An Efficient Methodology for Water Supply Pipeline Risk Index Prediction for Avoiding Accidental Losses

Muhammad Shuaib Qureshi[1], Ayman Aljarbouh[2],
Muhammad Fayaz[3]
Department of Computer Science, University of Central
Asia, Naryn 722918, Kyrgyz Republic

Wali Khan Mashwani[5]
Institute of Numerical Sciences
Kohat University of Science and Technology
Kohat, Pakistan

Muhammad Bilal Qureshi[4]
Department of Computer Science
Shaheed Zulfikar Ali Bhutto Institute of Science and
Technology, Islamabad, 46000, Pakistan

Junaid Khan[6]
Department of Computer Science
International Islamic University, Islamabad
Pakistan

Abstract—The accidents happening to buildings and other human facilitation sectors due to poor water supply pipelining system is a random phenomenon, but an efficient estimation system can help to escape from such accidents. Such a system can be useful in assisting the caretakers to take the initiative measures to avoid the occurrence of the accidents or at least reduce the associated risk. In this paper, we target this issue by proposing a water supply pipelines risk estimation methodology using feed forward backpropagation neural network (FFBPNN). For validation and performance evaluation, real data of water supply pipelines collected in Seoul, Republic of South Korea from 1987 to 2010 is used. A comprehensive analysis is performed in order to get reasonable results with both original and pre-processed input data. Pre-processing consists of two steps: data normalization and statistical moments computation. Statistical moments are mean, variance, kurtosis and skewness. Significant improvement in prediction accuracy is observed with data pre-processing in terms of selected performance metrics, such as mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean squared error (RMSE).

Keywords—*Neural networks; normalization; risk index; mean square error; statistical moments*

I. INTRODUCTION

Underground facilities are very important to be monitored, because these facilities are unforeseen hazards to buildings, bridges, railway lines, etc. In terms of reducing the hazards, underground water supply pipelines are one the most important facility which contributes many underground risks and difficulties. Water is the superlative part of life without which no one can survive in the world. Water facilities are fully dependent on pipe, without pipeline water supply is impossible. In order to provide and supply water to homes, buildings and commercial areas, a bundle of pipes is installed underground. In recent age, the cities become more and more congested and these water supply pipelines are danger warning to constructed buildings as well as more risks to be taken to supply water in these congested areas [1,2].

The affection of water supply pipelines is affected by different types of parameters such as leakage, age, depth and height, quality of the pipe, water temperature, soil electrical resistivity, soil temperature, soil moisture, etc. One of the most important features for estimation of water supply pipeline risk is aging. The fitness of pipelines may be degraded with the passage of time, and also depends on the quality of materials. The degradation of the water supply pipeline can affect and destruct the underground structure of the buildings; hence it is very important to fix and restore damaged water supply pipelines on time and reduce the chances of occurrence of accidents and risks [3]. In this paper, we examine the depth, leakage, length and age which are a very important factors in water supply pipeline failure or damage.

Water supply pipeline leakage is a very remarkable factor that can cause for unexpected underground menaces such as urban sinkhole and abrupt road-side subsidence due to water pipes leaks. This water supply pipeline leakage slowly and steadily destructs the underground structure of congested buildings, sub-ways, bridges, railways, etc. and because of this alarming situation, underground water facilities permanently remain a major and serious threat. Failure and depth of water supply pipeline also ultimately damages the underground structural damage or failure eventually. When a water pipeline is spread closely to the surface then any human activity or moving of vehicles can damage the surface over the pipe and can even break the pipe when heavy load is exerted on the pipe, if a pipe under the ground is buried deeper, affection does not cause on the surface because of deepness. Suddenly rupture in the water supply pipeline can bring serious damage to the near advantage people. Comparably another parameter that is length, plays an essential role in the protection of the water supply pipeline and ultimately at underground risk [4]. To analyze water supply risks, researchers have proposed different new techniques.

The objective of this paper is to compute the accurate estimated risk failure of water supply pipelines by using FFBPNN on data with statistical moments, original data and normalized data. We selected the FFBPNN for water supply pipeline risk estimation because it is the most important model

for estimation and prediction [5]. The data with statistical moments is also very important to increase the performance of ANN [5]. Normalization also increases the performance of an algorithm. In this study the pipeline failure risk is considered in term that how it impacts the underground structure.

Rest of the paper is organized as: Section 2 presents related work, Section 3 proposed model for water supply pipelines, Section 4 explains the experimental results and discussion. The conclusion of the paper is given in Section 5.

II. RELATED WORK

Numerous algorithms have been developed based for water supply risk index assessment in literature.

Hussam et al. in [6] suggested a hierarchical based fuzzy model for water supply pipeline hazard evaluation. 16 risk factors were considered in that system. They inferred that the age of the pipe extends a solid effect on failure. They further added that pipe segments length, diameter and material are also very important elements of water supply pipelines. Yan and Vairvamoorthy in [7] proposed a decision-making technique to assess pipeline conditions. The fuzzy numbers are the output of this model that reflects the state of each pipeline. Kleiner et al. recommended a technique for buried pipelines to model the decline process; the model is based on a fuzzy rule-based non-homogeneous Markova process [8]. They also proposed a fuzzy logic method for pipeline risk evaluation. They have three main chunks, namely failure possibility, failure consequences, and a mixture of both [9].

Kleta et al., [10] used to review the system of lining surface images by video recording in which cameras are moved for damage assessment and integrity observation. The recording and camera monitoring systems are only limited to the visible parts of the surface. Different tools are designed to inspect large diameters and water mains. The most accurate tool available to detect the pockets trapped gas, leak and structural faults complicated networks of huge diameter mater mains is the Sahara Pipeline Inspection System. Meng et al., [11] recommended a quantitative risk assessment (QRA) model which is a novel approach used for evaluation of non-homogeneous road tunnels risks, because the QRA models are inappropriate to apply to assess risk in road tunnels. The tunnel segmentation principle is used in this model in which the dissimilar urban road tunnel is segmented into numerous similar segments. The separate risk for road tunnel segments along with the combined risk indication for the entire road tunnel is elucidated. Duzgun et al., [12] suggested the decision analysis method for evaluation and managing the risk of underground coal mines and falling of coal mine roof. Possible consequences and cost of consequences, the probability was used for the risk assessment. Ustinovichius et al., [13] discussed various risk assessment methods. Assessment of risk can help decision-makers for ranking existing risk to take proper reaction suitably. Fault trees, monto carlo simulation, failure mode and effective analysis, event trees, game theory, fuzzy set, grey systems and multi-criteria verbal analysis are available numerous risk assessment methods. Multilayer perceptron (MLP) is an artificial neural network (ANN) with more than one hidden layer and a bias layer. For different types of modeling of ANN, different types of architectures have been used for many years in different research areas including mathematics, engineering, medicine, neurology, meteorology, economics, hydrology, psychology and different other areas [14-17]. The ANN has many variants like multilayer perceptron (MLP), self-organization map (SOM), support vector machine (SVM), recurrent neural network (RNN) and feed forwarded neural network (FFNN). In this work, feed forward back-propagation neural network (FFBPNN) is used which is a very famous ANN model for prediction and estimation [18].

Fayaz et al. [19] proposed a model called blended hierarchical fuzzy logic for water supply risk index assessment. The purposed of the proposed model was to reduce the number of rules in the developed model. Another model named as the cohesive hierarchical fuzzy inference system was developed in [20] to assess water supply risk index. The aim of this model was similar to the previous model and both models have the potential to decrease the number of rules as well as to improve accuracy.

III. PROPOSED METHODOLOGY OF WATER SUPPLY PIPELINES RISK INDEX ESTIMATION

Fig. 1 depicts the proposed method, comprised of four different kind of layers; data acquisition layer, pre-processing layer, estimation layer based on the neural network and performance evaluation layer. Each layer has its own functionality. Data layer contains the data related to water supply pipeline risks. In the pre-processing layer, statistical moments and normalization are used to pre-process the acquired data. In the estimation layer, FFBPNN is used for water supply pipeline risk index estimation. The performance of the FFBPNN is measured in the performance evaluation layer. The three-performance measurement used for evaluation of neural network such as mean absolute error (MAE), root means absolute error (RMSE) and mean absolute percentage error (MAPE), respectively.

A. Data Layer

The datasets used in this work are real datasets that have been acquired from Electronics and Telecommunications Research Institute (ETRI) working on the underground projects. This institute completed a lot of underground project globally. For our research, we collected water supply pipelines data from 1989-2010 at different places in Seoul, the Republic of South Korea. It is observed in the literature when some tweaks are added to the original data, the performance of machine learning algorithms improves, therefore in this study, we normalize the data and calculate statistical moments of the data to get better results.

B. Pre-Processing Layer

First, we take the dataset in pre-processing layer. The datasets comprise of the leakage, age, depth and height parameters of water supply pipelines. Using this data, first we calculate the statistical moments and then concatenate with the original data. The first four parameters, namely variance, mean, kurtosis and skewness [17] can be calculated using below Equations (1-4).

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

$$\sigma = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{2}$$

$$S = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{(x_j-\mu)}{\sigma}\right]^3 \tag{3}$$

$$K = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{(x_j-\mu)}{\sigma}\right]^4 \tag{4}$$

where σ, μ, K, S represent variance, mean, kurtosis and skewness, respectively. x_i denotes values of leak, depth, length and age of pipes values, i = 1,2,…. 4.

For trial and test purposes, the normalized data can be computed by using Equation (5).

$$x_{new} \frac{x-x_{min}}{x_{max}-x_{min}} \tag{5}$$

Where the output normalized value is denoted by x_{new}, the current value is indicated by x, the minimum value in the set is represented by x_{min}, and the maximum value is denoted by x_{max} [5].

C. Water Supply Risk Index Estimation Layer

The ANN method is characterized as the regression method, which signifies the state of the art nonlinearity between the dependent and independent variables [5]. In the recent decade, researchers have deployed NNs for analyzing different kinds of estimation problems in a variety of situations. The model we used in the proposed work is the FFNN model with back-error propagation as depicted in Fig. 2 and 3, respectively for original and normalized data, and for data with statistical moments. The ANN model, bind with the error propagation algorithm (FFBPNN) is a very popular ANN model for prediction and estimation [18]. ANN has normally three layers model such as input layer, hidden layer, and output layer as depicted in Fig. 2 and 4. Researcher always use more than one hidden layer, and a bias node can also be added to the hidden layer to reduce error in the model.

If we compute the hidden layer, we can use the below Equation (6).

$$v_j = \left(1 + \exp\left(-1 \times \sum_{i=1}^{1} x_i\, w_{ij}\right)\right)^{-1} \tag{6}$$

Where j nodes in the hidden layer can be denoted by vj, node i in the input layer is denoted by x_j, w_{ij} denotes the weight between the nodes, y represents the output layer node and can be computed by (7).

$$y = \left(1 + \exp\left(-1 \times \sum_{j=1}^{j} v_j\, w_{ij}\right)\right)^{-1} \tag{7}$$

The output layer node is denoted by y, (we have only taken two output nodes in this research work, multiple numbers of output nodes can be taken according to the requirements). Error can be computed between computed data and observed data by using equation (8):

$$\text{Error} = 0.5(d-y)^2 \tag{8}$$

The observed data propagation from the output layer is represented by d, and the hidden layer can be calculated by using equations 9 and 10, respectively.

$$\delta_y = (d-y)(1-y) \tag{9}$$

$$\delta_y = v_j(d-v_j)(1-y)\delta_y w_{j1}, j = 1, \ldots\ldots, J \tag{10}$$

The input and hidden layers and the adjustment of weight w between hidden layers and output layers can be computed by the following formulae (11, 12), respectively.

$$\Delta w_{ij} = \alpha\delta_y v_j, i = 1, \ldots\ldots, I; \ j = 1, \ldots., J \tag{11}$$

$$\Delta w_{ij}{}^n = \alpha\delta_y v_j\, j = 1, \ldots\ldots, J \tag{12}$$

The learning rate is represented by, and also momentum can be computed as (13-14);

$$\Delta w_{ij}{}^n = \alpha\delta_y v_{j,} + \beta\,\Delta w_{j1}{}^{n-1}, j = 1, \ldots, J \tag{13}$$

$$\Delta w_{ij}{}^n = \alpha\delta_y v_j + \beta\,\Delta w_{j1}{}^{n-1}, i = 1, \ldots, I; \ j = 1, \ldots., J \tag{14}$$

The iteration of error backpropagation is indicated by n, and momentum constant is represented by β. This momentum method accelerates the weights to avoid any fluctuations in the training process of error surface in the flat region.

For the validation of each of the models developed for a different number of inputs and hidden neurons, the percentage split method is applied in which the total data are separated, and experimentation is done by 70/30 random training-test splits, 70% for training the data and 30% for testing to validate the samples. This ratio is the standard ratio for splitting training and testing data [15].

D. Performance Evaluation Layer

Different parameters are available to calculate the performance of the model. Below three performance measurement equations have been used to measure the performance that is mean absolute error (MAE), root mean square error (RMSE) and the mean absolute percentage error (MAPE) [15]. For the assessment of regression accuracy, these performance measurement matrices are normally used in the literature. Numerical equations of MAE, RMSE, and MAPE are calculated by using below Equations (15, 16, 17), alternatively.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{k=0}^{n}(A-E)^2} \tag{15}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{n}|A_i - E_i| \tag{16}$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{n}\frac{|A_i-E_i|}{A_i} \times 100 \tag{17}$$

Where the total number of observations is represented by N, actual values are denoted by A, and estimated values is represented by P. As illustrated in Table I, we assessed the performance of the FFBPNN by using MAE, RMSE, and MAPE on different types of data, the values get from MAE, RMSE and MAPE show that the FFBPNN outperforms on data with statistical moments and normalized data.
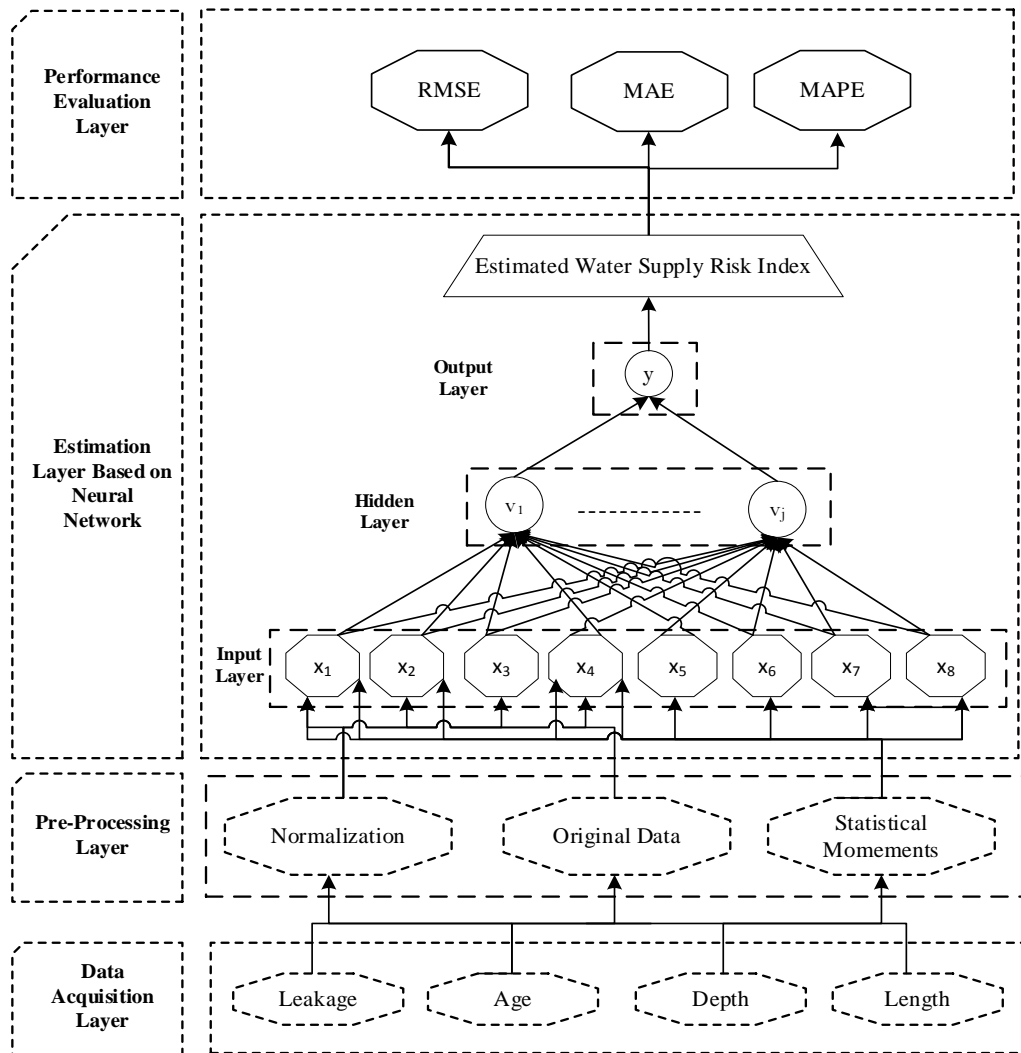
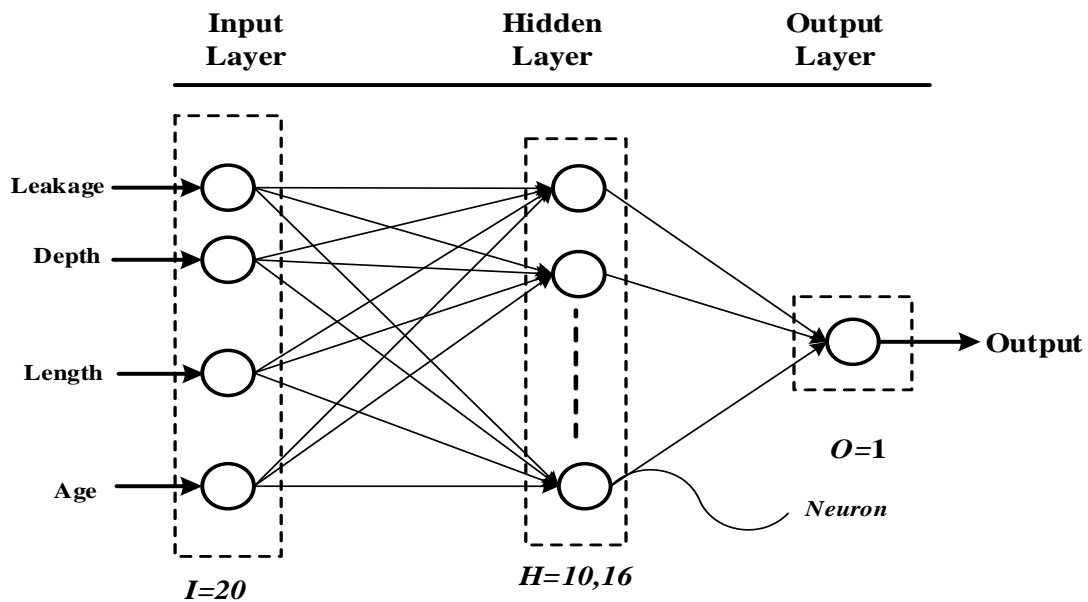Fig. 1. Proposed Water Supply Pipeline Risk Index Estimation Methodology.

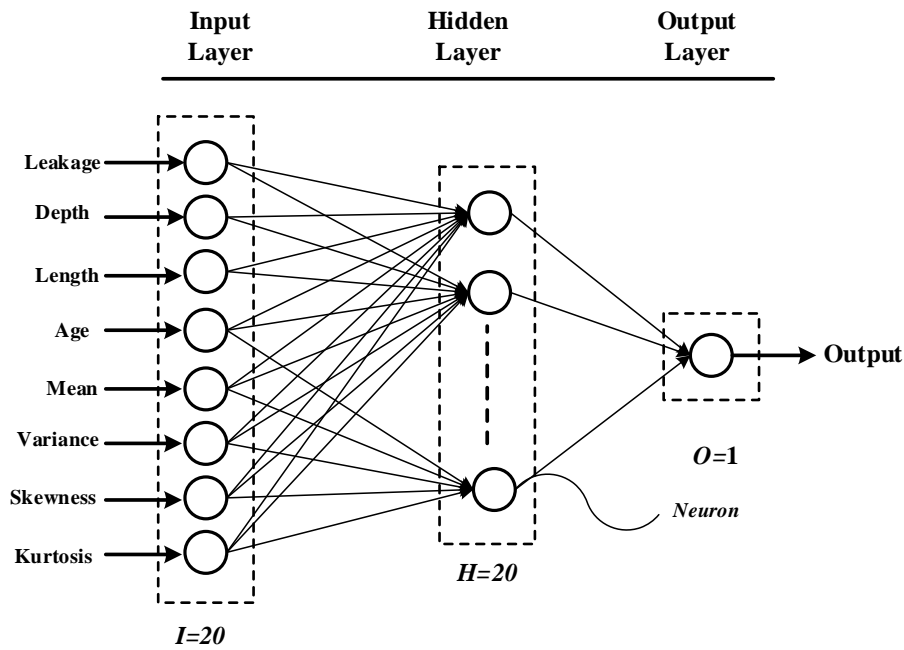Fig. 2. Structure of Model M1 for Four Inputs.

Input Layer **Hidden Layer** **Output Layer**

Fig. 3. Structure of Model M2 for Eight Inputs.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Setup

The proposed scheme experimentation is performed using windows 10 operating system with an Intel Core i5 processor using MATLAB R2019b version 9.7.0.1216025. In this study different types of experiments have been performed on the data in order to calculate the best estimated risk index for water supply pipelines. Typically input parameters play the most important role in the performance of any kind of machine learning algorithm technique. Therefore, in this research, first the leakage, depth, length and age values of water supply pipelines with the error correction neural network have been given as inputs to the feed forward backpropagation. The model we have tried in this research composed of input layer with different combination of neurons in the hidden layer and output layer. The number of neurons which is best suited to the proposed method is the combination of ten (10) neurons in hidden layer with four (4) neurons in input layer, and single neuron in output layer have been selected as shown in Fig. 4.

Secondly, Different sets of experiments are accomplished with normalized data. Four neurons in the input layer, sixteen (16) neurons in the hidden layer and single neuron in the output layer have been applied as shown in Fig. 5. In the same case, we also tried different combinations of maximum and minimum number of neurons in the hidden layer. We found that this combination (10 neurons in hidden layer) is best fit in combination with four inputs and one output layer.

Third, the experiment is performed by combining original data with statistical moments as shown in Fig. 6. Eight (8) neurons in the input layer and Twenty (20) neurons are configured in the hidden layer with one neuron in the output layer. Age, depth, height, leakage, variance, mean, kurtosis and skewness are inputs to the neural networks. In order to find the better combination of the number of neurons in the hidden layer with the input and output layer, we tested different number of neurons in the hidden layer and we concluded that this combination is more accurate combination as shown in Fig. 6, thus selected.

B. Results and Discussion

The graphical representations of the estimated results are presented in the following section. The actual risk, estimated risk and the errors observed in estimation for water supply pipelines using originally collected data are shown in Fig. 7 and 8. The estimated risk, actual risk and the errors observed in estimated risk and actual risk using FFBPNN on normalized data are shown in Fig. 9 and 10. The actual risk, estimated risk and the errors observed in estimation using FFBPNN for collecting water supply pipeline data with statistical moments are shown in Fig. 11 and 12.

Performance using three measurement; mean absolute percentage error (MAPE), mean absolute error (MAE) and root mean square error (RMSE) is calculated for FFBPNN on the normalized data (ND), original data (SD) and data with statistical moments (SMD) is shown in Table I and Fig. 13. The outcomes show that FFBPNN outperforms on both normalized data (ND) and data with statistical moments (SMD) as compared to original data that is provided to FFBPNN.
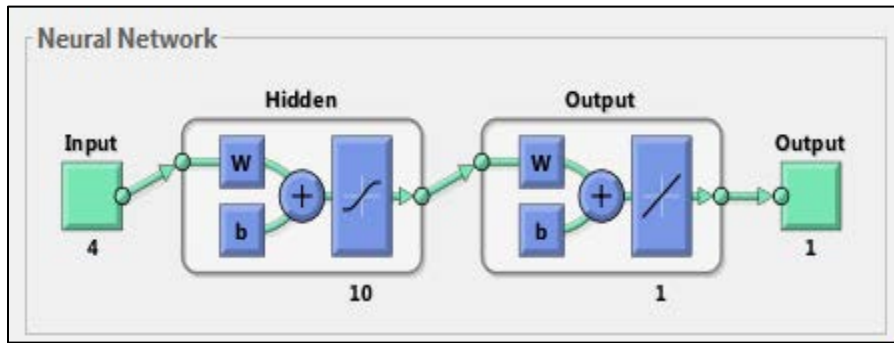
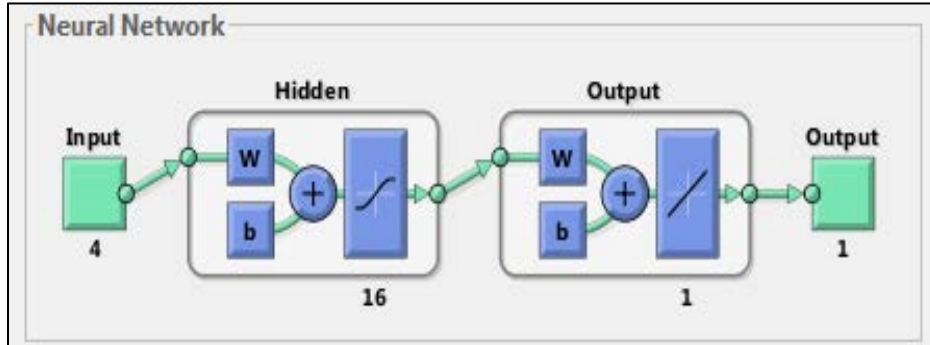Fig. 4.   Deployment of ANN Configuration on Original Data.

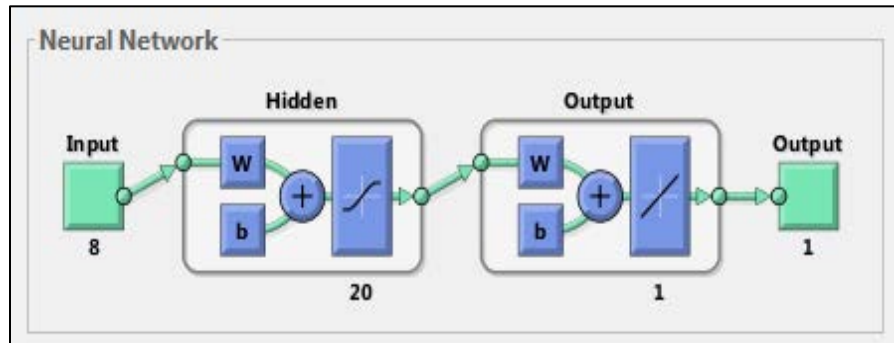Fig. 5.   Deployed ANN Configuration on Normalized Data.

Fig. 6.   ANN Configuration Applied on Data with Statistical Moments.
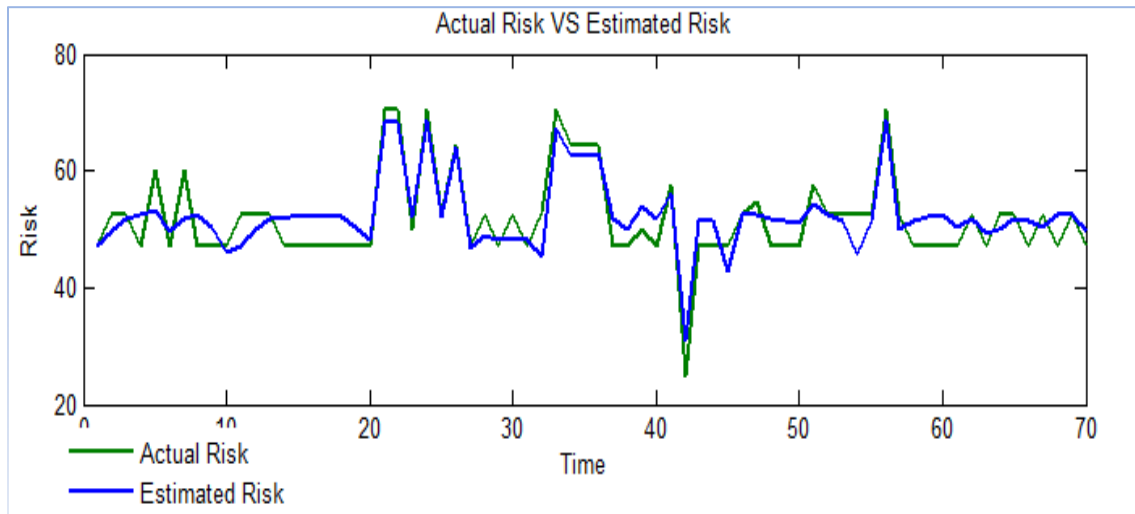

Fig. 7.   Actual Risk and Estimated Risk Values of Water Supply Pipelines using FFBPNN on Original Data.
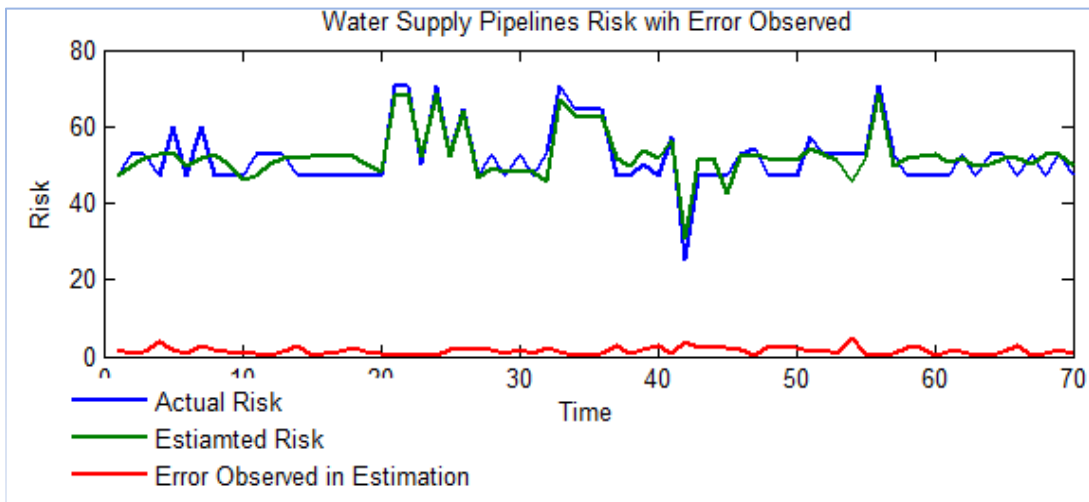
Fig. 8.   Actual and Estimated Risk Values of Water Supply Pipelines using FFBPNN on Original Data with Errors Observed.
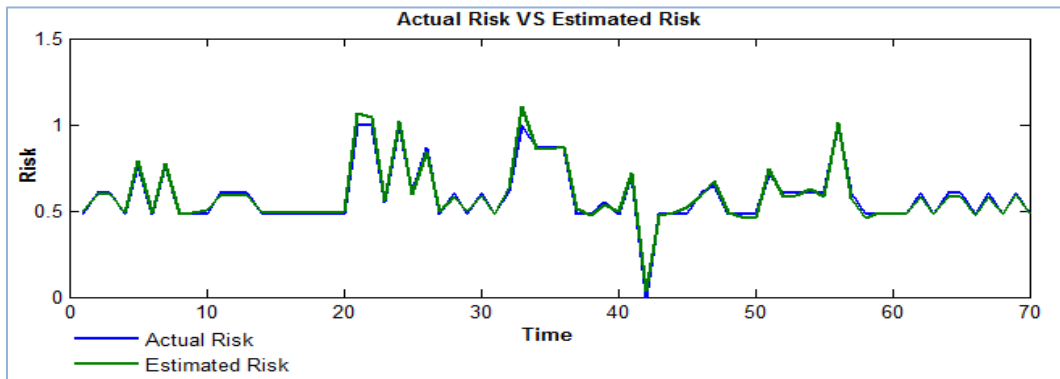

Fig. 9.   Actual Risk and Estimated Water Supply Pipelines Risk Values using Neural Network (NN) on Normalized Data.
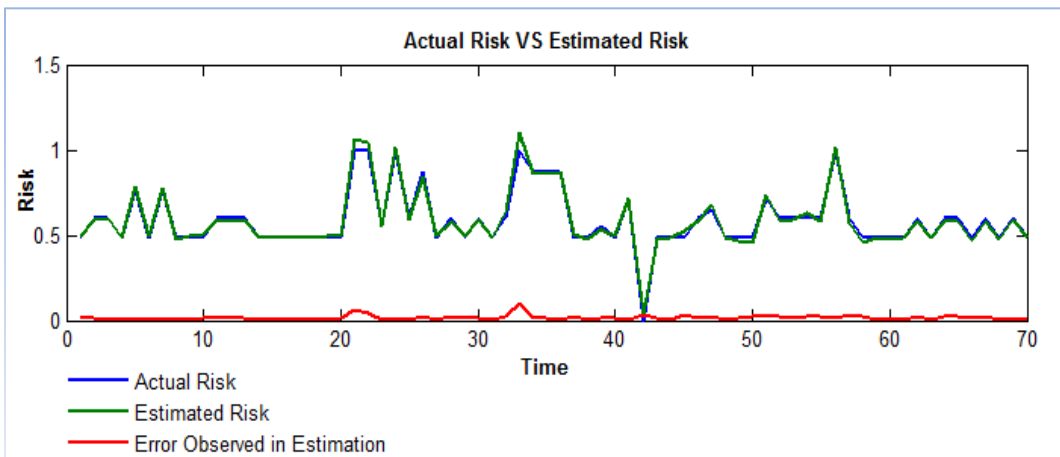

Fig. 10. Estimated Water Supply Pipeline Risk using Neural Network (NN) on Normalized Data with Errors Observed.
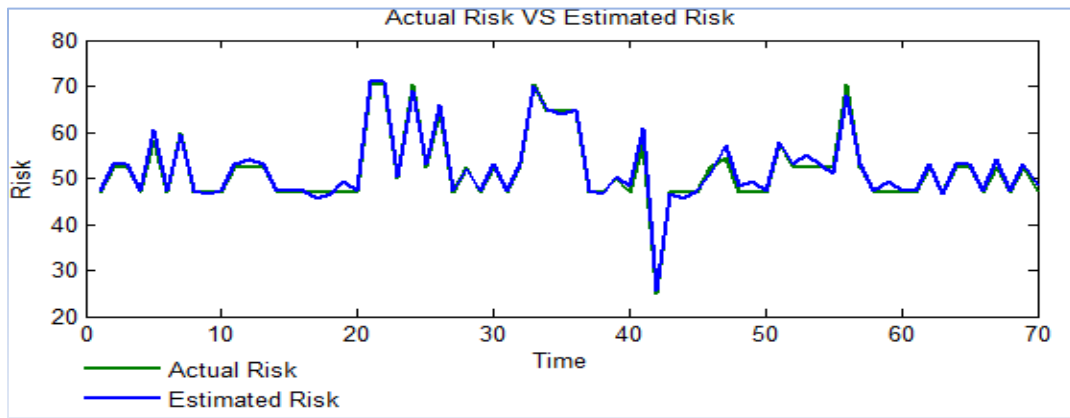
Fig. 11. Actual and Estimated Risk Values of Water Supply Pipeline using Neural Network (NN) on Data with Statistical Moments.
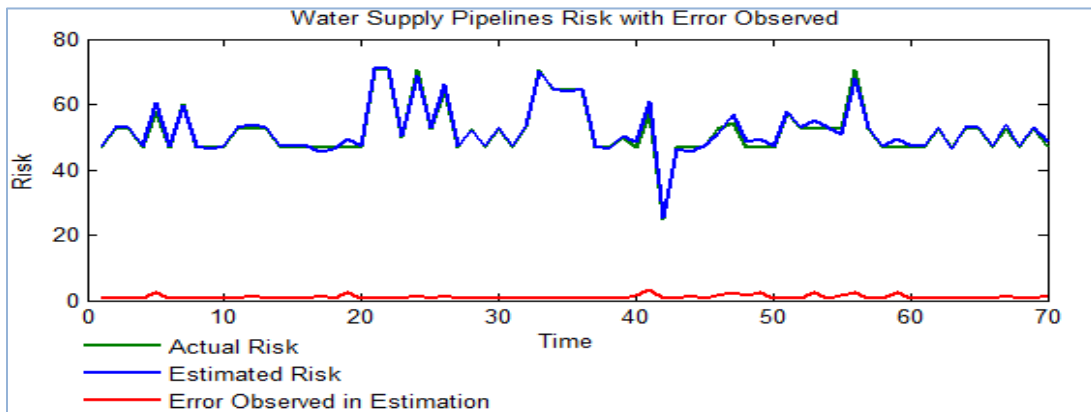

Fig. 12. Actual, Estimated Risk Values of Water Supply Pipelines using Neural Network (NN) on Data with Statistical Moments with Errors Observed.

TABLE I. PERFORMANCE OF FFBPNN ON ORIGINAL AND PREPROCESSED DATA

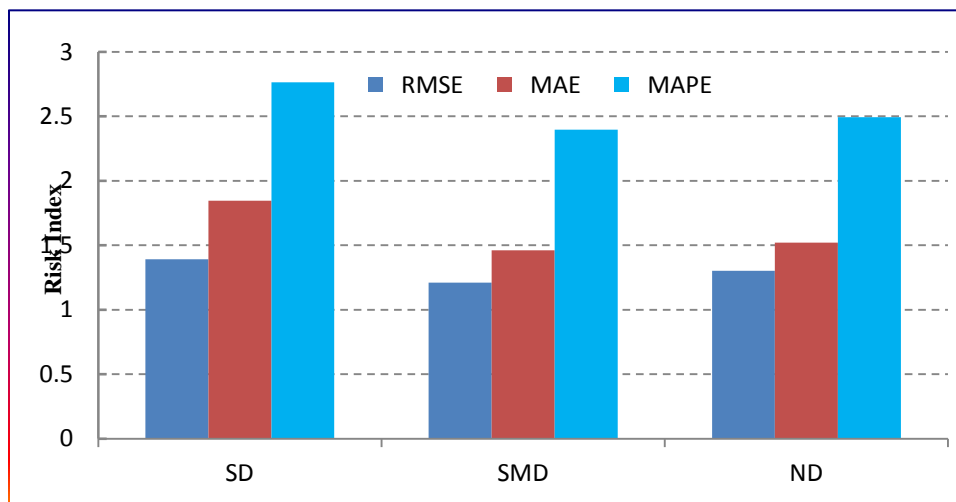|	Original Data (SD)	Data with Statistical Moments (SMD)	Normalized Data (ND)
MAPE	2.763	2.3965	2.4948
MAE	1.846	1.4619	1.5219
RMSE	1.3907	1.2106	1.3016

Fig. 13. Graphical Comparison of Estimation Performance of FFBPNN on different Types of Data.

V. CONCLUSION

In this paper, a multi-layer perceptron is applied to predict water supply pipeline risk. The multi-layer perceptron was randomly trained tested using historical data. The collected data are from 1989-2010 about water supply pipelines fitted in Seoul, Republic of Korea. Experimentation is done by 70/30 random training-test splits, where 70% for training and 30% for testing to validate the samples. For the performance and accuracy evaluation of the models, the root means square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) were used. The FFBPNN was applied to original collected data and we noticed that the outcomes attained for the original data were not prominent. Then we added a tweak by calculating the statistical moments of the original data and merged with the original data. After this process, we applied FFBPNN on this new data. The performance measure matrices indicate that the results provided by FFBPNN using the new data with statistical moments are comparatively better. Further, we normalized the original data and applied FFBPNN using this normalized data with statistical moments where noticeable outperformance is achieved. Overall the performance of FFBPNN on statistical moment data is slightly prominent as compared to normalized data and far better than original data. In future, we may apply some more tweaks on the data and may test more machine learning algorithms.

REFERENCES

[1] Kleiner, Y.; Rajani, B. In Considering time-dependent factors in the statistical prediction of water main breaks, AWWA Infrastructure Conference Proceedings, 2000; pp 12-15.

[2] Fayaz, M.; Pham, Q.B.; Linh, N.T.T.; Nhi, P.T.T.; Khoi, D.N.; Qureshi, M.S.; Shah, A.S.; Khalid, S. A Water Supply Pipeline Risk Analysis Methodology Based on DIY and Hierarchical Fuzzy Inference. Symmetry 2020, 12, 44.

[3] Tang, X.; Liu, Y.; Zheng, L.; Ma, C.; Wang, H. In Leak detection of water pipeline using wavelet transform method, Environmental Science and Information Application Technology, 2009. ESIAT 2009. International Conference on, 2009; IEEE: pp 217-220.

[4] Caleyo, F.; Velázquez, J.; Valor, A.; Hallen, J. Probability distribution of pitting corrosion depth and rate in underground pipelines: A monte carlo study. Corrosion Science 2009, 51, 1925-1934.

[5] Fayaz, M.; Shah, H.; Aseere, A.M.; Mashwani, W.K.; Shah, A.S. A Framework for Prediction of Household Energy Consumption Using Feed Forward Back Propagation Neural Network. Technologies 2019, 7, 30.

[6] Fares, H.; Zayed, T. Hierarchical fuzzy expert system for risk of failure of water mains. Journal of Pipeline Systems Engineering and Practice 1, 53-62.

[7] Yan, J.; Vairavamoorthy, K. Fuzzy approach for pipe condition assessment. In New pipeline technologies, security, and safety, 2003; pp 466-476.

[8] Kleiner, Y.; Sadiq, R.; Rajani, B. Modeling failure risk in buried pipes using fuzzy markov deterioration process. In Pipeline engineering and construction: What's on the horizon? 2004; pp 1-12.

[9] Kleiner, Y.; Rajani, B.; Sadiq, R. Failure risk management of buried infrastructure using fuzzy-based techniques. Journal of Water Supply: Research and Technology-AQUA 2006, 55, 81-94.

[10] Kleta, H.; Heyduk, A. Image processing and analysis as a diagnostic tool for an underground infrastructure technical condition monitoring. Underground Infrastructure of Urban Areas 3, 39.

[11] Meng, Q.; Qu, X.; Wang, X.; Yuanita, V.; Wong, S.C. Quantitative risk assessment modeling for nonhomogeneous urban road tunnels. Risk Analysis 31, 382-403.

[12] Duzgun, H. Analysis of roof fall hazards and risk assessment for zonguldak coal basin underground mines. International Journal of Coal Geology 2005, 64, 104-115.

[13] Ustinovichius, L.; Barvidas, A.; Vishnevskaja, A.; Ashikhmin, I.V. Multicriteria verbal analysis for the decision of construction problems. Technological and Economic Development of Economy 2009, 15, 326-340.

[14] Kumar, M.; Raghuwanshi, N.; Singh, R.; Wallender, W.; Pruitt, W. Estimating evapotranspiration using artificial neural network. Journal of Irrigation and Drainage Engineering 2002, 128, 224-233.

[15] Wahid, F.; Kim, D.H. Short-term energy consumption prediction in korean residential buildings using optimized multi-layer perceptron. Kuwait Journal of Science 44.

[16] Chow, T.; Zhang, G.; Lin, Z.; Song, C. Global optimization of absorption chiller system by genetic algorithm and neural network. Energy and buildings 2002, 34, 103-109.

[17] Firat, F.; Arslan, A.K.; Colak, C.; Harputluoglu, H. Estimation of risk factors associated with colorectal cancer: An application of knowledge discovery in databases. Kuwait Journal of Science 43.

[18] Gibbs, M.S.; Morgan, N.; Maier, H.R.; Dandy, G.C.; Nixon, J.B.; Holmes, M. Investigation into the relationship between chlorine decay and water distribution parameters using data driven methods. Mathematical and Computer Modelling 2006, 44, 485-498.

[19] Fayaz, M., Ahmad, S., Ullah, I., & Kim, D. A blended risk index modeling and visualization based on hierarchical fuzzy logic for water supply pipelines assessment and management. Processes, 2018, 6, 61.

[20] Fayaz, M., Ahmad, S., Hang, L., & Kim, D. Water supply pipeline risk index assessment based on cohesive hierarchical fuzzy inference system. Processes, 2019, 7, 182.