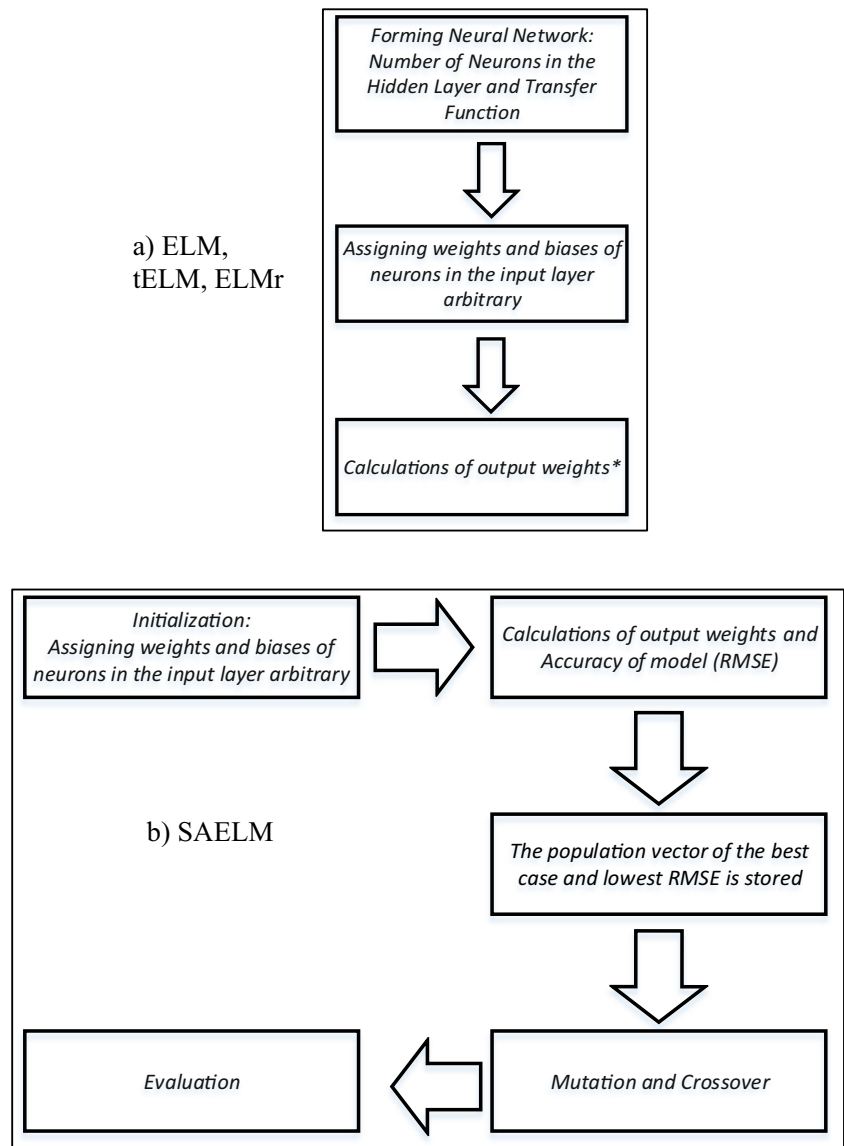
$$\text{RMSE} = \sqrt{E[(P-y)^2]} \quad (22)$$

$$E_{\text{sn}} = 1 - \frac{\sum_{i=1}^n (T_i - P_i)^2}{\sum_{i=1}^n (T_i - \bar{T})^2} \quad (23)$$

$$D = 1 - \frac{\sum_{i=1}^n (T_i - P_i)^2}{\sum_{i=1}^n (|P_i - \bar{T}| + |T_i - \bar{T}|)^2} \quad (24)$$

where $\bar{P} = 1/n \sum_{j=1}^n P_j$, P is the predicted value, and y is the observed value.

Fig. 2 Flow chart of **a** ELM and **b** SAELM networks



11. The performance of the developed ELM is validated by the following measures as utilized by Sattar [23] and Sattar and Gharabaghi [24]:

$$k = \sum_{i=1}^n (T_i \times P_i) / P_i^2 \text{ or } k' = \sum_{i=1}^n (T_i \times P_i) / T_i^2 \approx 1 \quad (25)$$

$$m = (R^2 - R_0^2) / R^2 \text{ and } n = (R^2 - R_0'^2) / R^2 < 0.1 \quad (26)$$

$$R_m = R^2 \times \left(1 - \sqrt{|R^2 - R_0^2|}\right) > 0.5 \quad (27)$$

where k and k' are the regression line gradients for observed versus predicted values, m and n are the regression line coefficient of determination, R_0^2 and $R_0'^2$ are the predicted and observed values correlation coefficients.

2.3 2.3. Uncertainty analysis of predictions of ELM models

Watermain failures are not a uniform process with constant rate but they are based on various parameters that lead to substantial variations between water distribution networks [25]. Therefore, it is expected that there will be some uncertainties in the predictions of any developed model. The availability of a watermain failure prediction model in addition to the expected uncertainty range of predictions would be a valuable tool for decision makers. Many recent models are reported [26–30] to have less uncertainty than other models. This can be accomplished by using the developed ELM models with the Monte Carlo simulation (MCS) method. The MCS is an easy to implement numerical method to determine the uncertainty of a model due to the combination of uncertainty

of various inputs. The MCS is capable of handling various probability distribution types of uncertain inputs [23, 31]. For running a stochastic analysis using MCS, thousands of realizations are needed and in each realization, the ELM model is used to predict a single deterministic output. Therefore, there are thousands of outputs which can be used to construct an output distribution and calculate the uncertainty associated with a parameter's median. The mean absolute deviation (MAD) is calculated as follows:

$$MAD = \frac{1}{250000} \sum_{i=1}^{250000} |P_i - \text{Median}(P)| \quad (28)$$

where the number of Monte Carlo realizations is taken 250,000 [1]. Afterwards, the predictive model uncertainty can be calculated as [32]:

$$\text{Uncertainty}\% = \frac{100 \times MAD}{\text{Median}(P)} \quad (29)$$

After calculating the prediction uncertainty, the least square linearization technique is used to determine the influence of various parameters on the output (details can be found in [22]). This is achieved by performing regression between the model output and each variable deviation from the mean.

$$y = w_1 \Delta v_1 + w_2 \Delta v_2 + \dots + w_i \Delta v_i + b \quad (30)$$

where y is the time to the next pipe failure; v_i are the pipe attribute inputs; $\Delta v_i = v_i - m_{vi}$ is the difference between v_i , the random pipe attribute input \underline{i} , and the mean value of all specific pipe attribute samples m_{vi} . Initially, random samples of input variables are used as inputs to the model yielding a single output y . This output (time to watermain failure for a particular pipe) is calculated for m Monte Carlo realizations. Using linear regression analysis, the regression coefficients w_i are calculated between the watermain time to failure and the input variables. Thus, the influence of each input variable \underline{i} (S_{V_i}) can be expressed as:

$$S_{V_i} = 100 \times w_i^2 \sigma_{\Delta v_i}^2 / \sum_{i=1}^n w_i^2 \sigma_{\Delta v_i}^2 \quad (31)$$

where $\sigma_{\Delta v_i}^2$ is the variance of Δv_i , and n is the number of random samples.

3.3. Results and discussions

3.1 Pipe failures in Greater Toronto Area

The Greater Toronto Area has more than 6000 km of drinking water network. The average age of the pipes is 50 years. Seventeen percent of the network is reaching 80 years in age

and 6.5% reaching more than 100 years. The data on watermain failure has been continuously recorded by the district of Scarborough, in the eastern part of the Greater Toronto Area. The data covers pipe failures from 1962 to 2005 with multiple breaks of the same pipe documented up to the 10th break for some pipes. The database consists of important data on pipe failures including the location of the failure, pipe length and diameter, and year of construction. The database includes important information regarding pipe coating or cathodic protection and year and the date of successive pipe failures. The pipe network of the district of Scarborough in Greater Toronto has 6342 watermains and has a cumulative length of more than 1000 km and installation began in 1905. The recording of pipe failures started in the year 1962 and contains data on successive breaks in an individual pipe till 10th break. The pipe material is either ductile iron (DI), cast iron (CI), or asbestos cement (AC). Pipe length ranges from 0.50 to 1.6 km and diameter from 30 to 500 mm. There are 3497 pipes that did not fail before, while there are 2845 pipes that have failed at least one time.

The majority of the Scarborough network pipes are cast iron (CI), with almost 60% of the network, and 30% for ductile iron (DI) and 10% for asbestos cement (AC). Therefore, the analysis of failures for cast iron and ductile iron pipes would cover 90% of the network. The statistics of the pipe failures are presented in Table 1.

Figure 3 shows the failure rates of watermains made of DI and CI normalized by specific pipes' length. The normalized pipe failure rate for the CI pipes is relatively higher than that for the DI with an average value of 0.32 for CI compared to 0.14 for DI. A similar value of normalized failure rate of 0.10

Table 1 General statistics of pipe failures database

Attributes	Asbestos cement (AC)	Cast iron (CI)	Ductile iron (DI)
Total number of pipes	463	3793	1747
Construction year	1950–1991	1921–2000	1905–1996
Pipe length (m)	1–831	0.3–978	0.3–1634
Average pipe length (m)	194	153	171
Total pipe length (km)	90	582	299
Diameter (mm)	150–400	42–1500	100–600
Number of pipes in sand/gravel	173	589	209
Number of pipes in silt/clay	290	3204	1538
Number of pipes with CP	9	514	785
Number of pipes with CML	0	1801	137
Number of pipes with zero failures	337	1530	1300
Number of pipes with at least one failure	126	2263	447
Number of pipe failures in total	248	8204	1454

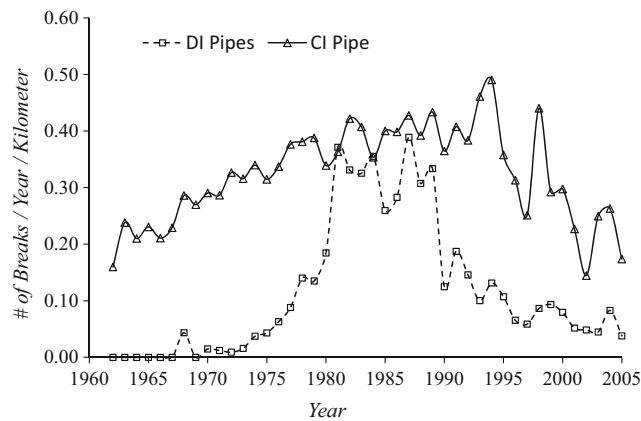


Fig. 3 The total number of DI and CI pipe failure rates per kilometer in Scarborough network

for DI has been reported in Canada [33]. While both pipe types experienced an increase in normalized failure rate with age, the gradient was steeper for DI pipes within the first 10 years after installation and remained steady afterwards until 1990. Moreover, a similar trend of decrease in normalized failure rates is observed for both pipe types starting from 1990, reaching to normalized rates of 40 years ago in 1960.

The city of Scarborough started implementing cathodic protection (CP) in 1986; this was accompanied by the application of cement mortar lining (CML) the following year. The CML process involves the cleaning of the rust from the inside of a pipe and applying a cement coating layer on the internal pipe surface. On the other hand, the CP attaches zinc anodes to the metallic surface of the pipes. The number of watermain failures started to decrease starting from 1990 after implementing these protection methods as shown in Fig. 4. The findings also show that these protection techniques are more effective in decreasing the DI pipe failure as compared to CI pipes with 80 and 60%, respectively.

Figure 5 shows the number of watermain failures per kilometer for DI and CI pipes versus the number of multiple pipe breaks for each pipe. The first pipe break is denoted by B1 and the second break by B2 and so forth. It is observed that the circumferential failure is the main failure type for CI pipes, while the hole failure is the main in DI pipes constituting more than 90% of the failure types in the network. The CI pipes tend to show higher failure rates in winter months, January and February [29, 34]. This is due to the external applied circumferential pressure exerted on the pipe circumference under the effect of frozen ground, where pipes tend to break more easily. Unlike the nonhomogeneous CI pipes, the DI pipes can resist externally applied pressure and thus experience fewer circumferential failures. DI pipes tend to break in localized areas caused by corrosion pitting and weakening the pipe material [34].

Considering the average age of various pipes when they first failed (Fig. 6), it can be seen that the DI has a lower age

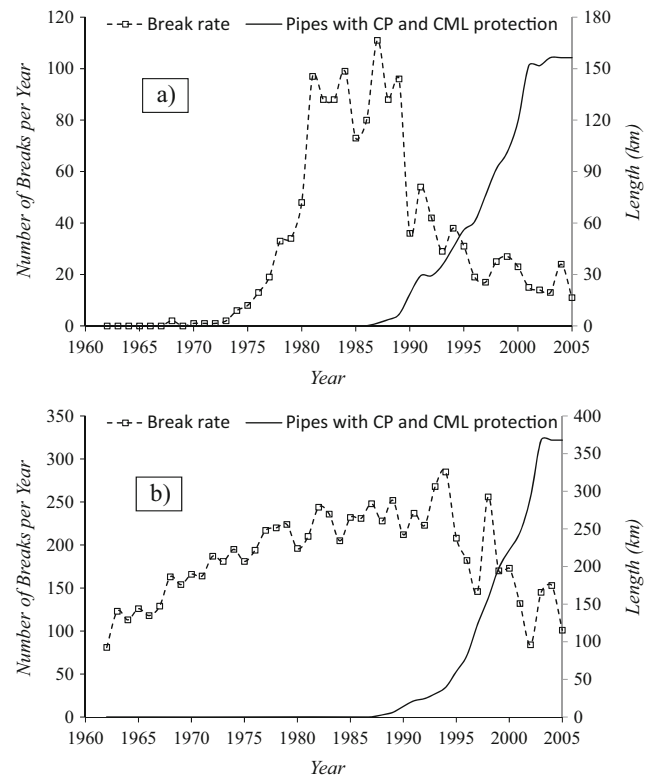


Fig. 4 Impact of CML and CP on the total number of watermain failures each year in Scarborough for **a** DI pipes and **b** CI pipes

at first failure than that of the CI. The DI pipe average age of 16 years was recorded versus 22 years for CI pipe. Folkman [33] and Rajani et al. [34] reported similar findings in other networks in the Greater Toronto. This finding is due to the nature of the soil around the pipes that triggers the pipe corrosion that mainly affects DI pipes leading to pitting and hole failures. This average age at the first recorded failure implies that the soil has moderate corrosion [34]. Figure 6 also shows that the average time between subsequent failures for CI is 2.5 years, which is relatively larger than that of 1 year for DI. This indicates that, when a DI breaks for the first time,

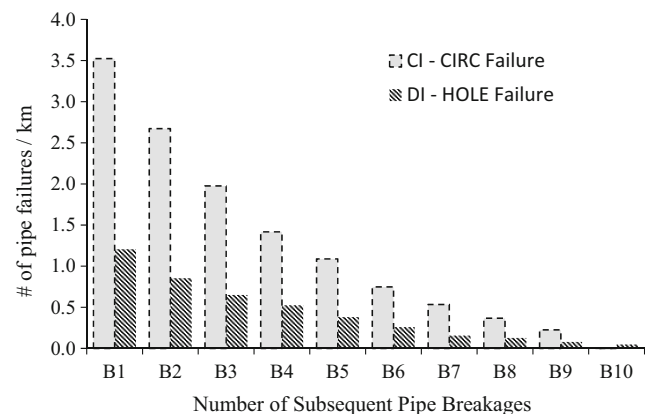


Fig. 5 Number of pipe failures per kilometer for main failure types in Scarborough network

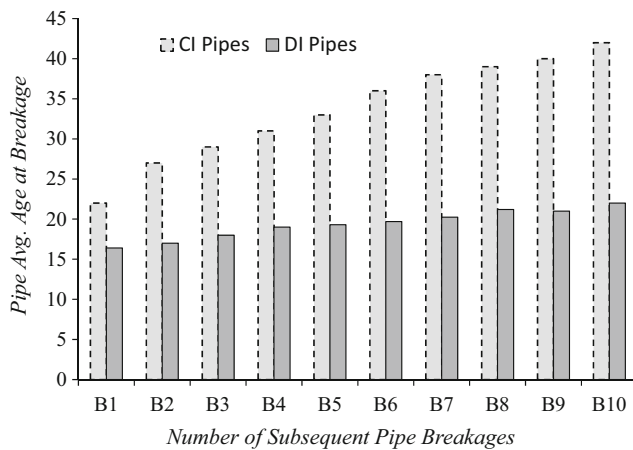


Fig. 6 Scarborough watermain average age at first and subsequent breakages

the frequency of subsequent breaks per year exceeds that of a CI in the same network.

Table 2 Pipe-specific attributes used in ELM model development

Material	Attribute	Type	Range
Asbestos cement (248 instances of failure)	Diameter	Numeric	150–400 mm
	Length	Numeric	18.79–828.55 m
	Construction year	Numeric	1952–1984
	Soil type	Nominal	0—Sand/gravel (109 instances) 1—Silt/clay (139 instances)
	Previous failures	Numeric	0–9 Failures
	Time to failure	Numeric	0–48 years
Cast iron (7795 instances of failure)	Diameter	Numeric	50–600 mm
	Length	Numeric	1.82–970.29 m
	Construction year	Numeric	1946–1985
	Soil type	Nominal	0—Sand/gravel (598 instances) 1—Silt/clay (7197 instances)
	CML protection	Nominal	0—no CML (7414 instances) 1—CML (381 instances)
	Cathodic protection	Nominal	0—No CP (7744 instances) 1—CP (51 instances)
Ductile iron (1453 instances of failure)	Previous failures	Numeric	0–9 Failures
	Time to failure	Numeric	0–56 years
	Diameter	Numeric	100–400 mm
	Length	Numeric	1.98–1260.10 m
	Construction year	Numeric	1957–1985
	Soil type	Nominal	0—Sand/gravel (99 instances) 1—Silt/clay (1354 instances)
	CML protection	Nominal	0—No CML (1446 instances) 1—CML (7 instances)
	Cathodic protection	Nominal	0—No CP (1377 instances) 1—CP (76 instances)
	Previous failures	Numeric	0–9 Failures
	Time to failure	Numeric	0–32 years

3.2 Development of new predictive equation

For pipe failure rate prediction, the objective is to construct an intelligent model employing the ELM algorithm that can perform better than available prediction models. The instances of watermain failures have been collected using pipes installed from 1946 up to 2005. Pipe failures are collected from the recorded dataset as per Harvey et al. [35–37] and Sattar et al. [1]. A total of 9508 watermain failures have been collected including all pipes that have failed at least one time during the period where data were collected. The watermain pipes types with their attributes are presented in Table 2.

According to Sattar et al. [1], the watermain failure is a function of the following variables:

$$\text{Watermain time to failure} = f(L, D, N_B, CML, CP) \quad (32)$$

where L is the pipe length, D is the pipe diameter, N_B is the number of previous pipe breaks, CML is the cement lining

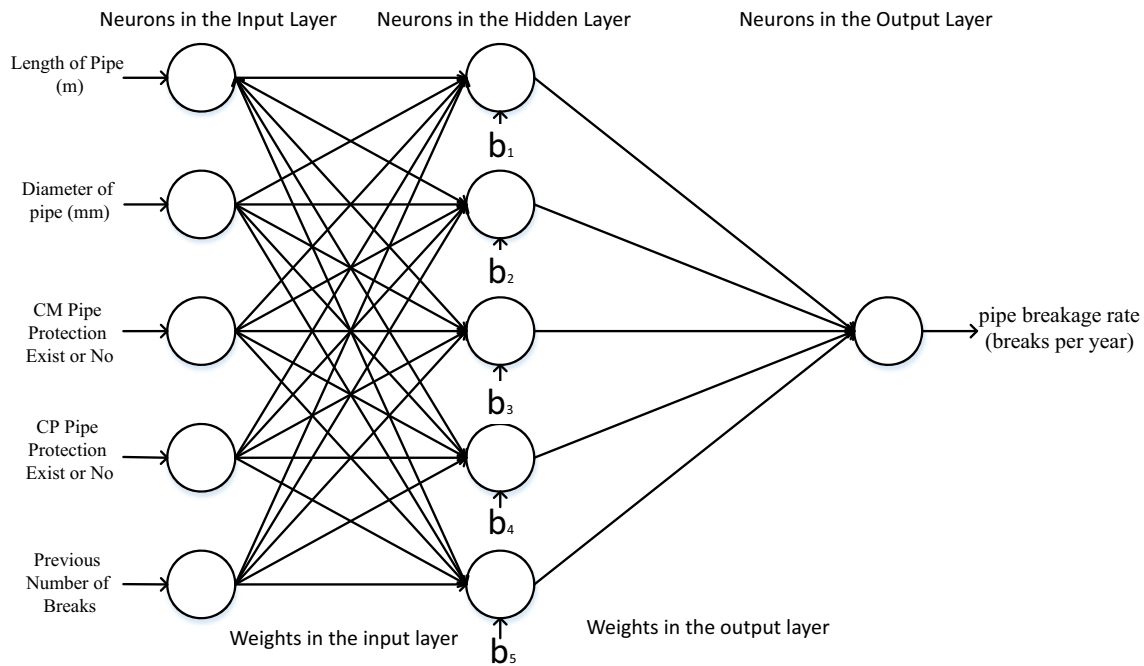


Fig. 7 Formulation of ELM network for prediction of watermain failure timing

protection, and CP is the cathodic protection. These are considered the input variables to the ELM network as shown in Fig. 7. The pipe failure recorded dataset has been split into training and test sets. Of the 9508 pipe break instance, 7131 (75%) were used to train the ELM network and 2377 (25%) were used to test and validate the developed model. Fourfold cross validation was utilized to validate the developed ELM.

3.3 Finding optimal ELM parameters

The optimum ELM network parameters can be grouped in the hidden layer neurons and the transfer function. Choice of such parameters is based on the user experience and the statistical performance of the developed model. The increase in the number of neurons increases the complexity of the developed model and, many times, it is at the expense of accuracy. The ELM models for failure time of AC and DI pipes were found to give the best results with a number of neurons less than 20,

while it required 50 neurons to produce the best results for DI. Regarding the transfer function, the hard limit function gave the least accurate ELM models, while the triangular basis function gave the best results. Other transfer functions such as the sin, the sigmoid, and the radial basis gave comparable or better results than the hard limit function. The chosen ELM model for predicting failure time of AC pipes had five neurons and based on the radial basis transfer function. However, the ELM models for predicting failure time for CI and DI pipes were based on the triangular basis transfer function with 50 and 20 neurons, respectively.

The developed ELM models performed well (Table 3), with R^2 of 0.46, 0.43, and 0.64 for AC, CI, and DI, respectively in training, and 0.43, 0.41, and 0.63 in testing. Testing R^2 and RMSE were based on fourfold cross validation where the ELM model is validated on the testing dataset (25% of total data) and then validated on the other three equal sets (each 25%). The RMSE associated with ELM models range

Table 3 Statistics of developed ELM and ELM variants on training and testing datasets

Pipe type	Error terms	ELM		SaELM		tELM		ELMr	
		Train	Test	Train	Test	Train	Test	Train	Test
AC	R^2	0.459	0.388	0.435	0.368	0.446	0.377	0.403	0.304
	RMSE	0.173	0.187	0.180	0.192	0.175	0.183	0.189	0.203
CI	R^2	0.426	0.418	0.416	0.381	0.399	0.366	0.384	0.377
	RMSE	0.134	0.135	0.148	0.156	0.157	0.177	0.151	0.169
DI	R^2	0.646	0.633	0.635	0.587	0.653	0.587	0.625	0.567
	RMSE	0.085	0.087	0.090	0.098	0.092	0.099	0.099	0.117

from 0.09 to 0.17, which are low and similar for both training and testing datasets. The R^2 and RMSE for training and testing cases are low and also similar for training and testing. These values indicate that the ELM model has an acceptable predictive performance. Other variants of the ELM have also been used. Accuracies obtained by SaELM, tELM, and ELMr are as shown in the same table. It is observed that the ELM scored the highest R^2 values and lowest RMSEs. For AC pipes, the tELM had the closest score to the ELM model, while the SaELM had close values to the ELM model in the case of CI pipe. In the case of DI pipe, tELM had higher R^2 than the chosen ELM model. On the other hand, the E_{sn} and D showed very good values for the ELM models compared to ANN and other methods with values of 0.44 (E_{sn}) and 0.82 (D) for AC, DI, and CI pipes.

Further testing and validation for the developed ELM models has been performed with results presented in Table 4. While the ELM model is considered a good one if it satisfied, one or more of the required validation conditions, it is observed that the developed ELM models for all pipe types satisfied all of the proposed tests confirming they have good prediction ability.

Consider now the performance of the developed ELM models in comparison with other machine learning methods, namely artificial neural networks (ANN), support vector machines (SVMs), and non-linear regression (NNR), as presented in Table 5. The developed ELM models show better performance than other machine learning models applied on the same dataset in terms of not only R^2 and RMSE but also, process time. All tests were completed in MATLAB with an Intel Core i7-2600 CPU, 3.4 GHz, 4 GB RAM, PC. SVM analysis was performed using the toolbox PrTools (www.prtools.org).

3.4 Sensitivity analysis

Further analysis is performed to test the sensitivity of various input parameters to the pipe failure time prediction ELM model. These inputs are the L , D , CP , CML , P , and N_B . These input parameters are fitted to probability distribution and unreal variables have been removed by truncating distributions. The truncated distribution limits have been constructed from the current dataset values. Various distributions for input variables

Table 5 Statistics of developed ELM model versus some popular machine learning machines

Pipe type	Measure	ELM	ANN	SVM	NNR
AC	R^2	0.46	0.42	0.08	0.17
	RMSE	0.17	0.24	0.24	0.26
	E_{sn}	0.51	0.48	0.10	0.15
	D	0.85	0.77	0.08	0.11
	PC time (s)	0.001	0.20	1.40	0.01
CI	R^2	0.43	0.40	0.21	0.18
	RMSE	0.13	0.14	0.16	0.19
	E_{sn}	0.46	0.44	0.21	0.20
	D	0.78	0.81	0.44	0.37
	PC time (s)	0.01	0.17	1.67	0.01
DI	R^2	0.65	0.61	0.33	0.42
	RMSE	0.09	0.11	0.13	0.12
	E_{sn}	0.68	0.59	0.32	0.37
	D	0.85	0.80	0.50	0.65
	PC time (s)	0.002	0.07	2.64	0.009

are ranked based on the Anderson Darling and chi-squared tests [31]. This resulted in using the exponential distribution for modeling L and D , and the Poisson distribution for CP and CML . Following the calculation of various realizations for time to failure, the multiple regression analysis is used to construct the following equation:

$$T_f = w_1 \Delta L + w_2 \Delta D + w_3 \Delta CP + w_4 \Delta CML + w_5 \Delta N_B + b \quad (33)$$

Results showed that using the developed ELM models, the predictions of time to next failure shows a MAD of 3, which is 36% of the median value. This is an acceptable uncertainty in model predictions according to Verbeeck et al. [38] and Sattar [23], with values up to 40% accepted.

Using the least square linearization to determine the parameter sensitivity showed the various input parameters importance (Table 6). The highest influential parameter on the ELM model prediction is shown to the number of previous pipe breaks. This agrees with what has been reported by Goulter and Kazemi [39], Asnaashari et al. [40–42], and Sattar et al. [1], where availability of previous failures increased pipe failure rates over time. Pipe diameter came

Table 4 External validation for developed ELM models

Model	K ($0.85 < K < 1.15$)	K' ($0.85 < K' < 1.15$)	m ($m < 0.1$)	n ($n < 0.1$)
AC-ELM- T_f	0.80	0.91	-0.49	-0.42
CI-ELM- T_f	0.80	0.91	-0.49	-0.42
DI-ELM- T_f	0.77	0.76	-0.25	-0.18

Table 6 Importance of various pipe parameters as predicted by ELM model

Attribute	CI	DI	AC
Length	2.1	0.30	0.75
Diameter	28.25	16.78	0.21
CML	0.12	0.63	3.70
CP	0.42	0.30	3.09
N_B	69.11	81.99	92.25

second to N_B with more influence on time to pipe failure than pipe length. Protection methods had less effect on the output uncertainty. This is confirmed by results that also show that the CP protective effect is more pronounced than that for CML for different pipe types. These outputs are generally in agreement with Harvey et al. [36] and Sattar et al. [1].

3.5 Parametric analysis of developed ELM model

This section presents the parametric analysis for the developed ELM model. This parametric analysis helps to determine the behavior of the model and influence of various input parameters, mainly pipe diameter, length, and previous failures, on the predicted time to failure. Figure 8 shows the time to pipeline failure as predicted by the ELM model with the ratio of diameter and length for the three types of pipes, CI, DI, and AC. The predicted time to failure of a pipe is observed to decrease with longer pipes than for shorter ones for three types of pipes. This is due to the fact that longer pipes are subject to

various possible external conditions that can affect their integrity such as traffic loads [43]. The same finding has been reported by Lei [44], Wang et al. [45], and [1, 46–48]. Predicted time-to-failure is higher for cast iron pipes than ductile iron and asbestos cement pipes. This is attributed to the non-homogeneity of CI pipe material, the same fact that makes this type of pipes prone to failures, unlike DI and AC. Furthermore, ELM model predictions showed that the time to next failure is directly proportional to the pipe diameter. This has been confirmed by Rostum [43] who related the larger time to failure with higher pipe diameter. He attributed this to reduced pipe strength and less reliable joints of smaller diameter pipes.

Predictions of the ELM are consistent with the historical trends observed in the city dataset. A significant increase in the next time of failure of a pipeline is predicted with the application of one or two types of pipe protection for three types of pipes. The CP protection is shown to be more effective at increasing the time of pipes before the next failure than the CML protection. This is true for pipe types DI and CI where the CP protection increased the time to next failure by more than 15% in case of CML protection. However, this is not true for the AC pipes where the application of CP showed the same effect as CML. In all pipe types, the effects of CP and CML protection is additive leading to an increase in the time to next failure. This specific behavior is related to the studied network since various impacts of CP and CML protection have been reported for other networks under different conditions of soil corrosiveness, temperature, and installation methods that have impacts on the coatings [34].

4 Conclusions

In this study, the extreme learning machine method has been used on more than 9500 pipe failure instances in the city of Scarborough, Canada to develop a new model that can predict the time to next watermain failure. The developed ELM model indicated results with coefficient of determination ranging from 0.67 to 0.82. This was achieved with a maximum of 50 neurons in the network hidden layer and the triangular basis function. Other variants of the ELM modeling methods were attempted namely tELM, ELMr, and SaELM. Error results showed the superiority of the ELM models over its variants on the study case. The ELM model has the advantage of including the type of protection of pipes and incorporates its influence on the predicted results in addition to pipe diameter, length, and previous failures. The number of previous pipe breaks is shown to be the most influential input parameter to the ELM predictions followed by pipeline diameter. Moreover, the CP pipe protection was found to be more effective in protecting pipes and decreasing their failure rate. The ELM model can be used as a tool to help decisions on

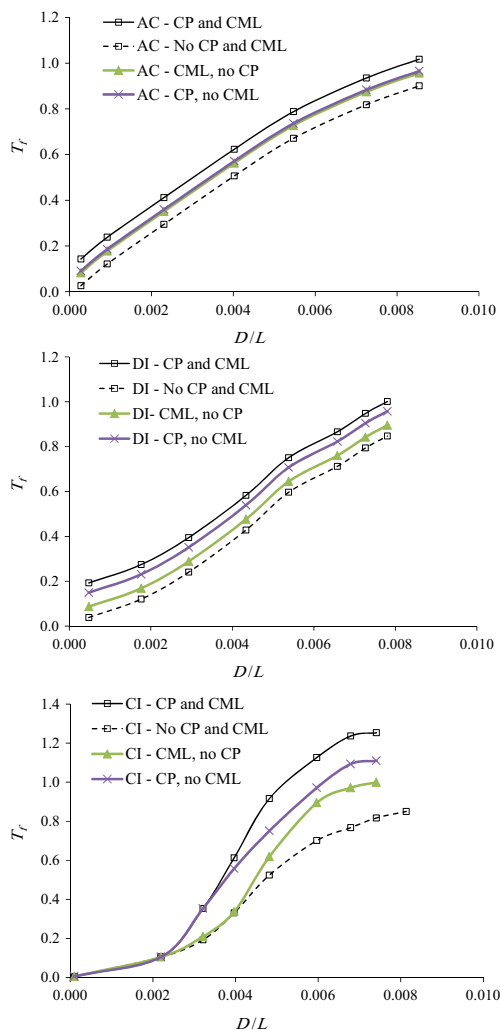


Fig. 8 T_f versus pipe length and pipe diameter for developed ELM models

optimum pipe inspection and maintenance schedule to proactively control the rising maintenance cost of the aging infrastructure and also improve the reliability and safety of the essential service to the public.

Acknowledgments The authors would like to thank the district of Scarborough for their contribution in the data collection phase and funding by the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chairs program.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Sattar AM, Gharabaghi B, McBean E (2016) Predicting timing of watermain failure using gene expression models for infrastructure planning. *Water Resour Manag* 30(5):1635–1651
- Schuster C, McBean E (2008) Impacts of cathodic protection on pipe break probabilities: a Toronto case study. *Can J Civ Eng* 35(2): 210–216
- Nishiyama M, Filion Y (2013) Review of statistical water main break prediction models. *Can J Civ Eng* 40:972–979
- Kleiner Y, Sadiq R, Rajani B (2006) Modelling the deterioration of buried infrastructure as a fuzzy Markov process. *J Water Supply Res Technol AQUA* 55(2):67–80
- Huang G, Zhu Y, Siew C (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70:489–501
- Atieh M, Mehlretter S, Gharabaghi B, Rudra R (2015a) Integrated neural networks model for prediction of sediment rating curve parameters for ungauged basins. *J Hydrol* 531(3):1095–1107. doi:10.1016/j.jhydrol.2015.11.008
- Atieh M, Gharabaghi B, Rudra R (2015b) Entropy-based neural networks model for flow duration curves at ungauged sites. *J Hydrol* 529(3):1007–1020. doi:10.1016/j.jhydrol.2015.08.068
- Atieh M., Taylor, G., Sattar, A. M., & Gharabaghi, B. (2017). Prediction of flow duration curves for ungauged basins. *Journal of Hydrology*. Volume 545, February 2017, Pages 383–394, DOI: 10.1016/j.jhydrol.2016.12.048.
- Cao J, Xiong X (2014) Protein sequence classification with improved extreme learning machine algorithms. *Biomed Res Int*. doi:10.1155/2014/103054. Epub 2014 Mar 30
- Ding S, Zhang J, Xu X, Yanan Z (2015) A wavelet extreme learning machine. *Neural Comput & Applic* 27(4):1033–1040
- Ding SF, Xu XZ, Nie R (2014) Extreme learning machine and its applications. *Neural Comput. Appl* 25(3):549–556
- Ertugrul O, Kaya M (2014) A detailed analysis on extreme learning machine and novel approaches based on ELM. *American Journal of Computer Science and Engineering* 1(5):43–50
- Gazendam, E., Gharabaghi, B., Ackerman, J., & Whiteley, H. (2016). Integrative neural networks models for stream assessment in restoration projects. *Journal of Hydrology*, 536 (2016) 339–350. DOI: 10.1016/j.jhydrol.2016.02.057.
- Lian C, Zeng Z, Yao W, Tang H (2014) Ensemble of extreme learning machine for landslide displacement prediction based on time series analysis. *Neural Comput & Applic* 24(1). doi:10.1007/s00521
- Luo M, Zhang K (2014) A hybrid approach combining extreme learning machine and sparse representation for image classification. *Journal Engineering Applications of Artificial Intelligence* archive 27:228–235
- Man Z, Huang G (2016) Guest editorial: special issue of extreme learning machine and applications. *Neural Comput & Applic* 27(1): 1–2
- Sabouri F, Gharabaghi B, Sattar A, Thompson AM (2016) Event-based stormwater management pond runoff temperature model. *Journal of Hydrology* 540(2016):306–316. doi:10.1016/j.jhydrol.2016.06.017
- Trenouth WR, Gharabaghi B (2016) Highway runoff quality models for the protection of environmentally sensitive areas. *Journal of Hydrology* Volume 542(November 2016):143–155
- Cao J, Lin Z, Huang G, Liu N (2012) Voting based extreme learning machine. *Inf Sci* 185(1):66–77
- Huang G, Zhou Y, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern B Cybern* 42(2):513–529
- Storn R, Price K (1997) Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J Glob Optim* 11(4):341–359
- Sattar AM (2014a) Gene expression models for the prediction of longitudinal dispersion coefficients in transitional and turbulent pipe flow. *J. Pipeline Syst. Eng. Pract.* ASCE 5(1):04013011
- Sattar AM (2016b) A probabilistic projection of the transient flow equations with random system parameters and internal boundary conditions. *J Hydraul Res*. doi:10.1080/00221686.2016.1140682
- Sattar AM, Gharabaghi B (2015) Gene expression models for prediction of longitudinal dispersion coefficient in streams. *J Hydrol* 524:587–596
- Al-Barqawi M, Zayed T (2006) Condition rating model for underground infrastructure sustainable water mains. *Journal of Performance of Constructed Facilities*, ASCE. 20(2):126–135
- El Hakeem M, Sattar AM (2015) An entrainment model for non-uniform sediment. *Earth Surf Process Landf*. doi:10.1002/esp.3715
- Najafzadeh M, Sattar AM (2015) Neuro-fuzzy GMDH approach to predict longitudinal dispersion in water networks. *Water Resour Manag* 29:2205–2219. doi:10.1007/s11269-015-0936-8
- Sattar AM, Dickerson J, Chaudhry M (2009) A wavelet Galerkin solution to the transient flow equations. *J Hydraul Eng* 135(4):283–295
- Sattar AM (2016a) Prediction of organic micropollutant removal in soil aquifer treatment system using GEP. *J Hydrol Eng*. doi:10.1061/(ASCE)HE.1943-5584.0001372 (in press)
- Thompson J, Sattar A, Gharabaghi B, Warner R (2016) Event-based total suspended sediment particle size distribution model. *J Hydrol* 536(2016):236–246
- Vose D (1996) Quantitative risk analysis: a guide to Monte Carlo simulation modeling. John Wiley, New York
- Walker H (1931) Studies in the history of the statistical method. Williams & Wilkins Co., Baltimore, MD, pp 24–25
- Folkman S (2012) Water main break rates in the USA and Canada: a comprehensive study, report, Utah State University buried structures laboratory, April 2012.
- Rajani B, Kleiner Y, Sink JE (2012) Exploration of the relationship between water main breaks and temperature covariates. *Urban Water* 9(2):67–84
- Harvey R, McBean EA, Murphy HM, Gharabaghi B (2015) Using data mining to understand drinking water advisories in small water systems: a case study of Ontario first nations drinking water supplies. *Water Resources Management* 29(14):5129–5139
- Harvey R, McBean E, Gharabaghi B (2014) Predicting the timing of water main failure using artificial neural networks. *J Water Resour Plan Manag* 140(4):425–434
- Harvey R, McBean EA, Gharabaghi B (2013) Predicting the timing of watermain failure using artificial neural networks. *J Water Resour Plan Manag* 140(4):425–434

38. Verbeeck H, Samson R, Verdonck F, Raoul L (2006) Parameter sensitivity and uncertainty of the forest carbon flux model FOUg: a Monte Carlo analysis. *Tree Physiol* 26:807–817
39. Goulter IC, Kazemi A (1998) Spatial and temporal groupings of water main pipe breakage in Winnipeg. *Can J Civ Eng* 15(1):91–97
40. Asnaashari A, McBean EA, Gharabaghi B, Tutt D (2013) Forecasting watermain failure using artificial neural network modeling. *Canadian Water Resources Journal* 38(1):24–33
41. Asnaashari A, McBean E, Gharabaghi B, Pourrajab R, Shahrou I (2010) Survival rate analyses of watermains: a comparison of case studies for Canada and Iran. *Journal of Water Management Modeling* 18(30):499–508
42. Asnaashari A, McBean EA, Shahrou I, Gharabaghi B (2009) Prediction of watermain failure frequencies using multiple and Poisson regression. *Water Sci Technol Water Supply* 9(1):9–19
43. Rostum J (2000) Master of Science Dissertation. In: Statistical modeling of pipe failures in water networks. Norwegian University of Science and Technology, Trondheim, Norway
44. Lei J (1997) Statistical approach for describing lifetimes of water mains - case Trondheim Municipality. STF22 A97320, SINTEF, Trondheim.
45. Wang Y, Moselhi O, Zayed T (2009) Study of the suitability of existing deterioration models for water mains. *Journal of Performance of Constructed Facilities*, ASCE 23(1):40–46
46. Sattar AM (2014b) Gene expression models for prediction of dam breach parameters. *Journal of Hydroinformatics*, IWA 16(3):550–571
47. Sattar AMA, El-Beltagy M (2017) Stochastic Solution to the Water Hammer Equations Using Polynomial Chaos Expansion with Random Boundary and Initial Conditions. *J Hydraul Eng* 143(2): 04016078
48. Ebtehaj I, Sattar AMA, Bonakdari H, Zaji AH (2017) Prediction of scour depth around bridge piers using self-adaptive extreme learning machine. *J Hydroinf* 19(2):207–224