ANN and ANFIS Modeling of Failure Trend Analysis in Urban Water Distribution Network

Libi P. Markose and Paresh Chandra Deka

Abstract Pipeline leakage is one of the crucial problems affecting urban water distribution system from both environmental and economical point of view. Unfortunately, necessary large databases are not maintained in India for proper replacement of pipes. In this situation, this research purports at using two artificial intelligence techniques such as artificial neural network (ANN) and adaptive neuro-fuzzy inference systems (ANFIS) to access the present condition and to predict the future trend of pipeline network of Peroorkada zone in Trivandrum city, Kerala, where a huge amount is spent every year for leakage rectification. Using different influential input variables, four models (all diameter pipes) have been developed. Also, the effect of each parameter (length, age, and diameter) along with previous year failures and previous year failures alone to current year failures are analyzed. Another two models of selective pipe diameter for considering the influence of prefailures up to last year alone and prefailures up to last year along with length to current year failures are constructed. Prioritizing the pipeline replacement is done for mains having 400 mm diameter and above since network details pertaining to those diameters are available. The performance of the models is evaluated using coefficient of correlation and mean absolute error and is compared to multiple linear regression (MLR) models. Three of them perform well and almost in kind, even though ANN is slightly having an upper hand. The applicability and usefulness of ANN and ANFIS will surely become beneficial for the authorities to take decisions regarding the replacement of pipes and this can in turn increase the efficiency of pipes.

Keywords Pipeline leakage · Failure prediction · ANN · Urban water distribution system · MLR · ANFIS

L.P. Markose (✉) · P.C. Deka
Department of Applied Mechanics and Hydraulics, National Institute of Technology, Srinivasanagar, Surathkal 575075, Karnataka, India
e-mail: libimarkose@gmail.com

P.C. Deka
e-mail: paresh_deka@sify.com

A.K. Sarma et al. (eds.), *Urban Hydrology, Watershed Management and Socio-Economic Aspects*, Water Science and Technology Library 73,
DOI 10.1007/978-3-319-40195-9_20

1 Introduction

Even with the improvement in technologies related to urban water supply, it has become common for cities to have a legion of water pipeline breaks each year. To ensure that the authorities can manage their water distribution systems to provide an adequate supply of water in a cost effective, reliable, and sustainable manner, it is essential that they develop a clear understanding of water pipelines deterioration process. Predictive modeling includes a collection of technologies that can be based on the massive data including pipe physical property, environmental factor, operational condition, and historical failure record.

Recently soft computing techniques like artificial neural networks (ANN) and adaptive neuro-fuzzy inference systems (ANFIS) have usurped many conventional methods in successful modeling of complex water resources systems. Unlike many hydrological applications, there are limited applications of ANN in modeling pipeline failure trend. Tabesh et al. (2009) used ANN and ANFIS to model pipe failure rate (number of accidents per year per unit length) by considering five input parameters and the results are compared with multivariate regression approach. RaedJafar and Juran (2010) presented an application of artificial neural networks (ANN) to model the failure rate of 4862 urban mains using a 14-year database collected in a city in the north of France and to estimate the optimal replacement time for the individual pipes in an urban water distribution system. Abdel Wahal M Budtiena et al. (2011) introduced the effectiveness of ANN to model pipe line breaks by considering ten input parameters.

Even though literatures have explored the usage of ANN to predict pipeline failure, the variation in performance of ANN by changing inputs and length of dataset are not considered and very few have used ANFIS for failure determination. Moreover no study regarding the analysis of failure trend in pipelines have been attempted in Peroorkada zone of Trivandrum city.

The main objectives of the present study are to analyze the trend in pipeline failure using ANN and ANFIS and to prioritize the renewal of network containing larger diameter pipelines with the best model.

2 Artificial Neural Network

Artificial neural networks are also referred to as “neural nets,” “artificial neural systems,” “parallel distributed processing systems,” and “connectionist systems.” A neural network paving its origin in nineteenth century is an information processing system which is biologically inspired, i.e., they are composed of elements that perform in a manner that is analogous to the most elementary function of the biological neuron. Feed-forward neural network (Fig. 1) with backpropagation algorithm is used for analysis. In this, the nodes are arranged in layers starting from the first input layer and ending at the final output layer. There can be several hidden

layers with each layer having one or more nodes. Information passes from the input to the outside side. The nodes in one layer are connected to those in the next, but not to those in the same layer. Thus output of a node in a layer is only dependent on the inputs it receives from previous layers and the corresponding weights. Backpropagation algorithm does its work in two steps. In the first step, each input pattern of the training dataset is passed through the network from the input layer to the output layer. The network output is compared with the desired target output and in the second step error is computed and is propagated back toward the input layer with the weights being modified. Backpropagation uses delta rule to adjust the connectivity weights. During training, weights need continuous adjustment from iteration t to $t + 1$. The adjustment $\Delta W(t+1)$, which is required in iteration $(t + 1)$, is assumed linearly related to the negative gradient of E with w in iteration t. The constant of proportionality in this linear relation is known as the learning rate (η). Mathematically this relation can be expressed as follows:

$$\Delta W(t+1) = \eta\left(-\frac{\partial E}{\partial W}\right)_{W=W(t)} \tag{1}$$

$$W(t+1) = W(t) - \eta\left(\frac{\partial E}{\partial W}\right)_{W=W(t)} \tag{2}$$

For improving the convergence the following modification is made in Eq. (2).

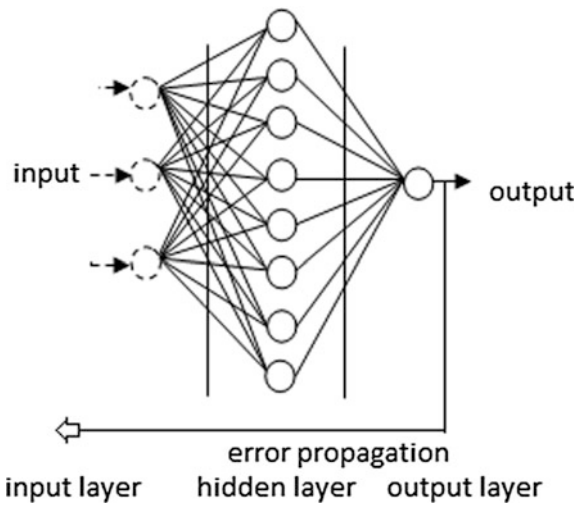

Fig. 1 Feed-forward backpropagation network

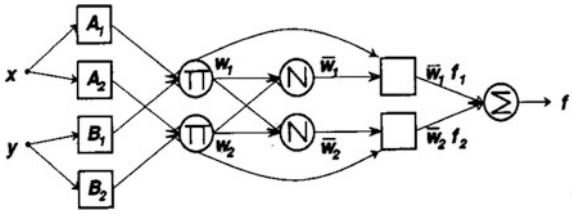

Fig. 2 Schematic of ANFIS architecture

$$W(t+1) = W(t) - \eta\left(\frac{\partial E}{\partial W}\right)_{W=W(t)} + \mu\Delta W(t) \tag{3}$$

where (μ) is the momentum factor as it imparts the momentum to the rate of convergence.

3 Adaptive Neuro-Fuzzy Inference System

Adaptive neuro-fuzzy inference system proposed by J.S.R Sang is a methodology to simulate complex nonlinear mappings utilizing neural network learning and fuzzy inference methodologies. It brings into service the learning capability of ANN for rule generation and parameter optimization. ANFIS architecture (Fig. 2) has five nodes out of which first and fourth are adaptive and others are fixed nodes. In the first layer each input (x/y) is converted to a membership values ranging from 0 to 1. Here A_1, A2, B1, B2 are linguistic labels. In the second layer all the incoming signals are multiplied together to get (w_1/w_2) and a normalized firing strength ($\bar{w}_1/\bar{w}_2$) is obtained in the third layer. In the subsequent layer firing strength will be multiplied with the function (f_1/f_2) and the last layer computes the summation (f) of all incoming signals.

4 Study Area

The study concerns the water distribution system of Peroorkada zone in Trivandrum city which comprises an area of 15.46 km^2. Total population as per 1991 census is 137714. Total water demand is 25.82 m^3/min. There is only one reservoir at Peroorkada which has a capacity of 8 million litres. The total domestic connection is 13700 and the number of household connections is approximately 18470. The Peroorkada network consists of 99 nodes and 114 pipes and 16 loops. The water distribution network map of Peroorkada zone is shown in Fig. 3.

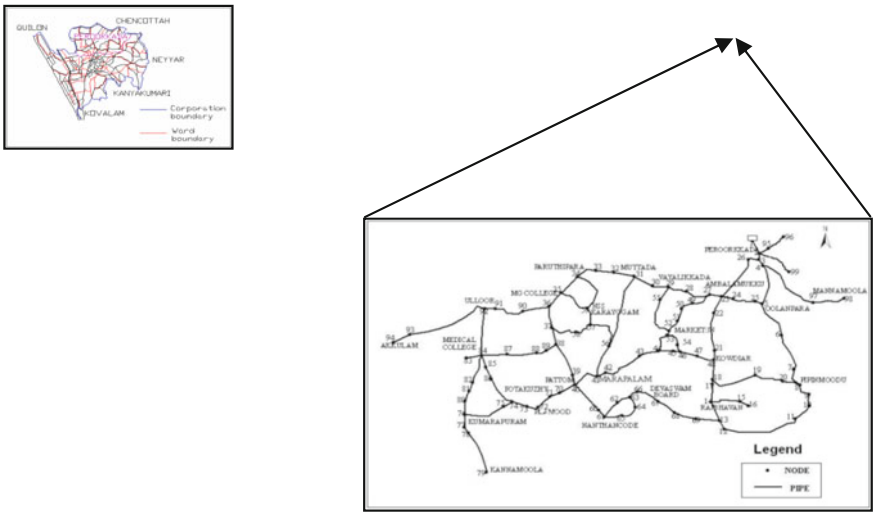

Fig. 3 Location and water distribution network of Peroorkada zone

5 Available Data and Model Structures

5.1 Available Data

The complete network containing diameters of pipe other than 400 mm prestressed concrete pipe is not available. The details (length, age, material, diameter, location, number of failures for the last ten years from 2000 to 2010) are available for 400 mm and for the main three pipes supplying water from Aruvikkara to Trivandrum city. These details are collected from the running contract leakage rectification available in Kerala Water Authority, Kowdiar, Aruvikkara, and Vellayambalam. For smaller diameter pipes, both the network details and the location of failure are not available even though failures happened in the past ten years for each of these pipes are there.

5.2 Model Structures

The dataset containing failures and prefailures are divided into two sets, i.e., testing and training data. As a total, 140 datasets are available for diameters lesser than 400 mm and 117 datasets are available for 400 mm diameter. Different lengths of testing data are considered to know the improvement in performance of the model.

In this study, for ANN a three-layered feed-forward backpropagation network (newff) has been selected based on the outcome of past literatures. To train the network the fastest training algorithm Levenberg Marquardt algorithm has been

selected to achieve the training speed. To know the optimal number of neurons in the hidden layer different trials are conducted. After training the networks are simulated for unknown values using testing sets. The predicted values obtained from the network are compared with the observed values using various performance evaluators and the best network is ranked.

The developed models are also trained using ANFIS. Different membership functions and epochs are considered to get the best possible results.

For the performance evaluation of the models, two statistical criteria namely mean absolute error and coefficient of correlation are used.

Mean absolute error

$$\mathrm{MAE} = \frac{\sum |X - Y|}{N} \tag{4}$$

Coefficient of correlation

$$\mathrm{CC} = \frac{\sum xy}{\sqrt{\sum x^2} * \sqrt{y^2}} \tag{5}$$

where
N number of datasets
X target output pattern value
Y output from neural net work model
x $X - X_{\mathrm{mean}}$, $y = Y - Y_{\mathrm{mean}}$

6 Results and Discussion

The efficacy of ANN and ANFIS to analyze the trend of failure is verified by means of six models and is shown in Table 1. The six models are constructed by varying the factors responsible for failure. The fifth and sixth models are constructed for 400 mm pipes. The modeling of ANN and ANFIS are coded using MATLAB 2010a. The feed-forward backpropagation network with single hidden layer is

Table 1 The structure of forecasting models

Model type	Inputs	Outputs
M1	Failures in the previous year	Failures in the present
M2	Diameter, failures in the previous year	Failures in the present
M3	Age, failures in the previous year	Failures in the present
M4	Diameter, age, failures in the previous year	Failures in the present
M5	Length, number of failures up to 2009	Failures up to 2010
M6	Number of failures up to 2009	Failures up to 2010

Table 2 Comparison of performance of ANN models in training with the length of dataset

Dataset (D)	10 % D		30 % D		25 % D	
Model	CC	MAE	CC	MAE	CC	MAE
M1 (1-7-1)	0.91	42.16	0.95	27.91	0.9	25.74
M2 (1-4-1)	0.91	35.56	0.95	29.73	0.94	26.63
M3 (1-2-1)	0.92	41.34	0.98	21.57	0.94	23.05
M4 (1-3-1)	0.94	32.84	0.95	31.42	0.93	25.28
M5 (1-3-1)	0.98	0.09	0.99	0.17	0.99	0.11
M6 (1-3-1)	0.98	0.15	0.99	0.14	0.98	0.18

Table 3 Comparison of performance of ANN models in testing with the length of dataset

Dataset (D)	10 % D		30 % D		25 % D	
Model	CC	MAE	CC	MAE	CC	MAE
M1 (1-7-1)	0.95	13.33	0.7	47.75	0.93	33.66
M2 (1-4-1)	0.95	15.39	0.81	36.94	0.93	36.24
M3 (1-2-1)	0.91	22.97	0.79	45	0.93	36.24
M4 (1-3-1)	0.93	23.72	0.81	35.2	0.93	33.1
M5 (1-3-1)	0.96	0.41	0.97	0.34	0.98	0.25
M6 (1-3-1)	0.95	0.51	0.95	0.28	0.96	0.26

performing best in majority of applications in the field of hydrology for ANN modeling, so it is been used here. The other parameters used are tangent sigmoid and linear activation function. The results obtained are shown by means of graphs and tables. Tables 2 and 3 show the variation in performance indices for testing and training while considering different lengths of testing dataset. The 10 and 30 % dataset for testing are taken after leaving first 90 and 70 % dataset for training, but the 25 % of dataset for testing is considered by starting with the second point and taking every fourth point. In majority of cases with 25 % dataset the models are giving best performance due to better training obtained by the model. The length of dataset which gives the best results for each of the models is considered. ANN is performing well in case of normalized data when compared with raw data for the first model (Fig. 4) so all the dataset has been normalized before giving it to the remaining models. When analyzed by ANN among the first four models, the fourth model is performing well both in case of training and testing and from the last two models model 5 excels others. Thus in the case of ANN, the models which took majority of parameters is performing well (Tables 4 and 5). In the case of ANFIS, model 2 does well in training and testing among the first four and model 6 among the last two (Tables 4 and 5). From this it can be underpinned that diameter, length, and prefailures are important parameters responsible for failure. Table 5 gives the values of various performance indices of different models evaluated by ANFIS and ANN and MLR. In majority of models ANN bettered both ANFIS and MLR, even though in case of model 6 an exception occurred. ANFIS performs well in training

Fig. 4 Comparison of using normalized and raw data in ANN

Table 4 Performance of models analyzed by ANFIS and ANN

Dataset (D)	ANFIS		ANN	
Model	CC	MAE	CC	MAE
M1	0.89	37.43	0.91	42.16
M2	0.93	32.54	0.91	35.56
M3	0.98	18.31	0.94	23.05
M4	0.96	22.8	0.93	25.28
M5	0.99	0.05	0.99	0.11
M6	0.99	0.12	0.99	0.14

Table 5 Performance of models analyzed by ANN, ANFIS, and MLR

Dataset (D)	ANFIS		ANN		MLR	
Model	CC	MAE	CC	MAE	CC	MAE
M1	0.93	16.93	0.95	13.33	0.92	17.7
M2	0.92	13.21	0.95	15.39	0.93	14.86
M3	0.91	37.46	0.93	36.24	0.9	40
M4	0.86	49.49	0.93	33.1	0.9	42.45
M5	0.88	0.62	0.98	0.25	0.91	0.36
M6	0.95	0.17	0.95	0.28	0.94	0.88

phase for all models compared to ANN. But in testing phase its capability is not up to the mark. This may be due to over training. To prioritize the pipe network considering 400 mm diameter, the model M5 (ANN) is used since it outfits model M4 by giving good correlation (Table 6). From the first four models model 1 predicts well the failures in the testing for both ANN and ANFIS and among the last two, model 5 performs better. So these two models are taken for the comparing ANN, ANFIS and MLR (Figs. 5 and 6). From Fig. 5, it can be observed that ANN and ANFIS are predicting lower and medium number of failures in an accurate manner than higher. Compared to ANN and MLR models, ANFIS predicts higher

Table 6 Prioritization order for renewal based on the predicted number of failures by model M5 (ANN)

Link	Length (m)	Order of prioritization	Diameter (mm)
0-a	2422	1	1200
a-b	4905	2	1000
4_5	687	2	400
3_4	117	3	400
22_23	307	3	400
30_31	322	3	400
71-72	252	3	400
6_7	634	3	400

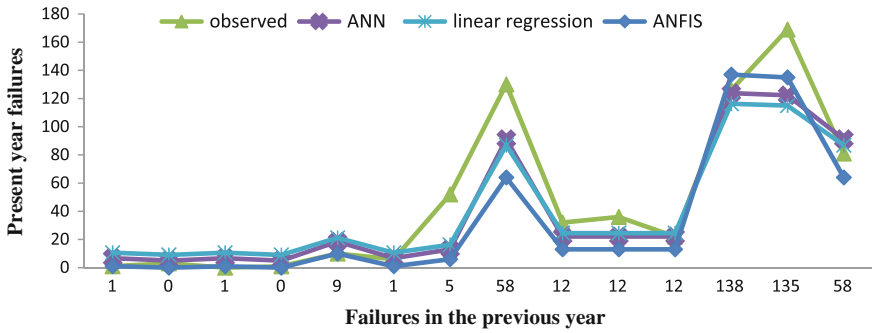

Fig. 5 Comparison between ANN, ANFIS, and MLR in predicting failures (model 1)

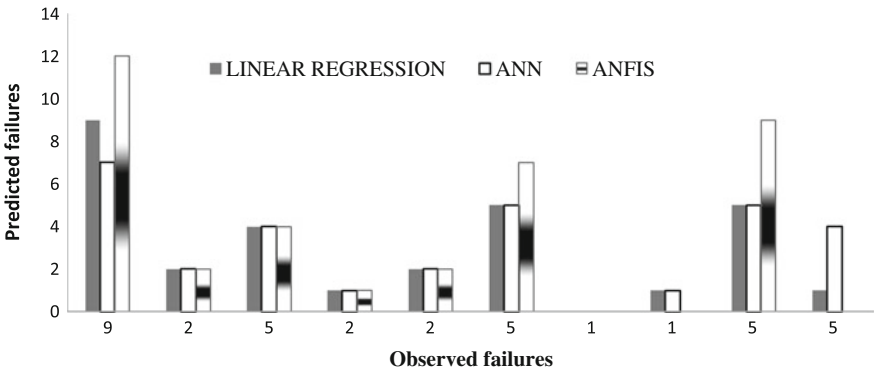

Fig. 6 Comparison between ANN, ANFIS, and MLR in predicting failures (model 5)

number of failures accurately in case of all diameter pipes. Figure 6 shows the betterment of ANN in prediction of failures than ANFIS and MLR for 400 mm and above prestressed concrete pipes.

7 Conclusion

This preliminary study looks into the capability of ANN and ANFIS for predicting failure in urban distribution network. Based on the leak detection study conducted by JBIC assisted Kerala Water Supply Project, the total physical loss of treated water is found to be 35.5 %. In this time of budget cuts and limited resources, the ability to optimize the use of maintenance money by employing predictive models in the planning stages is rapidly becoming a reality for urban underground infrastructure management. Planned maintenance facility in need of repair can yield significant savings over unscheduled or emergency repairs. The inputs to the model are selected based on relevant literature and data availability. The work can be extended with more number of inputs.

Acknowledgments The authors appreciate the enthusiasm taken by the officials of Kerala Water Authority, Kowdiar, Vellayambalam, and Aruvikkara.

References

Achim D, Ghotb F, McManus KJ (2007) Prediction of water pipe asset life using neural networks. J Infrastruct Syst 13(1):26–30

Ahmad Asnaashari IS (2007) Geostatistical analysis of water mains failure: a case study from Iran. Water Asset Manage Int 3(3):1–6

Bubtiena AM et al (2011) Application of artificial neural networks in modeling water networks. 7th international colloquium on signal processing and its applications

Ho CI, Lin MD, Lo SL (2010) Use of a GIS-based hybrid artificial neural network to prioritize the order of pipe replacement in a water distribution network. Environ Monit Assess 166:177–189

Jafar R, Shahrour I, Juran I (2010) Application of artificial neural networks to model the failure of urban water mains. Math Comput Model 51:1170–1180

Kleiner Y, Rajani B (2002) Forecasting variations and trends in water-main breaks. J Infrastruct Syst 8(4):4–122

Le Gat Y, Eisenbeis P (2000) Using maintenance records to forecast failures in water networks. Urban Water 2:173–181

Pelletier G, Mailhot A, Villeneuve JP (2003) Modeling water pipe breaks—Three case studies. J Water Res Plan Manage 129(2):2–115

Rogers PD et al (2009) Failure assessment modeling to prioritize water pipe renewal: two case studies. J Infrastruct Syst 15:162–171

Rogers PD (2011) Prioritizing water main renewals: case study of the denver water system. J Pipeline Syst Eng Pract 2:73–81

Sivanandam SN, Deepa SN (2011) Principles of soft computing. Wiley India Pvt. Ltd, Daryaganj, New Delhi-110002

Suja R, Deepthi N, Letha J (2011) Multi objective reliability based design of water distribution system. J Civil Eng (IEB) 39(1):19–31

Tabesh M et al (2009) Assessing pipe failure rate and mechanical reliability of water distribution networks using data driven modelling. J Hydroinformatics 11(1):1–17

Tian CH et al (2011) Pipe failure prediction. IBM Research, China Beijing 100193, 121–125