A hybrid spherical fuzzy AHP-MARCOS model for evaluating the condition of saltwater pipes in Hong Kong

Nehal Elshaboury
Construction and Project Management Research Institute, Housing and Building National Research Center, Giza, Egypt

Tarek Zayed
Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, and

Eslam Mohammed Abdelkader
Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hung Hom, Hong Kong and Structural Engineering Department, Cairo University, Giza, Egypt

Abstract
Purpose – Water pipes degrade over time for a variety of pipe-related, soil-related, operational, and environmental factors. Hence, municipalities are necessitated to implement effective maintenance and rehabilitation strategies for water pipes based on reliable deterioration models and cost-effective inspection programs. In the light of foregoing, the paramount objective of this research study is to develop condition assessment and deterioration prediction models for saltwater pipes in Hong Kong.

Design/methodology/approach – As a perquisite to the development of condition assessment models, spherical fuzzy analytic hierarchy process (SFAHP) is harnessed to analyze the relative importance weights of deterioration factors. Afterward, the relative importance weights of deterioration factors coupled with their effective values are leveraged using the measurement of alternatives and ranking according to the compromise solution (MARCOS) algorithm to analyze the performance condition of water pipes. A condition rating system is then designed counting on the generalized entropy-based probabilistic fuzzy C means (GEPFCM) algorithm. A set of fourth order multiple regression functions are constructed to capture the degradation trends in condition of pipelines overtime covering their disparate characteristics.

Findings – Analytical results demonstrated that the top five influential deterioration factors comprise age, material, traffic, soil corrosivity and material. In addition, it was derived that developed deterioration models accomplished correlation coefficient, mean absolute error and root mean squared error of 0.8, 1.33 and 1.39, respectively.

Originality/value – It can be argued that generated deterioration models can assist municipalities in formulating accurate and cost-effective maintenance, repair and rehabilitation programs.

Keywords Water pipes, Condition assessment, Deterioration model, SFAHP, MARCOS, GEPFCM

Paper type Research paper

1. Introduction
The water distribution system, a crucial element of urban infrastructure, aims at providing access to clean and safe drinking water to users. According to the Water Supplies
Department (WSD), more than 7 million people in Hong Kong have access to water through more than 8,000 kilometers of water pipes (WSD, 2022a). This city has three water supply sources: seawater for toilet flushing, imported water from Dongjiang, and rainwater from local catchments providing a total water consumption of about 1,377 million cubic meters in 2021 (WSD, 2022b). Since the 1950s, saltwater has been widely used for flushing, enhancing the management and sustainability of water usage in Hong Kong. Approximately 320 million cubic meters of salt water are delivered annually, saving an equivalent volume of fresh water, or about 20% of the total water supply (WSD, 2022c). To significantly reduce the demand for fresh water, WSD was able to expand the coverage of the seawater network from less than 80% to 85% of the population, with the target of providing its flushing system to 90% of the entire population (Water Conservation, 2021).

Due to the congested urban environment, a sizable quantity of subterranean utility services, and high water pressure brought on by the hilly geography, these pipes endure water loss because of leaks or bursts. Hong Kong currently has about 15% water leakage rate (Cheong, 2018). Water loss, erratic water delivery, traffic bottlenecks, and damage to other utilities are all consequences of water pipe failures (WSD, 2022a). Additionally, the WSD launched a water mains replacement and rehabilitation programme in 2000 to replace/rehabilitate about 3,000 km of aged water pipes, reducing the burst rate by 98.4% and the leakage rate by 40% in 2019 (WSD, 2022a). Additionally, the government aims to reduce the leakage rate to below 10% by 2030 (WSD, 2022d). The evaluation of the water pipe deterioration pattern can certainly help the government to set budget plans and maintenance priorities required to provide the required level of service.

Therefore, the main objectives of the current study are summarized as follows:

1. Identify and prioritize the importance of factors influencing the deterioration of water pipes
2. Develop an integrated model for assessing the current condition of water pipes
3. Establish a condition rating system for saltwater pipes
4. Simulate various deterioration models of water pipes considering their diverse physical, soil-related, operational, and environmental characteristics
5. Automate the previously developed models in a standalone computer-aided application

This research offers theoretical and practical contributions such that it identifies the factors affecting water pipe deterioration and develops condition assessment and deterioration prediction models for saltwater pipes from a theoretical perspective. Furthermore, real-world engineering practice will be significantly impacted by the implementation of a condition rating system and the automation of models into a stand-alone application. As such, the goal of this study is to address the gap between theory and practice by offering a thorough framework for the condition assessment and deterioration prediction of saltwater pipes. Although the saltwater network in Hong Kong is the subject of this study, the employed approaches have a great potential for modification and application in the world’s sea coastal cities facing comparable problems in their saltwater networks.

2. Literature review
Several research studies were conducted to assess the condition, deterioration, and/or risk of failure of water pipes. Zhou (2018) estimated the breakage rate and criticality of water distribution systems to optimize their rehabilitation and restoration using an improved genetic algorithm. The pipe breakage prediction model accounted for the physical characteristics and failure rates of the pipes, and it applied a weighted multiple nonlinear regression analysis to
group the pipes according to their condition and degradation propensity. The network criticality was established to account for its condition and hydraulic significance using the technique for order of preference by similarity to the ideal solution method. El-Abbasy et al. (2019) developed deterioration models for water pipes in Canada by accounting for physical, environmental, and operational factors. A fuzzy analytic network process was employed to weigh the factors based on the experts’ responses gathered using questionnaire surveys. Furthermore, the uncertainties of the derived weights were considered using Monte Carlo simulation. The proposed model was validated, and it showed an average validity rate of 93.59%. Chen and Guikema (2020) clustered pipe break data using density-based, Poisson-based, and locally weighted density for a real water pipe network. Spatial clusters have been applied to enhance pipe break machine learning models’ accuracy. The locally weighted density scan provided the highest precision level for locating high breakage zones. The application of clusters enhanced the performance of prediction models in prioritizing high-risk pipes.

Elshaboury et al. (2021) evaluated the mechanical reliability of a water distribution network in Egypt. The mechanical reliability was assessed for pipes, accessories, segments, and the whole network. The condition of network components was evaluated by estimating the weights and effect values of the influential factors. Based on the connection type (series or parallel) among pipes, the condition of segments was evaluated. A minimum cut set analysis was then employed to assess the reliability on the network level. Karimian et al. (2021) used the evolutionary polynomial regression technique to forecast the failure probability of water pipes in the Montreal water network. Pipes were grouped according to their characteristics (length, diameter, age, and material) following the best subset regression. The best-performing model had a determination coefficient of 89.35%. It was revealed that the diameter was the most sensitive factor and ferrous material was the most susceptible material type to aging. Almheiri (2022) proposed novel ensemble and deep learning models to forecast the failure of water distribution pipes in two Canadian municipalities under a variety of environmental and climatic situations. The consequences of critical factors influencing the failure prediction of water pipes were identified such as the minimum antecedent precipitation index, air temperature, and evaporation. The findings showed the influence of geographic location and environment on the failure of water pipes.

Barton et al. (2022) applied a gradient-boosted tree model to forecast the likelihood of pipe failure in the UK water network. The receiver operator curve and area under the curve (0.89), Briers score (0.007), and Mathews correlation coefficient (0.27) showed the accuracy of the prediction model. Additionally, a weighted risk analysis was utilized to graphically represent the high-risk pipes and to indicate the pipe failure consequences. Elshaboury and Marzouk (2022) employed optimized neural network models using genetic algorithm and particle swarm optimization (PSO) to predict the condition of water pipes. The performance of the proposed hybrid models was compared against that of the conventional neural network, group method of data handling, and adaptive neuro-fuzzy inference system. According to the results of the assessment metrics, the neural network trained using the PSO algorithm outperformed other machine learning models (mean bias error = 0.05, root mean square error = 0.09, index of agreement = 0.96, and fraction of prediction within a factor of two = 0.93). Lin et al. (2022) utilized a two-time-scale point method to simulate the degradation and determine the optimal preventative and corrective actions of water pipes. Determining the survival probability and the number of breaks before justifying a replacement action are the outcomes of the preventative and remedial techniques, respectively. These outcomes were investigated using Monte Carlo simulations while considering the degradation intensity and ratios of intervention strategies and consequence costs. Finally, a case study in a Canadian municipality was used to illustrate the applicability of both methodologies in real water systems.

Despite the contributions in the previous research studies, there exist certain gaps in the developed deterioration models for water pipes. The water supply department in Hong Kong
lacks condition assessment and deterioration models for their saltwater pipes. Besides, previous studies have often focused on single methodology or small subsets of factors to forecast pipe deterioration. This study integrates various advanced algorithms, which have not been widely explored within a single comprehensive framework. For better illustration, the spherical fuzzy analytic hierarchy process (SFAHP) is applied to compute the weights of the influential factors, while the effect values of factors are gathered from the literature review and revised by seniors in the WSD in Hong Kong. The current condition indices of water pipes are assessed using the measurement of alternatives and ranking according to the compromise solution (MARCOS) method. Several algorithms are employed to identify the sensitivity of the input factors to any change in the computed weights. The deterioration curves of different pipe materials, sizes, and soil types with respect to age are then developed to represent the change in the current condition over time. The generalized entropy-based possibilistic fuzzy C means (GEPFCM) method is performed to cluster the water pipes based on their condition indices and develop a condition rating scale. Furthermore, this study seeks to contribute by testing the developed models against a large dataset of saltwater pipes in Hong Kong. Developing a robust deterioration model helps decision-makers make more accurate decisions and prepare for appropriate maintenance and rehabilitation tasks.

3. Research methodology
As shown in Figure 1, the first step of the research flowchart is performing a thorough literature analysis to examine the previously developed condition assessment and deterioration models. The primary factors influencing the deterioration of water pipes are identified. Questionnaire surveys are then conducted to pairwise compare the major categories and factors and calculate the impact of each factor characteristic on the pipe deterioration. The pairwise comparisons of the collected responses were described based on Saaty’s fuzzifying scale. Therefore, the relative importance weights of the deterioration factors are computed using the SFAHP approach. The current condition of water pipes is assessed using the MARCOS method. MARCOS is a newly proposed multi-criteria decision-making method that is characterized by its ability to maintain appropriate stability and robustness when accommodating large attributes and alternatives when compared to current methods. In addition, it doesn’t suffer from rank reversal problems (Karaaslan et al., 2022; Stević et al., 2020). Information on saltwater pipes in Hong Kong districts is acquired from the WSD to validate the model. A sensitivity analysis is undertaken to determine the impact of changing the weight of each factor on the computed condition index. Besides, a database of deterioration models is developed while accounting for various material types, sizes, and soil corrosivity types. Finally, the GEPFCM clustering algorithm is performed to develop a condition rating scale for saltwater pipes. A computer program is written in C#.net language for automation of the developed condition assessment and deterioration prediction models.

4. Materials and methods
This section describes the SFAHP weighting technique used to determine the relative weights of importance of the factors influencing water pipe deterioration. Additionally, it provides background information on the MARCOS method used to compute the condition indices of water pipes. Several algorithms are then illustrated to identify the sensitivity of the input factors to any change in the computed weights. Finally, the GEPFCM clustering algorithm is explained to develop a condition rating scale.

4.1 SFAHP
Using a linguistic assessment scale based on spherical fuzzy sets, Kutlu Gündoğdu and Kahraman (2020) established the concept of spherical fuzzy sets to allow evaluators to
express their hesitancies independently. This unique combination enables a more thorough depiction of decision-makers’ ambiguities and uncertainties in the decision-making process. In this context, the SFAHP was applied in diverse engineering domains including the assessment of process mining technology (Dogan, 2021), optimum laminate flooring (Singer and Ozşahin, 2022), sustainable suppliers (Unal and Temur, 2022), operational capacities of airports (Yilmaz et al., 2022), and urban transport problem (Moslem, 2024).

Figure 1. Framework of the developed condition assessment and deterioration model for saltwater pipes

Source(s): Authors own work
Each spherical fuzzy number encompasses three parameters of membership ($\mu_{A^-}$), non-membership ($\pi_{A^-}$) and hesitancy degrees ($\nu_{A^-}$) that lie within the interval of $[0, 1]$. The fundamental procedures of spherical fuzzy sets are discussed as follows:

**Definition 1.** A spherical fuzzy set ($A^-_S$) over the universe of discourse ($U$) is denoted using Equations (1-2).

$$A^-_S = \left\{ x, \left( \mu_{A^-_S}(x), \nu_{A^-_S}(x), \pi_{A^-_S}(x) \right) | x \in U \right\}$$

(1)

where $\mu_{A^-_S}(x) : U \rightarrow [0, 1], \nu_{A^-_S}(x) : U \rightarrow [0, 1], \pi_{A^-_S}(x) : U \rightarrow [0, 1]$

In addition, the sum of parameters needs to satisfy the following condition.

$$0 \leq \mu^2_{A^-_S}(x) + \nu^2_{A^-_S}(x) + \pi^2_{A^-_S}(x) \leq 1 \forall x \in U$$

(2)

**Definition 2.** Assume $A^-_S = \{ \mu_{A^-_S}, \nu_{A^-_S}, \pi_{A^-_S} \}$ and $B^-_S = \{ \mu_{B^-_S}, \nu_{B^-_S}, \pi_{B^-_S} \}$ as two spherical fuzzy numbers, whereas the main arithmetic operations are presented using Equations (3-8).

Union:

$$A^-_S \cup B^-_S = \left\{ \max \left( \mu_{A^-_S}, \mu_{B^-_S} \right), \min \left( \nu_{A^-_S}, \nu_{B^-_S} \right), \right. \left. \min \left\{ \left( 1 - \left( \max \left( \mu_{A^-_S}, \mu_{B^-_S} \right) \right)^2 + \left( \min \left( \nu_{A^-_S}, \nu_{B^-_S} \right) \right)^2 \right) \right\}^{0.5}, \max \left( \pi_{A^-_S}, \pi_{B^-_S} \right) \right\}$$

(3)

Intersection:

$$A^-_S \cap B^-_S = \left\{ \min \left( \mu_{A^-_S}, \mu_{B^-_S} \right), \max \left( \nu_{A^-_S}, \nu_{B^-_S} \right), \right. \left. \max \left\{ \left( 1 - \left( \min \left( \mu_{A^-_S}, \mu_{B^-_S} \right) \right)^2 + \left( \max \left( \nu_{A^-_S}, \nu_{B^-_S} \right) \right)^2 \right) \right\}^{0.5}, \min \left( \pi_{A^-_S}, \pi_{B^-_S} \right) \right\}$$

(4)

Addition:

$$A^-_S \oplus B^-_S = \left\{ \left( \mu^2_{A^-_S} + \mu^2_{B^-_S} - \mu^2_{A^-_S} \mu^2_{B^-_S}, \nu_{A^-_S} \nu_{B^-_S}, \left( \left( 1 - \mu^2_{B^-_S} \right) \right) \pi^2_{A^-_S} + \left( 1 - \mu^2_{A^-_S} \right) \pi^2_{B^-_S} - \pi_{A^-_S} \pi_{B^-_S} \right) \right\}^{0.5}$$

(5)

Multiplication:

$$A^-_S \otimes B^-_S = \left\{ \mu^2_{A^-_S} \mu^2_{B^-_S}, \left( \nu^2_{A^-_S} + \nu^2_{B^-_S} - \nu^2_{A^-_S} \nu^2_{B^-_S} \right), \left( \left( 1 - \nu^2_{B^-_S} \right) \pi^2_{A^-_S} + \left( 1 - \nu^2_{A^-_S} \right) \pi^2_{B^-_S} - \pi_{A^-_S} \pi_{B^-_S} \right) \right\}^{0.5}$$

(6)
Multiplication $A_S^\sim$ by a scalar such that $\lambda > 0$:

$$\lambda A_S^\sim = \left\{ \left( 1 - \left( 1 - \mu_{A_S^\sim}^2 \right)^{\lambda} \right)^{0.5}, \nu_{A_S^\sim}, \left( 1 - \left( 1 - \nu_{A_S^\sim}^2 \right)^{\lambda} \right)^{0.5}, \left( 1 - \nu_{A_S^\sim}^2 - \pi_{A_S^\sim}^2 \right)^{\lambda^{0.5}} \right\}$$  (7)

Power of $A_S^\sim$ by a scalar $\lambda$ such that $\lambda > 0$:

$$A_S^\sim^{\lambda} = \left\{ \mu_{A_S^\sim}^{\lambda}, 1 - \left( 1 - \nu_{A_S^\sim}^2 \right)^{\lambda}, \left( 1 - \nu_{A_S^\sim}^2 \right)^{\lambda}, \left( 1 - \nu_{A_S^\sim}^2 - \pi_{A_S^\sim}^2 \right)^{\lambda^{0.5}} \right\}$$  (8)

Definition 3. For two spherical fuzzy numbers of $A_S^\sim = \{\mu_{A_S^\sim}, \nu_{A_S^\sim}, \pi_{A_S^\sim}\}$ and $B_S^\sim = \{\mu_{B_S^\sim}, \nu_{B_S^\sim}, \pi_{B_S^\sim}\}$, the following assumptions are valid under the conditions of $\lambda, \lambda_1$ and $\lambda_2 > 0$ (see Equations 9-14).

$$A_S^\sim \oplus B_S^\sim = B_S^\sim \oplus A_S^\sim$$  (9)

$$A_S^\sim \odot B_S^\sim = B_S^\sim \odot A_S^\sim$$  (10)

$$\lambda(A_S^\sim \oplus B_S^\sim) = \lambda B_S^\sim \oplus \lambda A_S^\sim$$  (11)

$$\lambda_1 A_S^\sim \odot \lambda_2 A_S^\sim = (\lambda_1 + \lambda_2)A_S^\sim$$  (12)

$$(A_S^\sim \odot B_S^\sim)^{\lambda} = A_S^\sim^{\lambda} \odot B_S^\sim^{\lambda}$$  (13)

$$A_S^{\lambda_1} \odot B_S^{\lambda_2} = A_S^{\lambda_1 + \lambda_2}$$  (14)

Definition 4. Spherical weighted arithmetic mean (SWAM) operator with regards to $w = (w_1, w_2, w_3, \ldots \ldots w_n)$; $w_i \epsilon [0, 1]$; $\sum_{i=1}^{n} w_i = 1$, can be defined using Equation (15).

$$\text{SWAM}_w(A_S^{\sim 1}, A_S^{\sim 2}, A_S^{\sim 3}, \ldots A_S^{\sim n}) = w_1 A_S^{\sim 1} + w_2 A_S^{\sim 2} + w_3 A_S^{\sim 3} + \ldots + w_n A_S^{\sim n}$$

$$= \left[ 1 - \prod_{i=1}^{n} \left( 1 - \mu_{A_S^{\sim i}}^2 \right)^{w_i} \right]^{0.5} \prod_{i=1}^{n} \nu_{A_S^{\sim i}}^{w_i}, \left[ \prod_{i=1}^{n} \left( 1 - \nu_{A_S^{\sim i}}^2 \right)^{w_i} - \prod_{i=1}^{n} \left( 1 - \nu_{A_S^{\sim i}}^2 - \pi_{A_S^{\sim i}}^2 \right)^{w_i} \right]^{0.5}$$  (15)

Definition 5. Spherical weighted geometric mean (SWGM) operator with regards to $w = (w_1, w_2, w_3, \ldots \ldots w_n)$; $w_i \epsilon [0, 1]$; $\sum_{i=1}^{n} w_i = 1$, can be defined using Equation (16).

$$\text{SWGM}_w(A_S^{\sim 1}, A_S^{\sim 2}, A_S^{\sim 3}, A_S^{\sim 4}, \ldots A_S^{\sim n}) = A_S^{\sim 1}^{w_1} + A_S^{\sim 2}^{w_2} + A_S^{\sim 3}^{w_3} + \ldots + A_S^{\sim n}^{w_n}$$

$$= \left[ \prod_{i=1}^{n} \mu_{A_S^{\sim i}}^{w_i} \right]^{0.5} \left[ 1 - \prod_{i=1}^{n} \left( 1 - \nu_{A_S^{\sim i}}^2 \right)^{w_i} \right]^{0.5}, \left[ \prod_{i=1}^{n} \left( 1 - \nu_{A_S^{\sim i}}^2 \right)^{w_i} - \prod_{i=1}^{n} \left( 1 - \nu_{A_S^{\sim i}}^2 - \pi_{A_S^{\sim i}}^2 \right)^{w_i} \right]^{0.5}$$  (16)

The following paragraphs provide a summary of the SFAHP computational steps:

As can be seen in Equation (17), spherical fuzzy pairwise comparison matrices are created by converting linguistic concepts to their corresponding spherical fuzzy integers. In this regard, the experts are asked to instill their preferences towards the deterioration factors using the spherical fuzzy scale given in Table 1 by Kutlu Gündogdu and Kahraman (2020).
where \( a_{ij} = (\mu_{ij}, \nu_{ij}, \pi_{ij}) \) denotes the pairwise comparison between the ith and jth attributes.

The score index \( (a_{ij}) \) for each pairwise comparison element is computed for the linguistic terms of absolutely more importance, very high importance, high importance, slightly more importance, and equal importance using Equation (18). The slightly low importance, low importance, and absolutely low importance are also accommodated by Equation (19).

The entries of paired comparison matrices must be checked and revised if the consistency ratio is less than 10%. The SWAM operator shown in Equation (20) is used to calculate the spherical fuzzy weights of each factor.

Equation (21) is then used to evaluate the significance level of each factor using the defuzzified spherical fuzzy weights of the attributes. These weights must be normalized to get the final and crisp weights of factors as shown in Equation 22.
Saltwater pipes in Hong Kong

4.2 MARCOS

The measurement of alternatives and ranking according to the compromise solution (MARCOS) method was proposed by Stević et al. (2020) to evaluate and rank alternatives based on their compromise solutions. This decision-making method offers several advantages, including the ability to consider various alternatives and criteria without compromising method stability, consider both ideal and anti-ideal solutions, determine utility degree in relation to both solutions, and propose a novel approach to determining utility functions and their aggregation. It was successfully applied in different fields such as performance of insurance companies amidst COVID-19 (Ecer and Pamucar, 2021), synthesis of the efficacy of transportation companies (Miškić et al., 2023), green machining optimization (Shanmugasundar et al., 2022), scrutinization of sources of dust storms (Hosseini Dehshiri et al., 2023), and virtual reality adoption for sustainable construction (Oke et al., 2023). In this research study, the MARCOS method is utilized to compute the condition indices of water pipes. The computational procedures of this method comprise the following steps:

Step 1: Establish the decision matrix and identify the best/ideal solution (AI) and worst/anti-ideal solution (AAI) as per Equations (23-24), respectively:

\[
\begin{align*}
\text{AI} & = \begin{cases} 
\max(X_{ij}) & \text{if } j \text{ is a benefit criterion} \\
\min(X_{ij}) & \text{if } j \text{ is a cost criterion}
\end{cases} \\
\text{AAI} & = \begin{cases} 
\min(X_{ij}) & \text{if } j \text{ is a benefit criterion} \\
\max(X_{ij}) & \text{if } j \text{ is a cost criterion}
\end{cases}
\end{align*}
\]  

(23) (24)

where \( x_{ij} \) refers to the measure of performance of the \( i \)th alternative with respect to the \( j \)th attribute for \( 1 \leq i \leq m, 1 \leq j \leq n \).

Step 2: Normalize the decision matrix using Equation (25).

\[
\begin{align*}
y_{ij} & = \begin{cases} 
\frac{X_{ij}}{X_{aij}} & \text{if } j \text{ is a benefit criterion} \\
\frac{X_{aij}}{X_{ij}} & \text{if } j \text{ is a cost criterion}
\end{cases}
\end{align*}
\]  

(25)

where \( y_{ij} \) refers to the normalized measure of performance of the \( i \)th alternative with respect to the \( j \)th attribute.

Step 3: Formulate the weighted matrix \((c_{ij})\) by multiplying the normalized values by the weights of the criteria \((w)\) using Equation (26).
Step 4: Compute the utility functions \( f(Z_i) \) of the alternatives using Equation (27).

\[
f(Z_i) = \frac{Z_i^+ + Z_i^-}{1 + \frac{1 - f(Z_i^+)}{f(Z_i^-)} + \frac{1 - f(Z_i^-)}{f(Z_i^+)}}
\]

where \( Z_i^- = \frac{S_i}{S_{ai}} \) and \( Z_i^+ = \frac{S_i}{S_{ai}} \) and they represent the utility degrees of alternatives \( Z_i \) concerning the ideal and anti-ideal solutions, respectively. It shall be noted that \( S_i = \sum_{j=1}^{n} c_{ij} \).

Besides, \( f(Z_i^+) = \frac{Z_i^-}{Z_i^+ + Z_i^-} \) and \( f(Z_i^-) = \frac{Z_i^+}{Z_i^+ + Z_i^-} \) and they represent the utility functions for the ideal and anti-ideal solutions, respectively. It is worth mentioning that the utility functions correspond to the overall condition indices of water pipes ranging from 0 to 10, denoting the worst and best pipe conditions, respectively. New pipes with the best pipe-related, soil-related, environmental, and operational factors have the best condition indices (i.e., 10). On the other hand, failed pipes that can no longer perform their primary objectives have the worst condition indices (i.e., 0).

### 4.3 Feature selection algorithms

In this research, eight feature selection algorithms are utilized to determine the importance and sensitivity of the input factors namely regression tree (REG-TR), bagged regression trees (BAG_REGTR), boosted regression trees (BOOS_REGTR), ReliefF, neighborhood component analysis (NCA), sensitivity analysis (SENS_ANA), Pearson correlation coefficient (PCC), maximum information coefficient (MIC). Each of these methods is described briefly in the following subsections. Besides, the sensitivity analysis and Pearson correlation analysis were explained and utilized extensively in previous research studies (El-Abbasy et al., 2019; Nasir et al., 2020; Elshaboury et al., 2021). Finally, an ensemble reciprocal ranking compiles the rankings retrieved from the aforementioned filter-based and embedded-based feature selection algorithms.

#### 4.3.1 Regression tree

Breiman et al. (1984) established a nonparametric statistical method known as classification and regression tree (CART) to address categorical and continuous dependent variables. This method is created by partitioning the original dataset (i.e., root node) into two other nodes following specific rules and properties of these nodes. This process is repeated until reaching terminal nodes with homogeneous elements of less prediction error and less variability. In such a way, the CART algorithm attempts to explain the variation of the target variable by linking it with the explanatory variables (predictors). The three key advantages of this approach are summarized as follows: 1) minimal conditions to arrive at a decision rule, 2) expression of continuous or categorical independent variables, and 3) simplicity of interpreting the outcome and simulating the decision-making process.

#### 4.3.2 Bagged regression trees

Bootstrap aggregation, also known as bagging, is a procedure for reducing the variance and enhancing the prediction accuracy of decision trees (Winkler et al., 2018). The fundamental background of this approach is that the variance of the mean would be \( \sigma^2 / n \) given a set of \( n \) independent observations \( z_1, z_2, \ldots, z_n \) with variance of \( \sigma^2 \). It is used to first build independent classifiers \( \hat{f}_1(x), \hat{f}_2(x), \ldots, \hat{f}_B(x) \) from \( B \) several training sets, which are then averaged using: \( \hat{f}_{\text{avg}}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}_b(x) \). Averaging a set of
observations reduces the variance, but it is impractical to have multiple training samples. In this case, bootstrapping may be used to generate \(B\) distinct bootstrapped training sets, and it is trained on the \(b\)th bootstrapped training set to get \(\hat{f}^{\text{bag}}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}^{b}(x)\). Finally, bootstrap aggregation involves averaging the predictions as follows: \(\hat{f}_{\text{bag}}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}^{b}(x)\).

4.3.3 Boosted regression trees. The boosted regression tree sometimes referred to as gradient boosting, is a decision tree technique that combines several decision trees to get superior results. It builds each tree using a random selection and replacement of the input. The boosting strategy entails giving each tree a weight such that data that was insufficiently described by a previous tree is more likely to be chosen for a subsequent tree. The model accounts for the inaccuracy of the precedent tree’s fit when fitting the subsequent tree, and it uses this sequential method to improve its outcomes over time. The optimum fit is automatically determined, and this modeling approach is stochastic, which enhances the outcomes of the prediction process (Aslani et al., 2021).

4.3.4 ReliefF. Relief, which is particularly effective in estimating the quality of features, served as inspiration for ReliefF. The input features that give different values to neighbors with the same response values are penalized, whereas those that give neighbors with different response values are rewarded (Gong et al., 2022). The condition indices serve as the response values, while the pipe-related, operational, environmental, and soil-related factors serve as the input/explanatory variables. The technique searches for the k-nearest observations to a randomly chosen observation. The weight of features that determines their priority ranking can be determined using this formula: 
\[
\hat{w} = \frac{w_{dy} \cdot w_{x}}{w_{dy} - \frac{w_{dy}}{w_{dr}} \cdot \frac{w_{dy}}{m \cdot w_{dy}}}.
\]
Where; \(w_{dy}\) and \(w_{dr}\) are the weights associated with having various values for the response \(y\) and feature \(r\), respectively, \(w_{dydr}\) is the weight associated with having different response and feature values, and \(m\) is the number of iterations.

4.3.5 Neighborhood component analysis. Yang et al. (2012) suggested a neighborhood component analysis (NCA) feature selection approach to learning feature weights by minimizing an objective function that gauges the average leave-one-out regression loss over the training data. The objective function is expressed as follows: 
\[
\hat{f}(w) = \frac{1}{n} \sum_{i=1}^{n} l_i + \lambda \sum_{i=1}^{p} w_i^2.
\]
Where; \(n\) is the number of observations, \(l_i\) is the distance between the response values and features, \(\lambda\) is the regularization parameter, \(p\) is the average accuracy, and \(w_i\) is the feature weight.

4.3.6 Maximum information coefficient. The maximum information coefficient (MIC) is a powerful feature selection algorithm as proposed by Reshef et al. (2011), and it is used to quantify the degree of correlation between two variables, \(r\) and \(y\). This approach makes the model more representative by removing the feature with less information. The MIC between the response values and features can be calculated as follows: 
\[
\text{MIC}(r; y) = \max_{a \times b \in B} \frac{I(r; y)}{\log_2 \min(a, b)} - \log_2 \min(a, b)
\]
where \(a\) and \(b\) are the number of grids in the \(r\) and \(y\) directions, \(I(r; y)\) is the mutual information between \(r\) and \(y\), and \(B\) is a variable that is set to the 0.6th power of the number of datasets.

4.3.7 Ensemble reciprocal ranking. For large-size datasets with many features that could be unimportant or irrelevant, feature selection algorithms have been developed to identify their subsets of pertinent features. However, due to the biases of some algorithms, ensemble feature selection approaches have emerged as an alternative to incorporate the benefits of single-feature selection algorithms and make up for their drawbacks (Mera-Gaona et al., 2021). In this regard, the developed model harnesses the use of reciprocal ranking to blend the rankings obtained from different feature selection algorithms. In this context, reciprocal Saltwater pipes in Hong Kong
ranking outperformed other reported ensemble feature selection algorithms such as instant runoff, Borda count, Bucklin, Coombs, and Condorcet (Toochaei and Moeini, 2023; Effrosynidis and Arampatzis, 2021). The reciprocal ranking, also known as the inverse rank position, is comparable to the harmonic mean rank. The final reciprocal rank \( r(f) \) of a feature \( f \) is calculated using Equation (28) (Effrosynidis and Arampatzis, 2021):

\[
r(f) = \frac{1}{\sum_{j \in r(f)} r_j(f)}
\]

where \( r_j(f) \) represents the rank of the feature according to the \( j \)th feature selection algorithm, and \( j = 1, 2, ..., N \) represents the investigated feature selection algorithm.

### 4.4 GEPFCM clustering

Using fuzzy clustering, data points are organized into fuzzy clusters, and they are assigned a membership grade based on the similarity of these points to that cluster. The fuzzy c-means (FCM) clustering approach seeks to achieve compact and well-separated clusters (Martino and Sessa, 2020). However, this approach exhibits a weak performance on noisy data. In this regard, a possibilistic FCM algorithm (PFCM) that combines FCM and possibilistic c-means (PCM) is proposed to address the coincident clusters and initialization sensitivity problems. Therefore, it is suggested as a suitable alternative for handling noisy data (Zeng et al., 2022).

The PFCM is formulated by minimizing the index given in Equation (29):

\[
E(B, D, Z; Y) = \sum_{j=1}^{N} \sum_{i=1}^{c} \left( c_{\text{FCM}} b_{ij}^m + c_{\text{PCM}} d_{ij}^\eta \right) \| \overline{y}_j - \overline{z}_i \|_A^2 + \sum_{i=1}^{c} \alpha_i \sum_{j=1}^{N} (1 - d_{ij})^\eta, \sum_{i=1}^{c} b_{ij} = 1
\]

where \( N \) is the number of data vectors, \( c \) is the number of clusters, \( c_{\text{FCM}} \) and \( c_{\text{PCM}} \) are the relative importance of fuzzy membership grades \( b_{ij} \) and \( d_{ij} \), \( b_{ij} \) is the membership grade of the \( j \)th data vector in the \( i \)th cluster \( b_{ij} = \left[ \sum_{k=1}^{c} \left( \frac{\| \overline{y}_j - \overline{z}_i \|_A}{\| \overline{y}_j - \overline{z}_k \|_A} \right)^{\frac{1}{m-1}} \right]^{-1} \), \( d_{ij} \) is the membership degree of data vectors \( d_{ij} = \left( 1 + \left( \frac{c_{\text{FCM}} b_{ij}^m + c_{\text{PCM}} d_{ij}^\eta}{\| \overline{y}_j - \overline{z}_i \|_A} \right)^{\frac{1}{m-1}} \right)^{-1} \), \( m \) is the fuzziness degree, \( \eta \) is a positive nonzero real number, \( \| \overline{y}_j - \overline{z}_i \|_A \) is the distance, \( \overline{y}_j \) is the \( j \)th data vector, \( \overline{z}_i \) is the center of the \( i \)th cluster \( \overline{z}_i = \frac{\sum_{j=1}^{N} (c_{\text{FCM}} b_{ij}^m + c_{\text{PCM}} d_{ij}^\eta) \overline{y}_j}{\sum_{j=1}^{N} (c_{\text{FCM}} b_{ij}^m + c_{\text{PCM}} d_{ij}^\eta)} \), \( A_{r \times r} \) is an appropriate norm matrix, \( r \) is the data dimension, and \( \alpha_i \) is the weighted distance between data vectors and the center of ith cluster \( \alpha_i = K \frac{\sum_{j=1}^{N} b_{ij}^m \| \overline{y}_j - \overline{z}_i \|_A^2}{\sum_{j=1}^{N} b_{ij}^m} \), where \( K = 1 \).
Entropy c-means (ECM) is a different clustering technique that maximizes both compactness and entropy by minimizing the index stated in Equation (30) (Gupta et al., 2019):

$$E(S, Z; Y) = \sum_{j=1}^{N} \sum_{i=1}^{c} s_{ij} ||y_j - z_i||_A^2 + \sum_{j=1}^{N} \sum_{i=1}^{c} s_{ij} \ln(s_{ij}), \sum_{i=1}^{c} s_{ij} = 1$$

where $s_{ij} = \frac{\exp\left(-\frac{||y_j - z_i||_A^2}{2\sigma^2}\right)}{\sum_{k=1}^{c} \exp\left(-\frac{||y_j - z_k||_A^2}{2\sigma^2}\right)}$, $z_i = \frac{\sum_{j=1}^{N} s_{ij} y_j}{\sum_{j=1}^{N} s_{ij}}$, and $\sigma$ is the Lagrangian multiplier.

Due to the high membership grades of noise points produced by the constraint $\sum_{j=1}^{c} s_{ij} = 1$, this ECM version performs poorly on noisy data. Therefore, a relaxed ECM paired with FCM and PCM is proposed for addressing noisy data. The GEPFCM is utilized to identify precise cluster centers of noisy data as it uses a function of distance $f_i(||y_j - z_i||_A^2)$ rather than the distance itself $||y_j - z_i||_A^2$ in the fuzzyistic, possibilistic, and entropy parts of the clustering objective function. The objective function of the GEPFCM algorithm is formulated as per Equation (31) (Askari et al., 2017).

$$E(B, D, S; Y) = \sum_{j=1}^{N} \sum_{i=1}^{c} \left( c_{i,PCM} b_{ij} f_{i,PCM}\left(||y_j - z_i||_A^2\right) + c_{i,PCM} d_{ij} f_{i,PCM}\left(||y_j - z_i||_A^2\right) \right) + c_{PCM} s_{ij} f_{i,E} \left(||y_j - z_i||_A^2\right) + \sum_{j=1}^{N} \sum_{i=1}^{c} s_{ij} \ln(s_{ij}) + \sum_{i=1}^{c} \alpha_i \sum_{j=1}^{N} (1 - d_{ij})^2, \sum_{i=1}^{c} b_{ij} = f_j$$

where $f_{i,PCM}(||y_j - z_i||_A^2)$, $f_{i,FCM}(||y_j - z_i||_A^2)$, and $f_{i,E}(||y_j - z_i||_A^2)$ are functions associated with possibilistic, fuzzy, and entropy terms, respectively.

Data cluster centers can then be determined as per Equation (32):

$$\frac{\partial E^*}{\partial z_i} = \sum_{j=1}^{N} \left( c_{i,PCM} b_{ij} f_{i,PCM}\left(||y_j - z_i||_A^2\right) + c_{i,PCM} d_{ij} f_{i,PCM}\left(||y_j - z_i||_A^2\right) \right) (A + A^T)(y_j - z_i) = 0 \Rightarrow z_i^{k+1} = y_j$$

$$= \frac{\sum_{j=1}^{N} \left( c_{i,PCM} b_{ij} f_{i,PCM}\left(||y_j - z_i^{k+1}||_A^2\right) + c_{i,PCM} d_{ij} f_{i,PCM}\left(||y_j - z_i^{k+1}||_A^2\right) \right) y_j}{\sum_{j=1}^{N} c_{i,PCM} b_{ij} f_{i,PCM}\left(||y_j - z_i^{k+1}||_A^2\right) + c_{i,PCM} d_{ij} f_{i,PCM}\left(||y_j - z_i^{k+1}||_A^2\right) + c_{PCM} s_{ij} f_{i,E}\left(||y_j - z_i^{k+1}||_A^2\right)}$$

(32)

The algorithm stops when the convergence criterion $||z_i^{k+1} - z_i^k|| \leq \epsilon$ is met, where $\epsilon$ is constant and $k$ is the iteration number.

The membership grade of the data vectors ($b_{ij}$) is computed using Equation (33). Meanwhile, the combination of the membership grade and the constraint $\sum_{i=1}^{c} b_{ij} = f_j$ results
in Equation (34). Additionally, the membership degree of data vectors \((d_{ij})\) is calculated using Equation (35).

\[
\frac{\partial E^*(B, D, S, Z; Y)}{\partial b_{ij}} = c_{FCM} mb_{ij}^{n-1} f_i \text{FCM}\left(\frac{||y_j - z_i||_A^2}{A}\right) + \lambda_j = 0 \Rightarrow b_{ij} = \left( c_{FCM} m f_i \text{FCM}\left(\frac{||y_j - z_i||_A^2}{A}\right) \right)^{1/\lambda},
\]

\[
b_{ij} = \left[ \sum_{k=1}^c \left( \frac{f_k \text{FCM}\left(\frac{||y_j - z_i||_A^2}{A}\right)}{f_k \text{FCM}\left(\frac{||y_j - z_k||_A^2}{A}\right)} \right)^{1/n-1} \right]^{-1} f^j
\]

\[
\frac{\partial E^*(B, D, S, Z; Y)}{\partial d_{ij}} = c_{FCM} d_{ij}^{n-1} f_i \text{FCM}\left(\frac{||y_j - z_i||_A^2}{A}\right) - \alpha_i \eta (1 - d_{ij})^{n-1} = 0 \Rightarrow d_{ij}
\]

\[
= \left( 1 + \left( \frac{c_{FCM} f_i \text{FCM}\left(\frac{||y_j - z_i||_A^2}{A}\right)}{\alpha_i} \right)^{1/n} \right)^{-1}, \quad \alpha_i = K \frac{\sum_{j=1}^N b_{ij}^m \left(\frac{||y_j - z_i||_A^2}{A}\right)}{\sum_{j=1}^N d_{ij}^m}
\]

The values of \(B, Z\), and \(\alpha\) are calculated by applying PFCM to the data. Besides, the \(s_{ij}\) for the relaxed GEPFCM is computed using Equation (36). Due to the uniform distribution of \(s_{ij}\) noise points in the clusters caused by the constraint, the relaxed GEPFCM that applies \(f_i \text{FCM}(y), f_i \text{FCM}(y),\) and \(f_i, E(y)\) is more suited for noisy data compared to the constrained GEPFCM.

\[
\frac{\partial E^*(B, D, S, Z; Y)}{\partial s_{ij}} = c_{E_i, E}\left(\frac{||y_j - z_i||_A^2}{A}\right) + \ln(s_{ij}) + 1 = 0 \Rightarrow s_{ij} = \exp\left( -1 - c_{E_i, E}\left(\frac{||y_j - z_i||_A^2}{A}\right) \right)
\]

Each cluster center is enclosed with a hyper-sphere whose radius is estimated using Equation (37).

\[
R^2 = \frac{\sum_{j=1}^N d_{ij}^p (\overline{y}_j - \overline{z}_i) A(\overline{y}_j - \overline{z}_i)}{\sum_{j=1}^N d_{ij}^p}
\]

where \(A\) is the covariance norm matrix, and it is calculated as follows:

\[
A = \left( \frac{1}{N} \sum_{j=1}^N (\overline{y}_j - \overline{z})(\overline{y}_j - \overline{z})^T \right)^{-1}, \quad \overline{z} = \frac{1}{N} \sum_{j=1}^N \overline{y}_j
\]
The general form \( f_i \left( \frac{\rho \| \bar{Y} - \bar{Z} \|_A^2}{R_i} \right) \) is considered for the functions \( f_{i,PCM}(y) \), \( f_{i,FCM}(y) \), and \( f_{i,E}(y) \), and the optimal value of \( \rho \) is chosen such that the \( J_\rho \) index given in Equation (38) is minimized to achieve maximum compactness.

\[
J_\rho = \sum_{j=1}^{N} \sum_{i=1}^{C} s_{ij} \| \bar{Y}^*_j - \bar{Z}^*_i \|_A^2, s_{ij} = \exp \left( -1 - \frac{\| \bar{Y}^*_j - \bar{Z}^*_i \|_A^2}{2} \right)
\]

where \( s_{ij} \) can be estimated by substituting \( c_E = 1 \) and \( f_{i,E}(y) = y, \forall y, i \) in Equation (36).

In this study, the functions \( f_{i,FCM}(\| \bar{Y}^*_j - \bar{Z}^*_i \|_A^2) = f_{i,PCM}(\| \bar{Y}^*_j - \bar{Z}^*_i \|_A^2) = 1 - \exp \left( -\frac{\| \bar{Y}^*_j - \bar{Z}^*_i \|_A^2}{2} \right) \) and \( f_{i,E}(\| \bar{Y}^*_j - \bar{Z}^*_i \|_A^2) = c_E \| \bar{Y}^*_j - \bar{Z}^*_i \|_A^2 \) are utilized to reduce noise contributions inside the hyper-ellipsoid, allowing the algorithm to calculate cluster centers with higher accuracy. It is assumed that \( \varepsilon = 0.00001, c_{FCM} = 1, c_{PCM} = 1, \eta = 2, m = 2, f_j = 1 \forall j \).

5. Data collection
In this study, data is acquired from four major sources: case study, literature studies, questionnaire surveys, and focus group. Each of these sources is described in the next subsections.

5.1 Case study description
The network of saltwater pipes in Hong Kong has 160,315 saltwater segments totaling 1854.4 kilometers in length. The network comprises different pipe materials; ductile iron, steel, stainless steel, polyethylene, cast iron, asbestos cement as well as lined and unlined galvanized iron. Nearly 44.6 and 36.7% are metallic and plastic pipes, respectively. Pipe diameters range between 20 and 1,500 millimeters. Around 10,773 (6.7%), 115,156 (71.8%), and 34,386 (21.4%) of the pipes are installed in highly corrosive, moderate corrosive, and noncorrosive soils, respectively. Moreover, 99.9% of the pipes (i.e., 160,282) are not cathodically protected.

Regarding the failure history, the failure records are available throughout the years 2010–2020, and around 4,275 pipes (2.7%) have previously experienced either a leak or burst. The average water pressure is 8.4 bar, with more than 70% of the pipes (i.e., 118,302) operating under pressure ranging between 5 and 15 bar. For the road type, nearly 39.7 and 42.9% of the pipes are installed under the carriageway and footway. Almost 90.8% of the pipes are installed in urban areas concerning land use. All these data were acquired from the WSD, which is responsible for managing the water distribution network in Hong Kong. These data were combined with weather and traffic data given by the Hong Kong Observatory and Transportation Department, respectively to develop a thorough database of the factors influencing the deterioration of water pipes. Considering the weather factors, the average total rainfall, mean air temperature, and relative humidity are 88.2 mm, 19.2 °C, and 72.7%, respectively. More than 50% of the pipes are subjected to 5550–24500 annual average daily traffic (AADT).

5.2 Previous literature studies
Literature studies serve as one of the main sources of data for this study. From the literature, a thorough list of factors influencing water pipe deterioration was compiled (El-Abbasy et al.,
These factors are grouped into four categories, namely pipe-related, soil-related, operational, and environmental. The pipe-related category refers to the construction and manufacturing process of pipes such as diameter, material, length, and age. Soil-related factors, such as soil corrosivity and cathodic protection, are among the most significant factors that determine pipe corrosion and degradation. Operational factors including failure history and water pressure have an impact on how the water network is operated and maintained. The environmental category incorporates factors related to rainfall, air temperature, relative humidity, traffic, road type, and land use. Table 2 provides categories, sources, and descriptions of the influential deterioration factors of water pipes.

It should be emphasized that other significant factors should have been considered while evaluating pipe deterioration. An example of these factors is the pipe internal protection method which is essential for resisting internal corrosion and enhancing the pipe condition. Other factors include pipe manufacturing, installation, and connection methods as well as bedding condition, backfilling material, and buried depth that can affect the pipe degradation pattern. However, these factors were not incorporated in the distributed survey since there does not exist sufficient information on them in the case study utilized to validate the developed models.

5.3 Questionnaire surveys and focus group

Questionnaire surveys were undertaken with a wide spectrum of water infrastructure asset management specialists in Hong Kong to identify the importance level of the identified factors. Between February and June 2022, a total of 27 questionnaire surveys were gathered using Google Forms. Academicians, governmental engineers, managers/coordinators, facility managers, site supervisors, or management consultants are among the experts’ backgrounds. In terms of experience, around 7%, 30%, 7%, and 56% of the participants have less than a year, one to five years, six to ten years, and more than ten years of experience, respectively. The pairwise comparison matrices are developed based on the gathered responses to show the relative significance between the deterioration factors and the category to which they belong, as well as between the factor categories and the primary objective. Each expert, based on his/her professional experience, determines the relative importance of each factor with respect to another through pairwise comparison at each level. In The questionnaire survey, the significance level of factors is expressed based on a linguistic scale ranging from “absolutely low importance” to “absolutely more importance”. Table 3 exemplifies the pairwise comparisons necessitated to analyze the importance levels of pipe-related factors.

Meanwhile, a focus group was conducted with the WSD specialists to estimate the effect values of the deterioration factors. Each factor may have a variety of characteristics with varying effects on the water pipe deterioration rate. The effect values range between 0 and 10, which correspond to the pipe’s worst and best influence on the pipe condition, respectively. Consider the “diameter” factor only, pipe condition with small, medium, and large sizes will be 3.5, 6, and 9, respectively.

6. Results and discussion

Using the 27 responses gathered through surveys, the SFAHP computational procedures are followed to establish the final weights of the deterioration factors as described as follows. Each member of the panel of experts is required independently to assess the level of importance of deterioration factors, and then the consolidated relative importance vector of deterioration factors is rendered by averaging the relative importance weights obtained by
The experts. The following demonstrates a sample of the results based on the evaluations dispensed by one of the experts. A pairwise comparison matrix is constructed for the deterioration categories predicated on the obtained expert’s evacuations (see Table 4). For instance, the expert delegated a preference of “slightly low importance” for the pipe-related factors when compared to operational factors. In addition, it was given that soil-related factors are slightly more important than environmental-related factors. A consistency test is carried out to measure the logicality of an expert’s judgments. The weighted sum vector (WSV) is obtained by transferring the linguistic variables to their corresponding scoring index and then multiplying it by the weights calculated using the classical analytical hierarchy process. The consistency vector (CV) is then generated by dividing the weighted sum vector by the weighting vector of AHP. Afterward, the consistency ratio (CR) is

<table>
<thead>
<tr>
<th>Category</th>
<th>Factor</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pipe-related</td>
<td>Diameter</td>
<td>This term refers to the pipe’s outer diameter. Large-diameter pipes have a thicker wall thickness, and they are subjected to reduced breakage rates.</td>
</tr>
<tr>
<td></td>
<td>Material</td>
<td>Some pipe material types perform better than others, implying that the failure patterns vary depending on the material.</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>Pipe length refers to how long the pipe is. The operational stress and traffic volume have a positive correlation with pipe length. As a result, the longer pipe is more likely to fail than the shorter one.</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>Pipe age refers to the length of time since installation. It is strongly correlated to the failure probability such that older pipes endure significant degradation compared to newly installed pipes.</td>
</tr>
<tr>
<td>Soil-related</td>
<td>Soil corrosivity</td>
<td>The corrosivity of the soil is a sign of how aggressive the soil is. Significant pipe breakage is caused by very corrosive soil. Soils that are acidic or very alkaline are prone to accelerate corrosion.</td>
</tr>
<tr>
<td></td>
<td>Cathodic protection</td>
<td>External corrosion brought on by pipe installation, soil characteristics, or the lack of cathodic protection is a common failure problem.</td>
</tr>
<tr>
<td>Operational</td>
<td>Failure history</td>
<td>The rate of pipe failure is directly related to pipe deterioration.</td>
</tr>
<tr>
<td></td>
<td>Water pressure</td>
<td>The pipe is affected by the internal water pressure, which leads to pipe failure and breakage.</td>
</tr>
<tr>
<td>Environmental</td>
<td>Rainfall</td>
<td>This is the total rainfall amount at different pipe sections. Pipe collapse frequently occurs in bad weather conditions (drought and heavy rain). These events cause unstable ground conditions and substantial damage to the pipe.</td>
</tr>
<tr>
<td></td>
<td>Air temperature</td>
<td>This is the average temperature at different pipe sections. The frequency of pipe breaks increases in cold weather and the resulting high tensions. Moreover, due to the soil shrinking and drying up in the summer, pipes frequently burst.</td>
</tr>
<tr>
<td></td>
<td>Relative humidity</td>
<td>Humidity is a crucial element in the corrosion of metals. Additionally, it has a big impact on how effectively coatings cure.</td>
</tr>
<tr>
<td></td>
<td>Traffic</td>
<td>This describes the traffic volume across each underground pipe. Heavy traffic sometimes causes water main breaks. However, it does not always result in higher external pressure on the pipe.</td>
</tr>
<tr>
<td></td>
<td>Road type</td>
<td>Road type demonstrates the type of road that crosses over the underground pipes. Due to the various surface types, pipes installed in various places experience varied degradation patterns (e.g., asphalt, seal, and unpaved).</td>
</tr>
<tr>
<td></td>
<td>Land use</td>
<td>Land use explains how land near the underground pipes is being exploited. Pipe failure is accelerated by rising water demand and population density.</td>
</tr>
</tbody>
</table>

Source(s): Authors’ own work

Table 2. Descriptions of the influential deterioration factors of water pipes in Hong Kong.
Significance of the factor in assessing pipe deterioration

<table>
<thead>
<tr>
<th>Pipe-related factor</th>
<th>Absolutely more importance</th>
<th>Absolutely more importance</th>
<th>Slightly more importance</th>
<th>High importance</th>
<th>Very high importance</th>
<th>Slightly low importance</th>
<th>Equally important</th>
<th>Slightly low importance</th>
<th>Low importance</th>
<th>Very low importance</th>
<th>Absolutely low importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Material</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Material</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source(s): Authors' own work
calculated by dividing the consistency index by the random index. It was found that the consistency ratio is 5.71% (<10%) which implies a satisfactory level of consistency in the judgments of the expert.

\[
WSV = \begin{bmatrix} 1 & 1/3 & 1/3 & 1 \\ 3 & 1 & 3 & 3 \\ 3 & 1/3 & 1 & 3 \\ 1 & 1/3 & 1/3 & 1 \end{bmatrix} \begin{bmatrix} 0.122 \\ 0.473 \\ 0.283 \\ 0.122 \end{bmatrix} = \begin{bmatrix} 0.496 \\ 2.054 \\ 1.173 \\ 0.496 \end{bmatrix}
\]

\[
CV = \begin{bmatrix} 0.122 & 0.496 \\ 0.473 & 2.054 \\ 0.283 & 1.173 \\ 0.122 & 0.496 \end{bmatrix} = \begin{bmatrix} 4.065 \\ 4.34 \\ 4.147 \\ 4.065 \end{bmatrix}
\]

\[
CR = \frac{4.154}{0.9} = 5.71\%
\]

Average consistency ratio is computed based on compiling pairwise comparison entries of all respondents. The results of consistency ratio tests are given in Table 5. It is found that the average consistency ratio of main deterioration factors is 0.17%. Likewise, the average consistency ratio of pipe-related factors, soil-related factors, operational-related factors and environmental-related factors are 1%, 0.14%, 2.5% and 1.54%, respectively. It can be seen that the average consistency ratios of deterioration categories and factors are less than 10%. Hence, the pairwise judgments of respondents are considered as consistent and reliable.

Hitherto the pairwise comparisons are consistent, and the weights of deterioration categories need to be calculated. Table 6 shows the spherical comparison matrix of deterioration categories that is subsequently used to derive the spherical fuzzy weights. Once the spherical fuzzy pairwise comparison matrix of deterioration categories is generated, spherical fuzzy weights are calculated using the SWAM operator. These weights are further de-fuzzified and normalized to calculate the normalized de-fuzzified weights of deterioration categories. In this context, this table presents spherical fuzzy weights, de-fuzzified, and normalized de-fuzzified weights.

<table>
<thead>
<tr>
<th>Deterioration category</th>
<th>Pipe-related factors</th>
<th>Soil-related factors</th>
<th>Operational factors</th>
<th>Environmental factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pipe-related factors</td>
<td>EI</td>
<td>SLI</td>
<td>SLI</td>
<td>EI</td>
</tr>
<tr>
<td>Soil-related factors</td>
<td>SMI</td>
<td>EI</td>
<td>SMI</td>
<td>SMI</td>
</tr>
<tr>
<td>Operational factors</td>
<td>SMI</td>
<td>SLI</td>
<td>EI</td>
<td>SMI</td>
</tr>
<tr>
<td>Environmental factors</td>
<td>EI</td>
<td>SLI</td>
<td>SLI</td>
<td>EI</td>
</tr>
</tbody>
</table>

**Source(s):** Authors’ own work

<table>
<thead>
<tr>
<th>Deterioration categories and factors</th>
<th>Average consistency ratio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main deterioration categories</td>
<td>0.17</td>
</tr>
<tr>
<td>Pipe-related factors</td>
<td>1.00</td>
</tr>
<tr>
<td>Soil-related factors</td>
<td>0.14</td>
</tr>
<tr>
<td>Operation-related factors</td>
<td>2.50</td>
</tr>
<tr>
<td>Environment-related factors</td>
<td>1.54</td>
</tr>
</tbody>
</table>

**Source(s):** Authors’ own work

<table>
<thead>
<tr>
<th>Table 4.</th>
<th>Pairwise comparison matrix for the deterioration categories</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Table 5.</th>
<th>Average consistency ratios of deterioration categories and factors</th>
</tr>
</thead>
</table>
Table 6. Spherical fuzzy weights, de-fuzzified weights, and normalized final weights of deterioration categories.

<table>
<thead>
<tr>
<th>Deterioration category</th>
<th>Pipe-related factors (PF)</th>
<th>Soil-related factors (SF)</th>
<th>Operational factors (OF)</th>
<th>Environmental factors (EF)</th>
<th>Spherical fuzzy weights ($\mu^-, \nu^-, \pi^-$)</th>
<th>De-fuzzified weights ($S(w^-_j)$)</th>
<th>SFAHP normalized final weights ($w^-_j$) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pipe-related factors</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.454, 0.49, 0.358)</td>
<td>11.778</td>
<td>22.28</td>
</tr>
<tr>
<td>Soil-related factors</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.578, 0.4, 0.326)</td>
<td>15.661</td>
<td>29.63</td>
</tr>
<tr>
<td>Operational factors</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.536, 0.443, 0.328)</td>
<td>14.397</td>
<td>27.27</td>
</tr>
<tr>
<td>Environmental factors</td>
<td>(0.4, 0.5, 0.4)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.428, 0.518, 0.357)</td>
<td>11.023</td>
<td>20.85</td>
</tr>
</tbody>
</table>

Source(s): Authors' own work
normalized de-fuzzified weights of deterioration categories. It can be observed that the category of “soil-related factors” (29.63%) is the most important followed by “operational factors” (27.27%) and then “pipe-related factors” (22.28%). An example of the calculation of the normalized crisp weights of the “pipe-related factors” is described in the following lines.

\[
\begin{align*}
\mu_{PF^*} &= \left[1 - \prod_{i=1}^{n} \left(1 - \mu_{A_S}^2ight)^{w_i}\right]^{0.5} \\
&= \left[1 - (1 - 0.5^2)^{0.25} (1 - 0.4^2)^{0.25} (1 - 0.2^2)^{0.25} (1 - 0.5^2)^{0.25}\right]^{0.5} = 0.454 \\
\nu_{PF^*} &= \prod_{i=1}^{n} \left(1 - \mu_{A_S}^2\right)^{w_i} = \left[0.4^{0.25} \times 0.6^{0.25} \times 0.6^{0.25} \times 0.4^{0.25}\right] = 0.49 \\
\pi_{PC^*} &= \left[\prod_{i=1}^{n} \left(1 - \mu_{A_S}^2\right)^{w_i} - \prod_{i=1}^{n} \left(1 - \mu_{A_S}^2 - \pi_{A_S}^2\right)^{w_i}\right]^{0.5} \\
&= \left[(1 - 0.5^2)^{0.25} (1 - 0.4^2)^{0.25} (1 - 0.2^2)^{0.25} (1 - 0.5^2)^{0.25} - (1 - 0.5^2 - 0.4^2)^{0.25}\right]^{0.5} = 0.358 \\
S(\overline{w}_{PF}) &= \sqrt{\left\lceil 100 \times \left[\left(3 \times 0.454 - \frac{0.49}{2}\right)^2 - \left(\frac{0.49}{2} - 0.358\right)^2\right]\right\rceil} = 11.778 \\
\overline{w}_{PF} &= \frac{11.778}{11.778 + 15.661 + 14.397 + 12.787 + 11.023} = 22.28%
\end{align*}
\]

The spherical comparison matrix, spherical fuzzy weights, de-fuzzified weights, and normalized de-fuzzified weights of pipe-related factors are then computed in Table 7. It can be noticed that age (32.48%) is the most important pipe-related factor followed by material with a relative importance weight of 25.84%. In addition, length is found to be the least important pipe-related factor with a relative importance weight of 18.48%. The relative importance weights of other deterioration factors are computed similarly. In this context, Table 8 depicts the comparison matrix, spherical fuzzy weights, de-fuzzified weights, and normalized de-fuzzified weights of environmental factors. It can be viewed that traffic volume (22.92%) exhibits the highest level of significance and road type (19.17%) is ranked in second place while rainfall (12.2%) sustains the lowest level of importance.

Another example for the calculation of the normalized weight of rainfall is delineated as follows.

\[
\begin{align*}
\mu_{tv^*} &= \left[1 - \prod_{i=1}^{n} \left(1 - \mu_{A_S}^2\right)^{w_i}\right]^{0.5} \\
&= \left[1 - (1 - 0.5^2)^{\frac{1}{4}} (1 - 0.4^2)^{\frac{1}{4}} (1 - 0.5^2)^{\frac{1}{4}} (1 - 0.2^2)^{\frac{1}{4}} (1 - 0.3^2)^{\frac{1}{4}}\right]^{0.5} = 0.388 \\
\nu_{tv^*} &= \prod_{i=1}^{n} \left(1 - \mu_{A_S}^2\right)^{w_i} = \left[0.4^{\frac{1}{4}} \times 0.6^{\frac{1}{4}} \times 0.6^{\frac{1}{4}} \times 0.8^{\frac{1}{4}} \times 0.7^{\frac{1}{4}}\right] = 0.579
\end{align*}
\]
<table>
<thead>
<tr>
<th>Pipe-related factors</th>
<th>Diameter</th>
<th>Material</th>
<th>Length</th>
<th>Age</th>
<th>Spherical fuzzy weights ($\mu^-\nu^\pi$)</th>
<th>De-fuzzified weights ($S(w^-_j)$)</th>
<th>SFAHP normalized final weights ($w^-_j$) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diameter</td>
<td>(0.5, 0.4, 0.3)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.487, 0.49, 0.331)</td>
<td>12.928</td>
<td>23.2</td>
</tr>
<tr>
<td>Material</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.536, 0.443, 0.328)</td>
<td>14.397</td>
<td>25.84</td>
</tr>
<tr>
<td>Length</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.2, 0.8, 0.1)</td>
<td>(0.395, 0.583, 0.308)</td>
<td>10.298</td>
<td>18.48</td>
</tr>
<tr>
<td>Age</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.8, 0.2, 0.1)</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.651, 0.336, 0.277)</td>
<td>18.101</td>
<td>32.48</td>
</tr>
</tbody>
</table>

**Source(s):** Authors’ own work
<table>
<thead>
<tr>
<th>Environmental factors</th>
<th>Rainfall</th>
<th>Air temperature</th>
<th>Relative humidity</th>
<th>Traffic</th>
<th>Road type</th>
<th>Land use</th>
<th>Spherical fuzzy weights ($\mu^<em>, \nu^</em>, \pi^*$)</th>
<th>De-fuzzified weights ($S(w_j)$)</th>
<th>SFAHP normalized final weights ($w'_j$) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rainfall</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.2, 0.8, 0.1)</td>
<td>(0.3, 0.7, 0.2)</td>
<td>(0.3, 0.7, 0.2)</td>
<td>(0.388, 0.579, 0.305)</td>
<td>10.117</td>
<td>12.2</td>
</tr>
<tr>
<td>Air temperature</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.3, 0.7, 0.2)</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.515, 0.439, 0.346)</td>
<td>13.664</td>
<td>16.47</td>
</tr>
<tr>
<td>Relative humidity</td>
<td>(0.4, 0.5, 0.4)</td>
<td>(0.4, 0.5, 0.4)</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.2, 0.8, 0.1)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.396, 0.554, 0.343)</td>
<td>10.159</td>
<td>12.25</td>
</tr>
<tr>
<td>Traffic volume</td>
<td>(0.8, 0.2, 0.1)</td>
<td>(0.7, 0.3, 0.2)</td>
<td>(0.8, 0.2, 0.1)</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.673, 0.324, 0.236)</td>
<td>19.009</td>
<td>22.92</td>
</tr>
<tr>
<td>Road type</td>
<td>(0.7, 0.3, 0.2)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.581, 0.408, 0.302)</td>
<td>15.899</td>
<td>19.17</td>
</tr>
<tr>
<td>Land use</td>
<td>(0.7, 0.3, 0.2)</td>
<td>(0.4, 0.5, 0.4)</td>
<td>(0.6, 0.4, 0.3)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.4, 0.6, 0.3)</td>
<td>(0.5, 0.4, 0.4)</td>
<td>(0.525, 0.453, 0.323)</td>
<td>14.107</td>
<td>17.01</td>
</tr>
</tbody>
</table>

**Source(s):** Authors’ own work
The weights of the remaining factors are computed in the same manner. The final global weights of the pipe-related, soil-related, operational, and environmental factors are depicted in Table 9. Among the four categories, the water pressure factor has the highest weight of importance (12.03%), soil corrosivity (11.40%), cathodic protection (11.08%), land use (9.53%), and failure history (9.21%). On the level of main categories, the environmental category is the most important (30.13%), followed by pipe-related (26.15%), soil-related (22.48%), and operational (21.24%) categories, respectively. Meanwhile, the effect values of the deterioration factors are estimated as depicted in the same table. The weights along with the effect values of the influential factors are utilized to compute the condition indices of water pipes.

The computational procedures of the MARCOS technique are described as follows: The extended decision matrix is established, and the best and worst solutions are identified. For instance, the ideal solutions (AI) for the diameter, material, length, age, soil corrosivity, cathodic protection, failure occurrence, water pressure, rainfall, air temperature, relative humidity, traffic, road type, and land use are 9.0, 8.5, 9.0, 9.5, 9.5, 9.0, 9.0, 9.5, 9.5, 9.5, 9.5, 7.5, and 9.5, respectively. Meanwhile, the anti-ideal solutions (AAI) for these factors are 3.5, 3.5, 5.0, 4.5, 1.5, 6.0, 3.5, 5.0, 7.5, 8.0, 7.5, 7.0, 4.5, and 6.0, respectively.

Consider a 150 mm diameter ductile iron water pipe that has a length of 50.3 meters and an age of 27 years old. Besides, the pipe is installed in moderate corrosive soil, carriageway road type, and urban land use. It is subjected to a water pressure of 6.5 bar, total rainfall of 42.6 mm, an average temperature of 17.8 °C, relative humidity of 75 %, and average traffic of 10,570 AADT. Additionally, this pipe does not have cathodic protection and experienced a failure (i.e., break or leak) before. It is assumed that all the assessment factors are beneficial criteria that need to be maximized. The extended decision matrix is then normalized, and the normalized values ($n_{ij}$) are multiplied by the weights of the criteria to determine the weighted matrix ($c_{ij}$). The $S_1$ value is estimated as follows: $S_1 = [6.6\% \times (\frac{7.8}{7.8})] + [8.1\% \times (\frac{7.5}{7.8})] + [2.3\% \times (\frac{7.9}{7.8})] + [9.2\% \times (\frac{7.5}{7.8})] + [11.4\% \times (\frac{7.5}{7.8})] + [11.1\% \times (\frac{7.9}{7.8})] + [9.2\% \times (\frac{7.5}{7.8})] + [12.0\% \times (\frac{7.5}{7.8})] + [2.1\% \times (\frac{8.5}{7.8})] + [2.9\% \times (\frac{8.5}{7.8})] + [2.8\% \times (\frac{8.5}{7.8})] + [5.0\% \times (\frac{9.0}{7.8})] + [7.8\% \times (\frac{8.0}{7.8})] + [9.5\% \times (\frac{8.9}{7.8})] = 0.69.

The utility degrees of alternatives concerning the ideal ($Z^+_i$) and anti-ideal ($Z^-_i$) solutions can be computed as follows: $Z^+_i = \frac{0.69}{1.00} = 0.69$ and $Z^-_i = \frac{0.69}{0.32} = 1.33$. Besides, the utility functions for the ideal ($f(Z^+_i)$) and anti-ideal ($f(Z^-_i)$) solutions can be calculated as follows: $f(Z^+_i) = \frac{0.32}{0.69+1.33} = 0.66$ and $f(Z^-_i) = \frac{0.69+1.33}{0.69 \times 1.33} = 0.34$. Finally, the utility function ($f(Z_i)$) of the pipe is estimated as follows: $f(Z_i) = \frac{0.69+1.33}{0.69 \times 1.33} = 0.59$. Therefore, it can be interpreted.
by examining the pipe’s characteristics that it has a condition index of 5.9. The same procedures are repeated for the other pipes. Table 10 displays a sample of the condition assessment database for saltwater pipes.
Table 10. Condition indices of a sample of saltwater pipes

<table>
<thead>
<tr>
<th>Diameter</th>
<th>Material</th>
<th>Length</th>
<th>Age</th>
<th>Soil corrosivity</th>
<th>Cathodic protection</th>
<th>Failure history</th>
<th>Water pressure</th>
<th>Rainfall</th>
<th>Temperature</th>
<th>Humidity</th>
<th>Traffic</th>
<th>Road type</th>
<th>Land use</th>
<th>Condition index</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>DI</td>
<td>50.3</td>
<td>27</td>
<td>Moderate</td>
<td>No</td>
<td>Yes</td>
<td>6.5</td>
<td>42.6</td>
<td>17.8</td>
<td>75.0</td>
<td>10,570</td>
<td>Carriageway</td>
<td>Urban</td>
<td>5.9</td>
</tr>
<tr>
<td>300</td>
<td>DI</td>
<td>2.1</td>
<td>11</td>
<td>Noncorrosive</td>
<td>No</td>
<td>Yes</td>
<td>8.3</td>
<td>157.3</td>
<td>18.1</td>
<td>80.0</td>
<td>26,780</td>
<td>Carriageway</td>
<td>Urban</td>
<td>6.5</td>
</tr>
<tr>
<td>80</td>
<td>PE</td>
<td>1.2</td>
<td>16</td>
<td>Noncorrosive</td>
<td>No</td>
<td>Yes</td>
<td>15.7</td>
<td>378.8</td>
<td>22.0</td>
<td>75.5</td>
<td>2940</td>
<td>Footway</td>
<td>Urban</td>
<td>6.5</td>
</tr>
<tr>
<td>80</td>
<td>PE</td>
<td>33.7</td>
<td>10</td>
<td>Noncorrosive</td>
<td>No</td>
<td>Yes</td>
<td>15.7</td>
<td>86.6</td>
<td>22.8</td>
<td>77.0</td>
<td>42,460</td>
<td>Footway</td>
<td>Urban</td>
<td>6.3</td>
</tr>
<tr>
<td>150</td>
<td>S</td>
<td>9.1</td>
<td>13</td>
<td>Moderate</td>
<td>No</td>
<td>Yes</td>
<td>6.8</td>
<td>86.8</td>
<td>26.0</td>
<td>74.5</td>
<td>51,805</td>
<td>Footway</td>
<td>Urban</td>
<td>6.0</td>
</tr>
<tr>
<td>400</td>
<td>S</td>
<td>48.1</td>
<td>32</td>
<td>Moderate</td>
<td>No</td>
<td>Yes</td>
<td>7.3</td>
<td>215.8</td>
<td>16.7</td>
<td>84.5</td>
<td>17,170</td>
<td>Carriageway</td>
<td>Urban</td>
<td>5.8</td>
</tr>
<tr>
<td>300</td>
<td>S</td>
<td>2.0</td>
<td>11</td>
<td>Moderate</td>
<td>No</td>
<td>Yes</td>
<td>4.3</td>
<td>226.8</td>
<td>21.4</td>
<td>91.0</td>
<td>7830</td>
<td>Carriageway</td>
<td>Urban</td>
<td>6.2</td>
</tr>
<tr>
<td>600</td>
<td>S</td>
<td>2.4</td>
<td>11</td>
<td>Moderate</td>
<td>No</td>
<td>Yes</td>
<td>8.3</td>
<td>38.0</td>
<td>19.8</td>
<td>74.5</td>
<td>71,645</td>
<td>Footway</td>
<td>Urban</td>
<td>6.4</td>
</tr>
<tr>
<td>200</td>
<td>DI</td>
<td>76.1</td>
<td>35</td>
<td>Moderate</td>
<td>No</td>
<td>Yes</td>
<td>0.2</td>
<td>116.8</td>
<td>15.3</td>
<td>67.5</td>
<td>52,590</td>
<td>Carriageway</td>
<td>Urban</td>
<td>5.8</td>
</tr>
<tr>
<td>600</td>
<td>S</td>
<td>40.1</td>
<td>11</td>
<td>Moderate</td>
<td>No</td>
<td>Yes</td>
<td>8.3</td>
<td>107.5</td>
<td>16.2</td>
<td>74.5</td>
<td>71,245</td>
<td>Footway</td>
<td>Urban</td>
<td>6.3</td>
</tr>
</tbody>
</table>

**Source(s):** Authors' own work
The rankings of the pipe-related, soil-related, operational, and environmental factors using the eight feature selection algorithms are evinced in Figure 2. The results illustrate the variations in the rankings across the feature selection algorithms. For instance, the water pressure factor is associated with the sixth, second, fourth, fourth, fourteenth, eighth, and tenth rankings, respectively according to regression tree, bagged regression trees, boosted regression trees, ReliefF, neighborhood component analysis, sensitivity analysis, Pearson correlation coefficient, and maximum information coefficient. In addition, soil corrosivity is ranked in the third, first, second, ninth, ninth, fifth, third, and twelfth based on regression tree, bagged regression tree, boosted regression tree, ReliefF, neighborhood component analysis, sensitivity analysis, Pearson correlation coefficient, and maximum information coefficient, respectively. It can be also seen that age is the most important deterioration factor according to regression tree, boosted regression trees, and Pearson correlation coefficient. However, traffic is found to be the most important based on ReliefF and neighborhood component analysis. Additionally, failure history, soil corrosivity, and diameter are the most influential deterioration factors counting on sensitivity analysis, bagged regression trees, and maximum information coefficient, respectively. It can be also viewed that a high level of agreement is yielded by the feature selection algorithms for the deterioration factors of cathodic protection and relative humidity. In this respect, seven out of the eight feature selection algorithms appended relative humidity as the eleventh most important deterioration factor, and six out of the eight feature selection algorithms selected cathodic protection as the fourteenth most important deterioration factor.

Figure 3 displays the correlation between the eight feature selection algorithms. It can be observed that the highest correlation coefficient (0.956) lies between ReliefF and NCA while the lowest absolute correlation coefficient (0.011) is exhibited between the regression tree and sensitivity analysis as well as between sensitivity analysis and Pearson correlation coefficient. Besides, it can be inferred that sensitivity analysis despite its wide use in the literature maintains low levels of correlation with filter-based and embedded-based feature selection algorithms. For instance, a negative correlation of $-0.2967$ exists between the pairs of (sensitivity analysis, ReliefF) and (sensitivity analysis, neighborhood component analysis). Furthermore, the correlation coefficients between the pairs (sensitivity analysis, Pearson correlation coefficient) and (sensitivity analysis, maximum information coefficient) are $-0.011$ and $-0.033$, respectively. This can be explained by the fact that sensitivity analysis manages to measure the influence of each factor independently without being able to capture the interdependencies between the deterioration factors. It is also observed that moderate levels of correlations lie between the rankings obtained from the pairs (regression tree, Pearson correlation) and (regression tree, boosted regression tree) with correlation coefficients of 0.67 and 0.604, respectively. The correlation matrix manifested that both bagged regression trees and boosted regression trees obtained high correlation with the reminder of embedded-based methods than the filter-based feature selection methods. To blend the rankings reported by the studied feature selection algorithms, an ensemble reciprocal ranking method is applied to provide the final rankings of the fourteen factors. As shown in Figure 4, the most influential five factors in order of importance are age, diameter, traffic, soil corrosivity, and material. On the contrary, the least influential five factors are cathodic protection, air temperature, rainfall, length, and relative humidity.

The acquired condition indices are clustered in this section using the GEPFCM algorithm to develop a condition rating scale. Figure 5 illustrates the existence of four clusters: excellent, very good, good, and poor. These clusters have x-coordinates and y-coordinates of (360.0, 4.7), (62.2, 5.5), (191.4, 6.8), and (332.1, 8.2), respectively. The midway through the cluster centers is leveraged to determine the threshold values of each rating class. A comprehensive description and recommended intervention action are added by analyzing the age and failure rate of each rating category shown in Table 11. The methodology of
computing the burst and leakage rates is described as follows: The dataset is divided based on three perspectives: water source (i.e., freshwater and saltwater), failure type (i.e., leak and burst), and material type (i.e., metallic and plastic). The dataset is then divided based on the diameter/size into small, medium, and large-sized pipes. The number of failures is computed for each category and divided by the total length of pipes to get the normalized number of failures per unit length. The outcome is then divided by the length of time over which failures have occurred (i.e., 12 years) to get the number of failures per unit length per year. The resultant failures/km/year are assigned to the critical category, and the values are acquired for the very good, good, and poor categories accordingly by assuming that the excellent category experiences no leaks and bursts.

Source(s): Authors own work
All the possible combinations of the different characteristics associated with the fourteen factors are constructed. Each factor has two or four characteristics, as previously stated. For instance, the pipe material factor comprises four classes: 1) ductile iron (DI), steel (S), and stainless steel (SS), 2) polyethylene (PE) and polyvinyl chloride (PVC), 3) cast iron (CI) and asbestos cement (AC), and 4) lined galvanized iron (GIL) and unlined galvanized iron (GI). Meanwhile, considering the pipe length factor, the pipe is considered either long (>100 meters), medium (30–100 meters), or short (<30 meters). When all the factors' characteristics are permuted, the number of possible combinations is 1259712. Therefore, deterioration curves are created for the common and worst-case scenarios representing the upper and lower boundaries, respectively. It is worth noting that the common scenario refers to small diameter, metallic material, short length, moderate corrosive soil, no cathodic protection, no failure history, intermediate water pressure, moderate rainfall, low temperature, low humidity, average traffic, footway road type, and urban land use. Besides, pipes having small diameters, other than plastic and metallic materials, long lengths, highly corrosive soil type, no cathodic protection, failure history, high water pressure, heavy rainfall, high temperature, high humidity, heavy traffic, carriageway road type, and sea/urban land use represent the worst base-case scenario. The developed deterioration curves calculate the anticipated condition index for each scenario to establish a database that could be utilized once a pipe’s characteristics are determined.

---

**Source(s):** Authors own work

---

**Figure 3.** Kendall tau rank correlation matrix among the investigated feature selection algorithms
Figure 6 displays the deterioration curves illustrating the change in material, soil corrosivity, water pressure, cathodic protection, failure history, and road type with pipe age while the remainder of the factors is considered constant based on the assessed scenarios. In each figure, there are three solid and dashed curves plotted representing the change in characteristics for the common and worst scenarios, respectively. The curves make it obvious that pipes installed in highly corrosive soils are more susceptible to deterioration compared to medium and low-corrosive soils. Moreover, CI, AC, GI, and GIL pipes have inferior conditions over their lifetimes than DI, S, and SS as well as PE and PVC material types. This can be attributed to the reduced breakage rate and resilience to corrosion of plastic pipes. It is evident that high-pressure pipes deteriorate faster than intermediate-pressure pipes and normal-pressure pipes. It can be also observed that pipes with cathodic protection maintain a notably better condition than pipes lacking cathodic protection. It is also noticed that water pipes with previous failure history are more vulnerable to significant degradation in their condition than the ones that did not experience previous failures of leaks or bursts. Moreover, it can be viewed that water pipes under carriage roads deteriorate slightly more rapidly than pipes under footways.

Numerous models may be developed as a function of the pipe age using the deterioration curves in Figure 6. This is accomplished by determining a correlation between the pipe age
and its condition index. As a result, the 12 degradation curves are utilized to build the deterioration models. Examples of the deterioration models used to represent the condition indices of different pipe materials for the common and worst scenarios are given below, where $t$ and $CI$ stand for the pipe’s age and condition index, respectively. A sample of the deterioration models is presented in Table 12. The accuracy and robustness of the degradation models are ensured by achieving an $R^2$ of 99.96%. Developing a database, which includes all combinations of the various input factors as well as the anticipated condition (output) is essential for establishing condition assessment models and capturing the deterioration patterns of water pipes. It is important to note that none of the repair, rehabilitation, or maintenance operations that often occur are included in the deterioration curves and models.

$$CI(t)_{DL-SS[C]} = 3.9259E - 8t^4 - 9.4602E - 6t^3 + 0.0012t^2 - 0.1107t + 9.9581$$
$$CI(t)_{DL-SS[W]} = 4.4194E - 8t^4 - 1.109E - 5t^3 + 0.0014t^2 - 0.1374t + 9.9486$$
$$CI(t)_{PE-PVC[C]} = 3.9116E - 8t^4 - 9.391E - 5t^3 + 0.0012t^2 - 0.1092t + 9.9586$$
$$CI(t)_{PE-PVC[W]} = 4.5034E - 8t^4 - 1.1159E - 5t^3 + 0.0014t^2 - 0.1363t + 9.9517$$
$$CI(t)_{CI-AC-GI-GIL[C]} = 4.1213E - 8t^4 - 9.9963E - 6t^3 + 0.0012t^2 - 0.1169t + 9.9601$$
$$CI(t)_{CI-AC-GI-GIL[W]} = 4.6231E - 8t^4 - 1.1588E - 5t^3 + 0.0015t^2 - 0.1437t + 9.9505$$

The developed deterioration models are verified and validated utilizing real data from existing water pipes in Hong Kong districts. The dataset comprises the factors influencing the deterioration of water pipes alongside the actual condition index of water pipes. The correlation coefficient ($R$), mean absolute error (MAE), and root mean squared error (RMSE) are shown in Equations (39-41). These metrics are characterized by their simplicity and accuracy in demonstrating the performance of the proposed model in forecasting pipe deterioration.
<table>
<thead>
<tr>
<th>Numeric scale</th>
<th>Linguistic scale</th>
<th>Description</th>
<th>Proposed action</th>
</tr>
</thead>
<tbody>
<tr>
<td>8–10</td>
<td>Excellent</td>
<td>Pipes in this category are newly installed within 7 years</td>
<td>No action is required</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metallic pipes:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small-size: Pipes in this category experience 0.002 bursts/km/year and/or 1 leak/km every 24 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium-size: Pipes in this category experience 0.006 bursts/km/year and/or 1 leak/km every 28 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Large-size: Pipes in this category experience 0.006 bursts/km/year and/or 1 leak/km every 32 years</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-metallic pipes:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small-size: Pipes in this category experience 0.0025 bursts/km/year and/or 1 leak/km every 8 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium-size: Pipes in this category experience 1 leak/km every 24 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Large-size: Pipes in this category experience 1 leak/km every 28 years.</td>
<td></td>
</tr>
<tr>
<td>6–8</td>
<td>Very good</td>
<td>Pipes in this category are installed with an average age of 7–20 years</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metallic pipes:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small-size: Pipes in this category experience 0.002 bursts/km/year and/or 1 leak/km every 20 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium-size: Pipes in this category experience 0.005 bursts/km/year and/or 1 leak/km every 28 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Large-size: Pipes in this category experience 0.006 bursts/km/year and/or 1 leak/km every 32 years</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-metallic pipes:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small-size: Pipes in this category experience 0.0025 bursts/km/year and/or 1 leak/km every 8 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium-size: Pipes in this category experience 1 leak/km every 28 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Large-size: Pipes in this category experience 1 leak/km every 32 years.</td>
<td></td>
</tr>
<tr>
<td>5–6</td>
<td>Good</td>
<td>Pipes in this category are installed with an average age of 20–35 years</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metallic pipes:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small-size: Pipes in this category experience 0.004 bursts/km/year and/or 1 leak/km every 12 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium-size: Pipes in this category experience 0.01 bursts/km/year and/or 1 leak/km every 14 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Large-size: Pipes in this category experience 0.015 bursts/km/year and/or 1 leak/km every 16 years</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-metallic pipes:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small-size: Pipes in this category experience 0.005 bursts/km/year and/or 1 leak/km every 4 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium-size: Pipes in this category experience 1 leak/km every 26 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Large-size: Pipes in this category experience 1 leak/km every 28 years.</td>
<td></td>
</tr>
<tr>
<td>3–5</td>
<td>Poor</td>
<td>Pipes in this category are installed with an average age of 35–50 years</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metallic pipes:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small-size: Pipes in this category experience 0.006 bursts/km/year and/or 1 leak/km every 8 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium-size: Pipes in this category experience 0.015 bursts/km/year and/or 1 leak/km every 10 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Large-size: Pipes in this category experience 0.019 bursts/km/year and/or 1 leak/km every 11 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-metallic pipes:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small-size: Pipes in this category experience 0.0075 bursts/km/year and/or 1 leak/km every 3 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium-size: Pipes in this category experience 1 leak/km every 18 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Large-size: Pipes in this category experience 1 leak/km every 19 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metallic pipes:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small-size: Pipes in this category experience 0.006 bursts/km/year and/or 1 leak/km every 8 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium-size: Pipes in this category experience 0.015 bursts/km/year and/or 1 leak/km every 10 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Large-size: Pipes in this category experience 0.019 bursts/km/year and/or 1 leak/km every 11 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-metallic pipes:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small-size: Pipes in this category experience 0.0075 bursts/km/year and/or 1 leak/km every 3 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium-size: Pipes in this category experience 1 leak/km every 18 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Large-size: Pipes in this category experience 1 leak/km every 19 years.</td>
<td></td>
</tr>
</tbody>
</table>

It is recommended to apply a pressure monitoring system to monitor the pressure drop in the water pipes, if any.

It is recommended to install sensors for a period of one week to detect leakage. The suggested intervention action is a spot or minor repair if a leak is detected. Apply the burst/leak prediction models to check their possibilities.

It is recommended to install sensors for a period of one month to detect leakage. The suggested intervention action is lining and/or cathodic protection (for metallic pipes) if leak(s) are detected. Apply the burst/leak prediction models to check their possibilities.

(continued)
Critical Pipes in this category are installed with an average age of more than 50 years.

*Metallic pipes:*
- Small-size: Pipes in this category experience 0.008 bursts/km/year and/or 1 leak/km every 6 years.
- Medium-size: Pipes in this category experience 0.02 bursts/km/year and/or 1 leak/km every 7 years.
- Large-size: Pipes in this category experience 0.025 bursts/km/year and/or 1 leak/km every 8 years.

*Non-metallic pipes:*
- Small-size: Pipes in this category experience 0.01 bursts/km/year and/or 1 leak/km every 2 years.
- Medium-size: Pipes in this category experience 1 leak/km every 13 years.
- Large-size: Pipes in this category experience 1 leak/km every 14 years.

It is recommended to permanently install sensors to detect leakage. The suggested intervention action is to plan for a replacement in the near future. Apply the burst/leak prediction models to check their possibilities and support the replacement action.

**Source(s):** Authors' own work
deterioration. The first metric measures the strength of a linear relationship between two variables, whereas the other two metrics represent the prediction error between the forecasted and actual conditions. Higher correlation coefficients with values closer to 1 as well as lower values of error metrics closer to 0 indicate the robustness and impreciseness of the models. For instance, the first developed model is validated, yielding R, MAE, and RMSE.
<table>
<thead>
<tr>
<th>ID</th>
<th>Diameter</th>
<th>Material</th>
<th>Length</th>
<th>Corrosivity</th>
<th>Cathodic protection</th>
<th>Failure occurrence</th>
<th>Pipe characteristics</th>
<th>Land use</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Small</td>
<td>DI-SS</td>
<td>Short</td>
<td>Moderate</td>
<td>No</td>
<td>No</td>
<td>Intermediate</td>
<td>Low</td>
<td>Average Footway Urban</td>
</tr>
<tr>
<td>2</td>
<td>Small</td>
<td>PE-PVC</td>
<td>Short</td>
<td>Moderate</td>
<td>No</td>
<td>No</td>
<td>Intermediate</td>
<td>Low</td>
<td>Average Footway Urban</td>
</tr>
<tr>
<td>3</td>
<td>Small</td>
<td>CI-AC-GI-GIL</td>
<td>Short</td>
<td>Moderate</td>
<td>No</td>
<td>No</td>
<td>Intermediate</td>
<td>Low</td>
<td>Average Footway Urban</td>
</tr>
<tr>
<td>4</td>
<td>Small</td>
<td>DI-SS</td>
<td>Long</td>
<td>Highly</td>
<td>No</td>
<td>Yes</td>
<td>High</td>
<td>High</td>
<td>Heavy Carriageway Sea/Urban</td>
</tr>
<tr>
<td>5</td>
<td>Small</td>
<td>PE-PVC</td>
<td>Long</td>
<td>Highly</td>
<td>No</td>
<td>Yes</td>
<td>High</td>
<td>High</td>
<td>Heavy Carriageway Sea/Urban</td>
</tr>
<tr>
<td>6</td>
<td>Small</td>
<td>CI-AC-GI-GIL</td>
<td>Long</td>
<td>Highly</td>
<td>No</td>
<td>Yes</td>
<td>High</td>
<td>High</td>
<td>Heavy Carriageway Sea/Urban</td>
</tr>
<tr>
<td>7</td>
<td>Medium</td>
<td>DL-S-SS</td>
<td>Short</td>
<td>Moderate</td>
<td>No</td>
<td>No</td>
<td>Intermediate</td>
<td>Low</td>
<td>Average Footway Urban</td>
</tr>
<tr>
<td>8</td>
<td>Large</td>
<td>DL-S-SS</td>
<td>Short</td>
<td>Moderate</td>
<td>No</td>
<td>No</td>
<td>Intermediate</td>
<td>Low</td>
<td>Average Footway Urban</td>
</tr>
<tr>
<td>9</td>
<td>Medium</td>
<td>DL-S-SS</td>
<td>Long</td>
<td>Highly</td>
<td>No</td>
<td>Yes</td>
<td>High</td>
<td>High</td>
<td>Heavy Carriageway Sea/Urban</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>ID</th>
<th>Diameter</th>
<th>Material</th>
<th>Length</th>
<th>Corrosivity</th>
<th>Cathodic protection</th>
<th>Failure occurrence</th>
<th>Water pressure</th>
<th>Rainfall</th>
<th>Temperature</th>
<th>Humidity</th>
<th>Traffic</th>
<th>Road type</th>
<th>Land use</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Large</td>
<td>DI-S-SS</td>
<td>Long</td>
<td>Highly corrosive</td>
<td>No</td>
<td>Yes</td>
<td>High</td>
<td>Heavy</td>
<td>High</td>
<td>High</td>
<td>Heavy</td>
<td>Carriageway</td>
<td>Sea/Urban</td>
<td>CI = 4.4586 E-8t^4–1.10947 E–5t^3 + 0.0014 t^2–0.1313 t+9.9524</td>
</tr>
<tr>
<td>11</td>
<td>Medium</td>
<td>PE-UPVC</td>
<td>Short</td>
<td>Moderate corrosive</td>
<td>No</td>
<td>No</td>
<td>Intermediate</td>
<td>Moderate</td>
<td>Low</td>
<td>Low</td>
<td>Average</td>
<td>Footway</td>
<td>Urban</td>
<td>CI = 3.8522 E-8t^4–9.2158 E–6t^3 + 0.0011 t^2–0.1064 t+9.9618</td>
</tr>
<tr>
<td>12</td>
<td>Large</td>
<td>PE-UPVC</td>
<td>Short</td>
<td>Moderate corrosive</td>
<td>No</td>
<td>No</td>
<td>Intermediate</td>
<td>Moderate</td>
<td>Low</td>
<td>Low</td>
<td>Average</td>
<td>Footway</td>
<td>Urban</td>
<td>CI = 4.4103 E-8t^4–1.0913 E–5t^3 + 0.0014 t^2–0.1333 t+9.9526</td>
</tr>
<tr>
<td>13</td>
<td>Medium</td>
<td>PE-UPVC</td>
<td>Long</td>
<td>Highly corrosive</td>
<td>Yes</td>
<td>Yes</td>
<td>High</td>
<td>Heavy</td>
<td>High</td>
<td>High</td>
<td>Heavy</td>
<td>Carriageway</td>
<td>Sea/Urban</td>
<td>CI = 4.3895 E-8t^4–1.0815 E–5t^3 + 0.0014 t^2–0.1313 t+9.9551</td>
</tr>
<tr>
<td>14</td>
<td>Large</td>
<td>PE-UPVC</td>
<td>Long</td>
<td>Highly corrosive</td>
<td>No</td>
<td>Yes</td>
<td>High</td>
<td>Heavy</td>
<td>High</td>
<td>High</td>
<td>Heavy</td>
<td>Carriageway</td>
<td>Sea/Urban</td>
<td>CI = 4.1255 E-8t^4–1.0009 E–5t^3 + 0.0012 t^2–0.1171 t+9.9584</td>
</tr>
<tr>
<td>15</td>
<td>Small</td>
<td>DI-S-SS</td>
<td>Short</td>
<td>Highly corrosive</td>
<td>No</td>
<td>No</td>
<td>Intermediate</td>
<td>Moderate</td>
<td>Low</td>
<td>Low</td>
<td>Average</td>
<td>Footway</td>
<td>Urban</td>
<td>Source(s): Authors' own work</td>
</tr>
</tbody>
</table>
values of 0.8, 1.33, and 1.39, respectively. These outcomes provide a strong indication of the model’s capability to reliably forecast the water pipe condition index.

\[
R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}
\]  

(39)

\[
MAE = \frac{\sum_{i=1}^{n} |x_i - y_i|}{n}
\]  

(40)

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2}
\]  

(41)

where \(x_i\) and \(y_i\) represent the actual and predicted condition indices, respectively. Meanwhile, \(\bar{x}\) and \(\bar{y}\) represent the average actual and predicted condition indices, respectively.

As mentioned earlier, a computer-aided application is developed to ease the implementation of the developed condition assessment and deterioration prediction model by the users. The automated tool comprises two modules, namely “pairwise comparison-based” and “user’s weight-based”. In the former module, the user is asked to fill in the pairwise comparisons for the deterioration categories and deterioration factors. In the latter module, the user is asked to identify the relative importance of deterioration factors based on his/her intuition and expertise. As for the first module, the interface of the developed weight interpretation model for pipe-related factors is depicted in Figure 7a. In it, the user is required to specify the importance levels of pipe-related factors against each other using a drop-down list. Then, by clicking the “calculate” button, two types of output are appended. The first is the local relative importance weights of pipe-related factors. The second type of output encompasses maximum eigenvalue, consistency ratio, and a message box popping up indicating whether pairwise comparison entries are consistent or need revision.

The interface of the developed condition assessment model is illustrated in Figure 7b. The automated tool allows the user to determine the features of the designated water pipe through either a drop-down list or a text box control. The outcome of the condition assessment model encapsulates the condition rating of the water pipe, respective condition category, and recommended intervention action. The interface of the developed deterioration prediction model is presented in Figure 7c. The user is asked first to identify the designated pipe features by the user. The output of the model constitutes the condition ratings of water piper over time in common and worst-case scenarios, which are displayed in a data-grid view. In addition, the deterioration curves of the aforementioned scenarios are illustrated.

7. Conclusions
This research paper houses the development of integrated condition assessment and deterioration models for saltwater pipes in Hong Kong. Fourteen different factors were found to implicate the condition of water pipes, and they were categorized into pipe-related, soil-related, operational, and environmental categories. The priority weights of these factors were determined using the SFAHP algorithm. MARCOS algorithm was then exploited to merge the relative importance weights of deterioration factors along with their effect values into an integrated condition rating. The GEPPFCM algorithm was harnessed to design a condition rating system for water pipes. Eight filter-based and embedded-based feature selection algorithms were deployed to scrutinize and prioritize the deterioration factors. Fourth-order regression functions were compiled to project the deterioration patterns in pipe condition over time meanwhile accounting for different characteristics of factor combinations and the Saltwater pipes in Hong Kong
accompanying anticipated conditions. A computer program written in C#.net language was designed to ease the implementation of the condition assessment and deterioration models.

Analysis of questionnaire surveys evidenced that the environmental-related (30.13%) category had the highest relative importance followed by pipe-related (26.15%), soil-related (22.48%), and operational (21.24%). In addition, it was found that water pressure, soil corrosivity, cathodic protection, land use, and failure history were determined to be the top five most important factors with 12.03%, 11.4%, 11.08%, 9.53%, and 9.21%, respectively. Feature selection results substantiated that sensitivity analysis fails to render a reliable ranking of deterioration factors based on their implication on pipe condition. Additionally, the developed ensemble feature selection model appended age, diameter, traffic, soil

![Figure 7.](image)

User interface of the developed (a) weight interpretation model for pipe-related factors, (b) condition assessment model, and (c) deterioration prediction model

Source(s): Authors own work
corrosivity, and material as the top five influential factors. It was also observed that a high level of correlation lies between (ReliefF, NCA) while low levels of correlation existed between (sensitivity analysis, ReliefF) and (sensitivity analysis, NCA). The deterioration model was trained and tested using 160,315 saltwater pipes across Hong Kong. Performance analysis showed that they attained correlation coefficient, mean absolute error and root mean squared error of 0.8, 1.33, and 1.39, respectively. Municipalities should be able to accurately plan for future maintenance and rehabilitation activities using the developed models. This research study can be extended in the future to accommodate other factors influencing pipe deterioration to develop a more comprehensive model. These studies could consider real-time sensor data such as material wear and fatigue, corrosion rates, and water quality parameters to enhance the predictive accuracy of deterioration models. Future research efforts may be also directed towards anticipating the probabilities of occurrence leakage and bursts of saltwater pipes, and analyzing the implicating factors on their occurrences.

References


**Corresponding author**
Eslam Mohammed Abdelkader can be contacted at: eslam.abdelkader@polyu.edu.hk

For instructions on how to order reprints of this article, please visit our website: [www.emeraldgrouppublishing.com/licensing/reprints.htm](http://www.emeraldgrouppublishing.com/licensing/reprints.htm)
Or contact us for further details: permissions@emeraldinsight.com