14th International Conference on Structural Safety and Reliability - ICOSSAR'25
1-6 June, 2025 - Los Angeles California, USA
R. Ghanem, E. Johnson, S. Masri and A. Olivier (Eds)

RELIABILITY ASSESSMENT OF INDIVIDUAL WATER SUPPLY PIPELINES USING MACHINE LEARNING TECHNIQUES

CHENGCAI ZONG[1], CHAOLIE NING[2] AND JIE LI[3]

[1] College of Civil Engineering, Tongji University
1239 Siping Road, Shanghai 200092, PR China
E-mail address: zongcc@tongji.edu.cn

[2] College of Civil Engineering, Tongji University
1239 Siping Road, Shanghai 200092, PR China
E-mail address: clning@tongji.edu.cn

[3] College of Civil Engineering, Tongji University
1239 Siping Road, Shanghai 200092, PR China
E-mail address: lijie@tongji.edu.cn

Key words: Water Supply Pipelines, Reliability Assessment, Failure Count, Failure Probability, Homogeneous Grouping, Machine Learning.

Abstract. *To achieve a quantitative assessment of the reliability of individual water supply pipelines, this study proposes a machine learning-based analytical framework. First, individual pipeline samples are constructed based on pipeline attributes and historical failure records. Each sample comprises feature variables that influence failures, along with a corresponding failure count label. Subsequently, the samples are aggregated into homogeneous groups based on feature similarity, and predictive models-including Random Forest (RF), Artificial Neural Networks (ANN), and Support Vector Regression (SVR)-are developed to estimate the failure counts for each group. Building on this, the failure process of the pipelines is modeled as a non-homogeneous Poisson process. By incorporating the individual pipeline length and the average failure rate of the corresponding homogeneous group, the failure probability is estimated, thereby enabling a quantitative assessment of the reliability of individual pipelines. A case study on gray cast iron water supply pipelines in certain administrative regions of Shanghai is conducted to validate the effectiveness and applicability of the proposed framework. The results demonstrate that all three models can effectively predict the failure counts of homogeneous groups, and that these predictions can be used to accurately estimate the failure probabilities of individual pipelines. These findings highlight the promising potential of the proposed method for risk mitigation and asset management in the water industry.*

1 INTRODUCTION

Pipeline networks form the backbone of urban water supply systems, accounting for nearly 80% of total system investment[1]. As urbanization progresses and pipeline infrastructure expands, long-term environmental corrosion and material aging have led to an increasing number of failures, such as leaks and bursts[2,3]. These failures not only result in substantial water losses but also pose serious risks to public safety and property[4]. In this context, the assessment of pipeline reliability has become an essential aspect of daily operations, enabling more effective risk control and asset management for water utilities[5,6].

Reliability assessment models for water supply pipelines are generally classified into two categories: physics-driven models and data-driven models[7,8]. Physics-driven models primarily rely on mechanical analysis of external loads and the deterioration of structural resistance over time. Methods such as Monte Carlo simulation[9,10] and fuzzy analysis[11] are often employed to incorporate and quantify various sources of uncertainty. However, due to the limited understanding of pipeline deterioration and failure mechanisms[12], as well as the fact that water supply pipelines are often buried in geologically complex environments, it is challenging to accurately obtain and quantify key physical parameters[1]. These difficulties constrain the accuracy and applicability of physics-driven models in practical applications.

Data-driven models estimate the failure risk of water supply pipelines by analyzing historical failure records and establishing a mapping between input features and target variables, thereby enabling a quantitative assessment of pipeline reliability. In recent years, the growing efforts by governmental agencies and utility companies to collect and manage data throughout pipeline service life have led to an increasingly rich reservoir of failure records. At the same time, advances in machine learning techniques have provided powerful tools for modeling the complex behavior of infrastructure systems, further promoting the application of data-driven models in pipeline failure prediction. Methods such as Artificial Neural Networks (ANN)[5,13,14], Support Vector Machines (SVM)[15], Random Forests (RF)[16,17], Logistic Regression (LR)[18], and Bayesian Belief Networks (BBN)[6,19] have been widely applied to predict key indicators such as failure rates, remaining useful life, and failure counts of water supply pipelines. Compared to physics-driven models, data-driven models do not rely on specific physical parameters, offer greater flexibility in selecting input features, and can function effectively even when the underlying deterioration mechanisms are not fully understood. As a result, they provide a more economical and practical solution for rapid failure prediction and intelligent management in large-scale water supply networks.

This study proposes an analytical framework for assessing the reliability of individual water supply pipelines using machine learning techniques. Initially, a single sample is constructed for each pipeline using the available attribute data and historical failure records. Each sample comprises a set of feature variables related to pipeline failure and a label representing the corresponding failure count. To enhance statistical significance, the samples are aggregated into homogeneous groups based on feature similarity, enabling more effective analysis. Leveraging the strengths of machine learning in modeling nonlinear relationships, predictive models based on RF, ANN, and SVR are developed to estimate the total failure counts within each homogeneous group. Furthermore, the failure process of the pipelines is modeled as a non-homogeneous Poisson process. By incorporating the length of individual pipelines and the average failure rate of their respective homogeneous groups, the failure probability is estimated, providing a quantitative assessment of the pipeline's reliability. Finally, a case study on gray cast iron water supply pipelines in certain administrative regions of Shanghai is conducted to validate the effectiveness and applicability of the proposed framework.

2 METHODOLOGY

The flowchart of the proposed framework is shown in Figure 1. Detailed descriptions of the data collection and processing procedures, as well as the pipeline sample grouping strategy, are provided in Sections 3.1 and 3.2. This section introduces the three machine learning regression algorithms employed in this study, the method for estimating the failure probability of individual pipelines, and the evaluation metrics used to assess the performance of both the prediction of failure counts in homogeneous groups and the estimation of failure probabilities for individual pipelines.

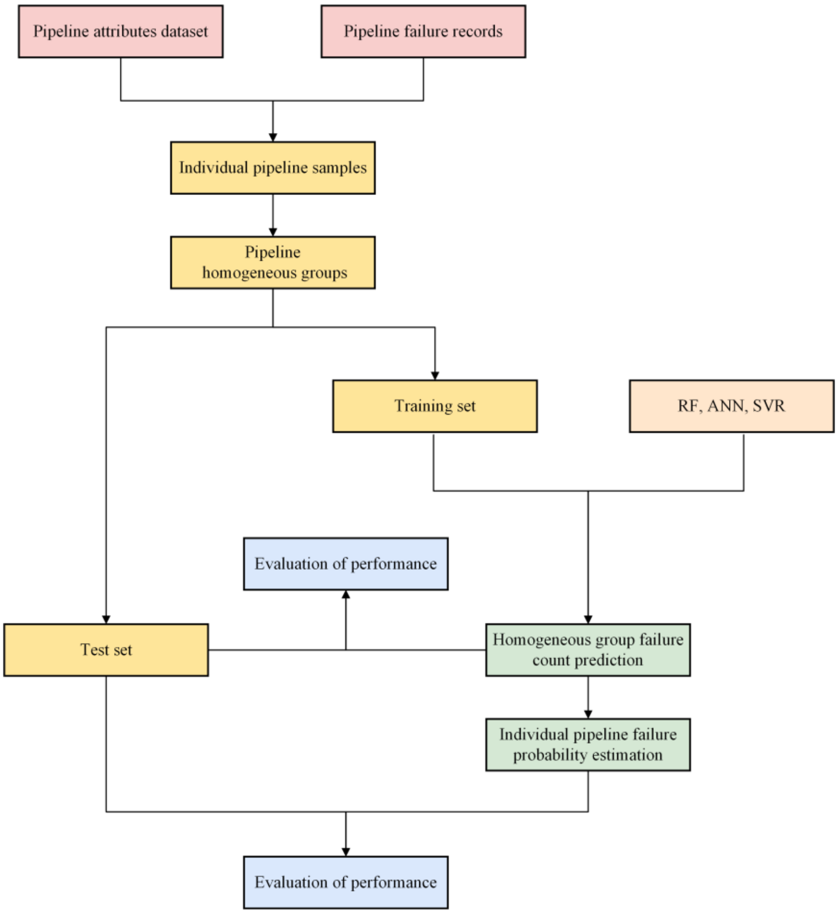

Figure 1: Flowchart of the proposed framework

2.1 Overview of machine learning methods

RF: RF is an ensemble learning method proposed by Breiman[20], which uses decision trees as base learners. It combines the Bagging strategy with random feature selection to increase the diversity among base learners and reduce overfitting, thereby improving the model's robustness and generalization ability. In this study, binary regression trees are constructed using the CART algorithm[21] as the base learners of the RF. The final regression output is obtained by averaging the predictions of all individual trees.

ANN: ANN simulates the structure of biological neural networks in the human brain and consists of interconnected artificial neurons capable of transmitting information[22,23]. Typically, the network comprises an input layer, one or more hidden layers, and an output layer. Each layer consists of multiple neurons, and the neurons between adjacent layers are connected by weights that enable information propagation. Given an input feature vector $\boldsymbol{a}^{(0)}$, a neural network with S layers (including all hidden and output layers) propagates information through successive iterations according to Equation (1):

$$\boldsymbol{a}^{(s)} = f_s\left(\boldsymbol{W}^{(s)}\boldsymbol{a}^{(s-1)} + \boldsymbol{b}^{(s)}\right) \tag{1}$$

Where $\boldsymbol{a}^{(s)} \in \mathbb{R}^{M_s}$ denotes the output of the neurons in the s-th layer; M_s is the number of neurons in the s-th layer; $\boldsymbol{W}^{(s)} \in \mathbb{R}^{M_s \times M_{s-1}}$ represents the weight matrix between the (s−1)-th and s-th layers; $\boldsymbol{b}^{(s)} \in \mathbb{R}^{M_s}$ denotes the bias vector from the (s−1)-th layer to the s-th layer; and $f_s(\cdot)$ is the activation function applied to the neurons in the s-th layer. In this study, the ReLU function is used as the activation function in the hidden layers, while the identity function is employed in the output layer.

During the training process of the ANN model, the weights and biases are optimized by minimizing the error between the predicted and actual values. Additionally, regularization terms such as the L_2 norm can be introduced to constrain model complexity, thereby effectively reducing the risk of overfitting.

SVR: SVR is an extension of the Support Vector Machine (SVM) proposed by Cortes and Vapnik[24], designed to address regression problems. The core idea is to find an optimal regression hyperplane that can estimate the target variable as accurately as possible. The optimization objective of SVR is defined as follows:

$$\min_{\boldsymbol{w},b,\xi_i,\hat{\xi}_i} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \hat{\xi}_i\right) \tag{2}$$

Subject to:

$$\begin{cases} y_i - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i - b \le \varepsilon + \xi_i, \\ \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b - y_i \le \varepsilon + \hat{\xi}_i, \\ \xi_i \ge 0, \hat{\xi}_i \ge 0, \quad i = 1, 2, \cdots, n \end{cases} \tag{3}$$

Where $\boldsymbol{w}$ denotes the weight vector; b is the bias term; n represents the number of samples; C is the regularization parameter, which balances model complexity and error tolerance; ξ_i and ξ_i are slack variables that allow sample i to deviate within a certain error margin; and ε is a hyperparameter that controls the model's tolerance to prediction errors.

SVR incorporates kernel functions[25] to construct nonlinear regression models, allowing it to accommodate more complex data distributions. In this study, the Radial Basis Function (RBF) kernel is employed.

2.2 Assessment of failure probability for individual pipelines

If the total failure count FC of a pipeline homogeneous group has been modeled as a function of the group's feature vector $\boldsymbol{X}$ using machine learning regression algorithms, it can be written as:

$$FC = g(\boldsymbol{X}) \tag{4}$$

Suppose that the failure events of pipeline P_{ij} , which is the *j*-th pipeline in the *i*-th homogeneous group G_i , follow a Poisson distribution along its length. Then, its failure probability FP_{ij} can be estimated as:

$$FP_{ij} = 1 - e^{-\lambda_{G_i} l_{ij}} \tag{5}$$

Where l_{ij} is the length (km) of pipeline P_{ij} , and λ_{G_i} represents the average failure rate (failures/km) within the homogeneous group G_i :

$$\lambda_{G_i} = \frac{FC_{G_i}}{\sum_{P_{ij} \in G_i} l_{ij}} \tag{6}$$

where FC_{G_i} is the total failure count for all samples within homogeneous group G_i .

2.3 Evaluation of performance

Four performance evaluation metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2), as defined in Equations (7)-(10), are used to assess the performance of the three algorithms in predicting the total failure counts of homogeneous groups.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(FC_{G_i} - FC_{G_i} \right)^2 \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(FC_{G_i} - FC_{G_i} \right)^2} \tag{8}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| FC_{G_i} - FC_{G_i} \right| \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(FC_{G_i} - FC_{G_i} \right)^2}{\sum_{i=1}^{n} \left(FC_{G_i} - \frac{1}{n} \sum_{i=1}^{n} FC_{G_i} \right)^2} \tag{10}$$

Where n is the number of homogeneous groups, FC_{G_i} is the observed total failure count for group G_i , and FC_{G_i} is the corresponding predicted value.

Three performance evaluation metrics, including Brier Score (BS), Cross-Entropy Loss (L_{CE}), and Ranking Quality (RQ), as defined in Equations (11)-(13), are used to assess the performance of individual pipeline failure probability estimation. Among them, BS and L_{CE} primarily focus on the absolute accuracy of the predicted probabilities, while RQ emphasizes the quality of the relative ranking of the predictions.

$$BS = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - FP_i \right)^2 \tag{11}$$

$$L_{\mathrm{CE}} = -\frac{1}{n}\sum_{i=1}^{n}\left[Y_i \cdot \log\left(FP_i\right) + \left(1-Y_i\right)\cdot \log\left(1-FP_i\right)\right] \tag{12}$$

$$RQ = \frac{\sum_{i=1}^{n}\sum_{j=i+1}^{n}\xi\left(i,j\right)}{CN} \tag{13}$$

$$\xi\left(i,j\right) = \begin{cases} 1, \left(FP_i - FP_j\right)\cdot\left(FC_i - FC_j\right) > 0 \\ 0.5, \left(FP_i - FP_j\right)\cdot\left(FC_i - FC_j\right) = 0 \\ 0, else \end{cases} \tag{14}$$

Where n represents the total number of pipeline samples; Y_i denotes the actual failure status of the *i*-th pipeline sample (0 for non-failure and 1 for failure); FP_i is the estimated failure probability of the *i*-th sample; FC_i is the actual failure count of the *i*-th sample; and CN indicates the number of sample pairs where $FC_i \neq FC_j$. When $FC_i = FC_j$, the effectiveness of the failure probability estimation cannot be determined; therefore, such cases are excluded from Equations (13) and (14).

3 CASE STUDY

This section presents a case study on gray cast iron water supply pipelines in certain administrative regions of Shanghai, China, to verify the applicability and effectiveness of the proposed framework.

3.1 Data collection and preprocessing

The total length of gray cast iron pipelines in the selected study area of Shanghai is approximately 340 km. Available data include attribute information such as installation date, location, diameter, and length of the pipelines, as well as failure records (including failure date and location) from 2004 to 2018. The raw data were preprocessed to generate individual pipeline samples following the steps below:

(1): Pipelines with erroneous or missing attribute information, as well as those with abnormal failure records, were excluded to ensure data completeness and accuracy.

(2): Only failure events occurring in the pipeline body or its joints were retained, while those associated with other components of the water supply network—such as hydrants and valves—were excluded.

(3): The lifecycle of each pipeline was divided into several one-year intervals. To reduce correlation between samples, one interval was randomly selected for each pipeline during the failure observation period. Intervals containing failure events were prioritized to make full use of available failure information. For the selected interval of each pipeline, relevant features were extracted-including failure history over the past five years (FH), diameter (D), age (A), and length (L)-along with the corresponding failure count, forming a single sample.

As a result of this preprocessing, a total of 1583 pipeline samples and 631 corresponding failure records were obtained.

3.2 Homogeneous grouping of pipeline samples

To enhance statistical significance and enable effective analysis, pipeline samples need to be aggregated into homogeneous groups. Prior to the grouping process, a visual analysis of all sample features was conducted to examine their distribution patterns, identify potential outliers, and determine whether preprocessing steps such as binning were needed to reduce noise and improve grouping quality. The distribution of each feature is presented in Figure 2.

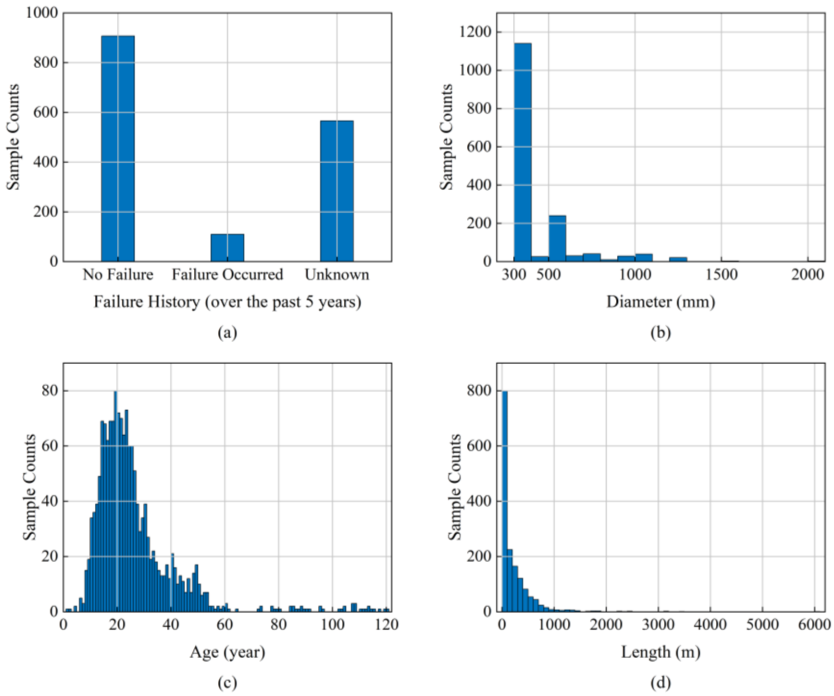

Figure 2: Statistical distribution of features for pipeline samples

The distribution of failure history over the past five years is shown in Figure 2(a). The results indicate that most samples experienced no failures over the past five years, with only a small number having recorded failures. Due to the left-truncated nature of failure data[26]-where failure events prior to the failure observation period are unobservable for pipelines installed earlier-the failure history over the past five years of some samples is incomplete. For these samples, the failure history is labeled as “Unknown,” representing an intermediate state between “No Failure” and “Failure Occurred.”

The distribution of diameter is shown in Figure 2(b). The results show that pipelines with a diameter of 300 mm constitute the vast majority. Given that most samples have diameters in multiples of 100 mm, and to mitigate the impact of outliers on model performance, the diameters of 10 samples with non-standard sizes (e.g., 350 mm, 625 mm, 750 mm, 950 mm) were rounded down. For instance, a diameter of 625 mm was recorded as 600 mm.

The distribution of age is shown in Figure 2(c). It can be observed that the vast majority of samples are less than 60 years old, with relatively few exceeding this threshold. To handle older samples more appropriately and reduce noise, samples with pipeline age over 60 years were divided into two intervals: (60, 90] and (90, 120], and were assigned representative values of

75 and 105 years, respectively.

Figure 2(d) shows the distribution of individual pipeline lengths. Most samples are shorter than 100m, although a few extremely long samples exist, with the maximum reaching nearly 6100m.

The pipeline samples were aggregated into 377 homogeneous groups based on three features: failure history over the past five years, diameter, and age. All pipelines within each group share identical values for these three features. The distributions of group-level features are shown in Figure 3.

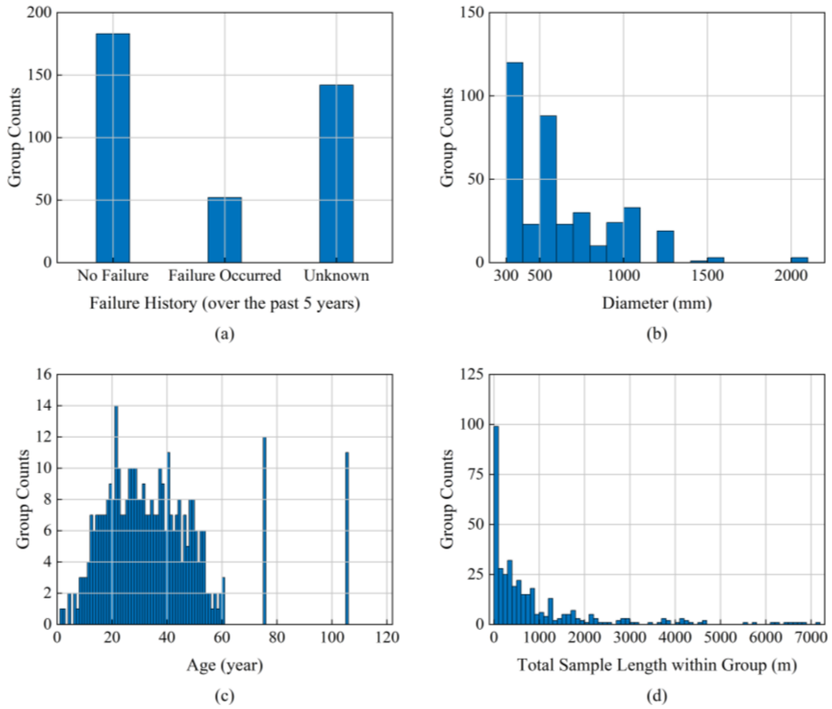

Figure 3: Statistical distribution of features for homogeneous groups

All homogeneous groups were randomly divided into a training set (comprising 80% of the total groups, including 1,280 pipeline samples) and a testing set (comprising the remaining 20%, including 303 pipeline samples). To ensure a fair and meaningful comparison among the predictive models built using different machine learning methods, all models were trained on the same training set and evaluated on the same testing set. Figure 4 presents the distribution of failure counts across homogeneous groups in the training and testing sets, while Figure 5 shows the distribution of failure counts at the individual pipeline sample level.

3.3 Model training and results

Three predictive models-RF, ANN, and SVR-were developed using the failure history over the past five years, diameter, age, and total pipeline length within each homogeneous group as input features, and the total failure count of each group as the output variable.

During model development, several feature transformation techniques were applied to improve model compatibility and training performance. For the categorical feature-failure history over the past five years-the three levels ("No Failure," "Unknown," and "Failure Occurred") were encoded as [-1 -1], [1 -1], and [1 1], respectively. This encoding scheme helps capture the increasing trend of failure risk across categories, thereby enhancing the models' ability to learn from the data. For the numerical features-diameter, age, and length-logarithmic transformation was applied to diameter and length to mitigate the impact of long-tailed distributions. All numerical features were then standardized to eliminate scale differences and facilitate efficient model training.

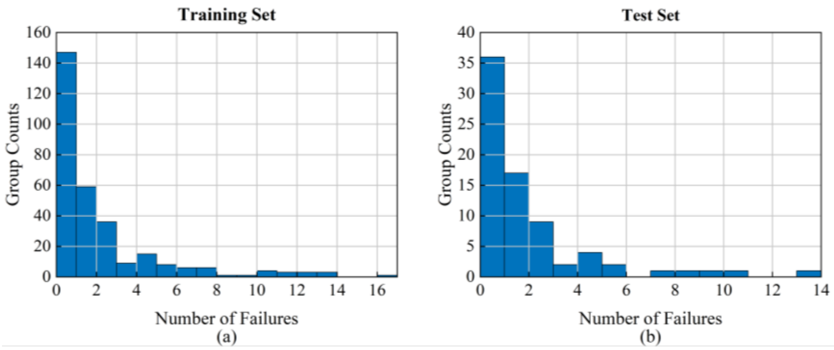

Figure 4: Statistical distribution of the failure counts in homogeneous groups for the training and test sets

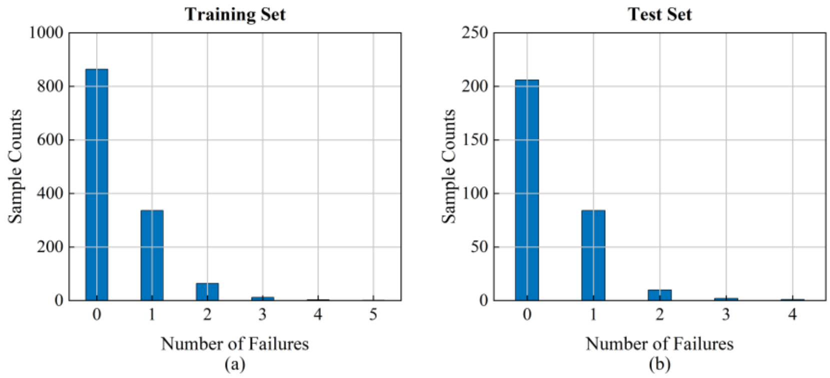

Figure 5: Statistical distribution of the failure counts in individual pipeline samples for the training and test sets.

Table 1: Hyperparameter search ranges for each ML algorithm

`ML algorithms	Hyperparameters
RF	n_ estimators: [10, 500], min_sample_leaf: [1, *n*/2], max_numSplits: [1, *n*-1], max_features: [1, *m*]
ANN	Number of hidden layers: [1,2], Nodes of hidden layers: [1,128], λ_r: [0.00001,100000] /*n*
SVR	*C*: [0.01,10000], kernel scale: [0.001,1000], ε: [0.001,100]×iqr(y)/1.349

In Table 1: $n = 302$ represents the number of homogeneous groups in the training set, $m = 4$ represents the number of features in the training set, and iqr(y) represents the interquartile range of the output variable in the training set.

To develop predictive models with strong generalization performance, we optimized the hyperparameters of three machine learning algorithms using five-fold cross-validation. The optimization was conducted by minimizing the MSE, with Bayesian optimization employed to efficiently explore the hyperparameter space[27]. Tables 1 and 2 present the search ranges and the final optimized values of the hyperparameters for each algorithm.

Table 2: Optimized hyperparameters for each ML algorithm

`ML algorithms	Hyperparameters
RF	n_ estimators=498, min_sample_leaf=5, max_numSplits=44, max_features=2
ANN	Number of hidden layers=1, Nodes of hidden layers=128, λ_r:0.0353
SVR	C=6.5333, kernel scale=2.3044, ε=0.3997

Table 3 presents the performance evaluation results of the three homogeneous group failure count prediction models on both the validation and test sets, where the validation set metrics represent the average values from five-fold cross-validation. The results indicate that all three models exhibit strong predictive capabilities. Overall, except for the MAE metric on the test set, where the ANN model achieves the best performance, the RF model outperforms the others on all metrics for both the validation and test sets. In contrast, the SVR model shows the poorest performance across these metrics.

Table 3: Regression performance of the three models on validation and test sets

`ML algorithms	MSE		RMSE		MAE		R^2	
	Validation set	Test set	Validation set	Test set	Validation set	Test set	Validation set	Test set
RF	**2.0326**	**1.1139**	**1.4257**	**1.0554**	**0.8437**	0.7512	**0.7497**	**0.8240**
ANN	2.1633	1.2458	1.4708	1.1162	0.8831	**0.7404**	0.7335	0.8032
SVR	2.1859	1.3270	1.4785	1.1519	0.9290	0.7906	0.7308	0.7904

Table 4: Performance evaluation of the three models for predicting individual pipeline failure probabilities on validation and test sets

`ML algorithms	BS		L_{CE}		RQ	
	Validation set	Test set	Validation set	Test set	Validation set	Test set
RF	**0.1583**	0.1582	**0.5008**	0.5118	**0.8113**	0.8060
ANN	0.1599	**0.1566**	0.5277	**0.5043**	0.8096	**0.8109**
SVR	0.1663	0.1601	0.5729	0.5170	0.7923	0.8028

To evaluate the applicability of these models for estimating the failure probability of individual pipelines, the failure probability for each pipeline was calculated using Equation (5). Table 4 shows the performance evaluation results of the three models for this task, where the metrics for the validation set again represent the average values from five-fold cross-validation. The results demonstrate that all three models effectively predict the failure probability of individual pipelines, validating the feasibility and effectiveness of the proposed framework. A comprehensive comparison reveals that the RF model achieves the highest accuracy for failure probability estimation on the validation set, while the ANN model performs best on the test set.

4 CONCLUSIONS

This study proposes a machine learning-based analytical framework for assessing the reliability of individual water supply pipelines. The framework begins by integrating each pipeline's attributes with its historical failure records to form samples, where feature variables represent the factors influencing failures and labels represent failure counts. To enhance the statistical robustness of the analysis, the samples are aggregated into homogeneous groups based on feature similarity. Three machine learning models are then developed to predict failure counts for each group. Subsequently, the failure probability of an individual pipeline is estimated by combining its length with the average failure rate of its homogeneous group, enabling a quantitative assessment of its reliability. To validate the effectiveness of the proposed framework, a case study was conducted on gray cast iron water pipelines located in certain administrative regions of Shanghai. The results demonstrate that the three machine learning models can effectively predict failure counts within homogeneous groups and, more importantly, that their outputs can be further utilized to derive sufficiently accurate estimates of failure probabilities for individual pipelines.

Future research will focus on exploring feature selection methods for homogeneous grouping, with the aim of improving the accuracy of reliability estimates for individual pipelines and providing more precise technical support for maintenance and replacement decisions in water utility management.

ACKNOWLEDGMENTS

The authors are thankful for the support of the National Key Research and Development Program (Grant No. 2016YFC0802400).

REFERENCES

[1] Kleiner, Y. and Rajani, B. Forecasting variations and trends in water-main breaks. *J. Infrastruct. Syst.* (2002) 8(4):122–131.

[2] Jiao, H., Hu, Z., Yang, Z., Zeng, W., Xu, F. and Han, C. Hierarchical structure-based model for importance and reliability assessment of water distribution networks. *Reliab. Eng. Syst. Saf.* (2025) 253:110542.

[3] Robles-Velasco, A., Cortés, P., Muñuzuri, J. and De Baets, B. Prediction of pipe failures in water supply networks for longer time periods through multi-label classification. *Expert Syst. Appl.* (2023) 213:119050.

[4] Berardi, L., Giustolisi, O., Kapelan, Z. and Savic, D.A. Development of pipe deterioration models for water distribution systems using EPR. *J. Hydroinform.* (2008) 10(2):113–126.

[5] Tabesh, M., Soltani, J., Farmani, R. and Savic, D. Assessing pipe failure rate and mechanical reliability of water distribution networks using data-driven modeling. *J. Hydroinform.* (2009) 11(1):1–17.

[6] Kabir, G., Tesfamariam, S., Francisque, A. and Sadiq, R. Evaluating risk of water mains failure using a Bayesian belief network model. *Eur. J. Oper. Res.* (2015) 240(1):220–234.

[7] Rajani, B. and Kleiner, Y. Comprehensive review of structural deterioration of water mains: physically based models. *Urban Water* (2001) 3(3):151–164.

[8] Kleiner, Y. and Rajani, B. Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water* (2001) 3(3):131–150.

[9] Sadiq, R., Rajani, B. and Kleiner, Y. Probabilistic risk analysis of corrosion associated failures in cast iron water mains. *Reliab. Eng. Syst. Saf.* (2004) 86(1):1–10.

[10] Davis, P., Burn, S., Moglia, M. and Gould, S. A physical probabilistic model to predict failure rates in buried PVC pipelines. *Reliab. Eng. Syst. Saf.* (2007) 92(9):1258–1266.

[11] Rajani, B. and Tesfamariam, S. Estimating time to failure of ageing cast iron water mains under uncertainties. *ICE Water Manag. J.* (2005) 160(2):83–88.

[12] Xu, Q., Chen, Q., Li, W. and Ma, F. Pipe break prediction based on evolutionary data-driven methods with brief recorded data. *Reliab. Eng. Syst. Saf.* (2011) 96(8):942–948.

[13] Fahmy, M. and Moselhi, O. Forecasting the remaining useful life of cast iron water mains. *J. Perform. Constr. Facil.* (2009) 23(4):269–275.

[14] Jafar, R., Shahrour, I. and Juran, I. Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. *Math. Comput. Model.* (2010) 51(9-10):1170–1180.

[15] Fan, X., Wang, X., Zhang, X. and Yu, X.B. Machine learning based water pipe failure prediction: The effects of engineering, geology, climate and socio-economic factors. *Reliab. Eng. Syst. Saf.* (2022) 219:108185.

[16] Liu, W., Wang, B. and Song, Z. Failure prediction of municipal water pipes using machine learning algorithms. *Water Resour. Manag.* (2022) 36(4):1271–1285.

[17] Cen, H., Huang, D., Liu, Q., Zong, Z. and Tang, A. Application research on risk assessment of municipal pipeline network based on random forest machine learning algorithm. *Water* (2023) 15(10):1964.

[18] Robles-Velasco, A., Cortés, P., Muñuzuri, J. and Onieva, L. Prediction of pipe failures in water supply networks using logistic regression and support vector classification. *Reliab. Eng. Syst. Saf.* (2020) 196:106754.

[19] Francis, R.A., Guikema, S.D. and Henneman, L. Bayesian belief networks for predicting drinking water distribution system pipe breaks. *Reliab. Eng. Syst. Saf.* (2014) 130:1–11.

[20] Breiman, L. Random forests. *Mach. Learn.* (2001) 45:5–32.

[21] Breiman, L., Friedman, J., Olshen, R. and Stone, C.J. Classification and regression trees. *Wadsworth* (1984).

[22] Fausett, L.V. Fundamentals of neural networks: architectures, algorithms and applications. *Pearson Educ. India* (2006).

[23] Rumelhart, D.E., Hinton, G.E. and Williams, R.J. Learning representations by back-propagating errors. *Nature* (1986) 323(6088):533–536.

[24] Cortes, C. and Vapnik, V. Support-vector networks. *Mach. Learn.* (1995) 20:273–297.

[25] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* (2011) 12:2825–2830.

[26] Scheidegger, A., Leitao, J.P. and Scholten, L. Statistical failure models for water distribution pipes–A review from a unified perspective. *Water Res.* (2015) 83:237–247.

[27] Snoek, J., Larochelle, H. and Adams, R.P. Practical bayesian optimization of machine learning algorithms. *Adv. Neural Inf. Process. Syst.* (2012) 25.