

Topology-informed survival modelling for predicting future pipe failures in water distribution networks

Downloaded from: https://research.chalmers.se, 2026-11-24 03:57 UTC

Citation for the original published paper (version of record):

Sangroula, U., Kristiansson, E., Bergstedt, O. et al (2027). Topology-informed survival modelling for predicting future pipe failures in water distribution networks. Water Research, 309(Part A). http://dx.doi.org/10.1016/j.watres.2026.127187

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Water Research 309 (2027) 127187

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Topology-informed survival modelling for predicting future pipe failures in water distribution networks

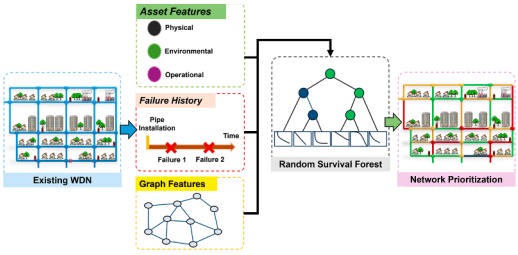
Uchit Sangroula [a,*], Erik Kristiansson [b], Olof Bergstedt [a,c], Thomas J.R. Pettersson [a]

[a] Department of Architecture and Civil Engineering, Water Environment Technology, Chalmers University of Technology, Gothenburg, SE-412 96, Sweden
[b] Department of Mathematical Sciences, Chalmers University of Technology & University of Gothenburg, Gothenburg, Sweden
[c] Sustainable Waste and Water, City of Gothenburg, Gothenburg, Sweden

HIGHLIGHTS

- Graph features convert network topology into pipe failure model inputs.
- Temporal evaluation reflects practical utility decision-making.
- Network topology adds value beyond conventional pipe asset data.
- Predicted high-risk pipe groups capture most future failures.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Water distribution networks
Pipe failure prediction
Survival analysis
Recurrent failures
Network topology
Graph-based features

ABSTRACT

Pipe failure prediction models are increasingly used to support maintenance and renewal planning in water distribution networks (WDNs), yet most existing models treat pipes as independent assets. Hence, they do not sufficiently capture the influence of network structure on failure risk. Moreover, model evaluation is often based on randomly selected or retrospective data, which may not reflect utility decision-making where historical information is used to prioritize future interventions. Here, we present a topology-informed survival modelling framework for multi-year pipe failure risk prediction. Pipe failures are treated as recurrent events within a gap-time framework and modelled using random survival forests to estimate future failure risk. Network structure is represented through a pipe connectivity line graph, from which graph-based features are derived and combined with conventional pipe input features. The framework was evaluated using pipe asset and failure records from a Swedish water utility, and temporally consistent training and testing data were used to reflect real-world application. The survival outputs separated pipes into distinct risk groups and identified a small high-risk subset in which observed failures were concentrated. In addition, graph-based features provided information beyond conventional pipe attributes and improved the identification of high-risk pipes. The proposed framework provides a practical way to incorporate network topology into pipe failure prediction, supporting risk-based asset prioritization in WDNs.

* Corresponding author.
E-mail address: uchit@chalmers.se (U. Sangroula).

https://doi.org/10.1016/j.watres.2026.127187
Received 22 June 2026; Received in revised form 1 October 2026; Accepted 21 October 2026
Available online 21 October 2026
0043-1354/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Ageing water distribution networks (WDNs), increasing break rates, and constrained renewal budgets are driving water utilities to shift from reactive repairs toward proactive, risk-informed asset management (Delnaz et al., 2023; Malm et al., 2012). However, renewal capacity remains limited, with replacement rates varying across utilities, but commonly reported to be around 1% of the network per year (Daulat et al., 2024; Forero-Ortiz et al., 2026). Proper maintenance of such critical infrastructure is essential to minimize service disruption, reduce water losses, and protect public health (Marlow et al., 2013; Ramos-Salgado et al., 2022). Accordingly, utilities aim to prioritize a limited subset of pipes for inspection, rehabilitation, or replacement. Therefore, a systematic ranking of network assets using available information is necessary to identify which pipes should be prioritized for intervention over the planning horizon.

To support such prioritization, many studies have focused on predictive modelling of WDN pipe failure. Such failure prediction models are commonly categorized as physical, statistical, or machine learning (ML) approaches (Fan et al., 2022). Physical methods typically rely on mechanistic representations of structural deterioration and material behaviour (Makar, 2000; Rajani and Kleiner, 2001). Statistical approaches include regression-based formulations and survival analysis for hazard estimation and replacement planning (Barton et al., 2022a; Kleiner and Rajani, 2001; Scheidegger et al., 2015; Scholten et al., 2014). In recent years, ML methods have become increasingly prevalent for pipe failure prediction, offering improved predictive capacity and the ability to handle complex nonlinear relationships (Chen et al., 2025; Forero-Ortiz et al., 2023). Applications include random forests (Raspati et al., 2022), ensemble methods (Taiwo et al., 2025a), support vector machines (Robles-Velasco et al., 2020), and artificial neural networks (Taiwo et al., 2025b). Recent studies have also combined survival analysis with ML, including random survival forests (RSF) (Snider and McBean, 2021) and deep survival learning approaches (Wang et al., 2025), enabling longer-term time-to-failure and survival probability estimations.

In these models, factors influencing pipe failure are commonly grouped into physical, operational, and environmental factors (Barton et al., 2019), which form the basis of model inputs used for failure prediction. Physical factors include pipe characteristics such as age, diameter, material, and length. Operational factors include service conditions such as pressure and flow, and environmental factors include external conditions such as soil type, climate, and surrounding ground conditions.

Beyond these factors, growing attention has been directed toward spatial dependence and network structure (Abokifa and Sela, 2019). Pipe failures are not spatially independent and often cluster in localized hotspots associated with shared environmental and operational conditions (Wéber et al., 2020). To account for such spatial effects, some studies have incorporated spatial and system-level characteristics into failure modelling. Barton et al. (2022b) utilized a grouping strategy by aggregating pipes into segments with similar material, diameter, and age characteristics, thereby representing shared local pipe conditions in the failure model. Other studies have applied spatial clustering approaches to identify regions with high pipe failure rates to incorporate geographic proximity into prediction models (Aslani et al., 2021; Chen and Guikema, 2020). More recently, graph-based metrics and network representation learning have been explored to capture structural connectivity and neighbourhood dependence within WDNs (Rukmi et al., 2023; Yu et al., 2024). Notably, several studies have employed graph neural networks (GNNs) to model topology-aware pipe failure prediction within supervised classification frameworks (Liang et al., 2021; Xu and He, 2024).

Despite these advances, important methodological and practical challenges remain in operationalizing predictive models as reliable decision-support tools for utilities. Many studies evaluate predictive performance under artificial settings, e.g. by using randomly selected or retrospective evaluation data (Taiwo et al., 2025b; Winkler et al., 2018). This differ substantially from real-world deployment, where utilities must use information available at the time of decision-making to prioritize pipes for intervention over a future planning horizon. Consequently, results from such evaluations will not reflect true prospective performance and are likely to overlook important temporal changes in failure risk.

Furthermore, pipe failure prediction is typically formulated as a classification problem within a fixed time window (Robles-Velasco et al., 2020; Xu and He, 2024). While this may be suitable for short-term prediction, it requires redefining the target variable (pipe failure) for each planning horizon. In reality, pipe failures often occur as recurrent processes: assets may experience multiple breaks over their lifetime, with possible changes in failure risk after repair (Tscheikner-Gratl et al., 2016). Hence, modelling approaches that account for recurrent failures may better represent the temporal nature of pipe failure.

Finally, despite increasing attention to spatial dependence and network structure, a key challenge remains in how these relationships are represented in pipe failure modelling. Cluster-based and hotspot approaches typically represent spatial effects using predefined geographic zones or simple distance measures, which may not fully reflect the connectivity structure of WDNs. As a result, local pipe interactions are often approximated using spatial distances rather than actual network topology. However, pipes that are geographically close may not necessarily be connected in the network, while connected pipes may share operational and failure-related dependencies that are not captured by such distance-based representations. Collectively, these limitations highlight a gap between current modelling practices and the operational realities faced by utilities.

To address these knowledge gaps, we develop and evaluate a topology-informed prediction framework to rank WDN pipes by their future failure risk. We structure the framework to reflect the operational question faced by utilities: using all information available up to a defined decision year, which pipes should be prioritized for intervention over the next planning horizon? Accordingly, the modelling objective is to estimate failure risk over a future horizon using information available up to a decision year and to evaluate whether the model provides a useful ranking of pipes for decision support.

We formulate pipe failure within a gap-time survival formulation (Lin et al., 1999) and implement the model using RSF (Ishwaran et al., 2008). This allows repeated failures of the same pipe to be represented over time and enables failure probability to be estimated over future multi-year horizons. Additionally, we incorporate network topology by deriving graph-based features from the WDN connectivity, describing the local network context and historical failures. When applied to pipe asset data from a Swedish water utility, the proposed topology-informed RSF framework provided useful failure-risk ranking, achieving a mean holdout C-index of 0.89 and capturing 81% of actual failures among the top 20% of pipes predicted to be at highest risk over a five-year prediction window.

2. Methodology

This section describes the graph representation and modelling process utilized in the study. An overview of the framework of the study is presented in Fig. 1.

2.1. Graph representation of a WDN

A WDN consists of physical components such as reservoirs, tanks, junctions, pipes, pumps, and valves. To capture the structural and functional relationships between these components, the network can be represented as a graph. Mathematically, a WDN is defined as $G = (V,E)$,

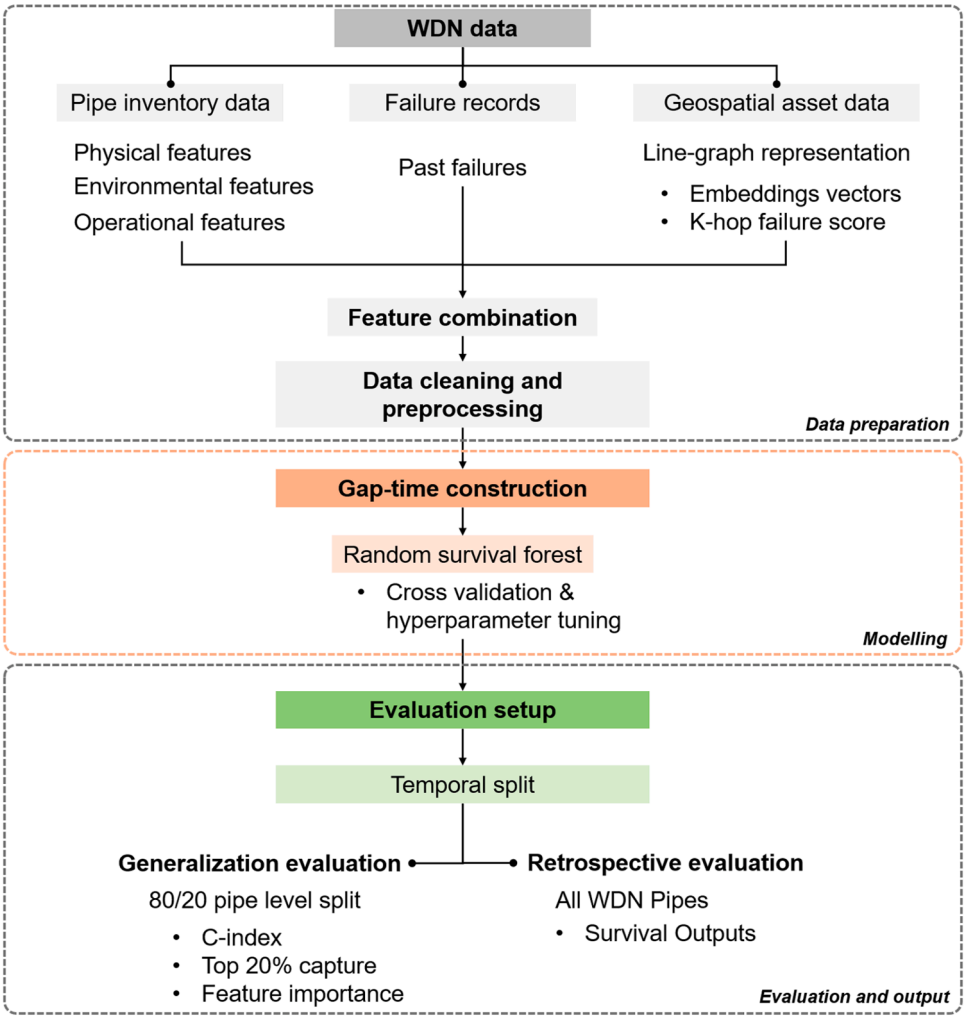


Fig. 1. Study framework summarizing data preparation, gap-time survival modelling with RSF, and evaluation through temporal pipe-level generalization and retrospective WDN-wide analysis.

where $V = \{v_1, v_2, \ldots, v_n\}$ is a set of nodes and $E = \{(v_i, v_j) | v_i, v_j \in V\}$ is a set of edges. Nodes represent elements, such as junctions, tanks, or reservoirs, while edges correspond to pipes, pumps, or valves connecting these elements. The connectivity of the network can be expressed through an adjacency matrix A of size $n \times n$ where n is the number of nodes. Each entry a_{ij} in A represents the connection between node i and node j. In its simplest form, $a_{ij} = 1$ if a link exists between nodes i and j, and $a_{ij} = 0$ otherwise. This binary representation reflects only the network topology and is commonly referred to as an unweighted graph. For a weighted graph, the entries a_{ij} can be assigned values that capture physical properties of the connecting elements such as pipe diameter or material (Zhou et al., 2025).

2.1.1. Line-graph transformation

While the node-edge representation provides a natural description of the connectivity in a WDN, many applications in failure modelling require analysis at the pipe level. To enable analysis at the pipe level, we transformed the original graph G into a line graph $L(G) = (V_L, E_L)$. In the line graph $L(G)$, each node corresponds to one pipe in the original WDN. Two nodes in a line graph are connected if their corresponding pipes share a common junction.

This transformation provides a direct way to associate pipe-specific attributes (e.g., diameter, material, installation year, etc.) with nodes of the line graph. Moreover, this enables the computation of features that reflect the relationships between pipes. For example, the degree of a node in a line graph corresponds to the number of pipes that are directly connected via a junction in the WDN.

2.2. Feature generation from line-graph

Next, we constructed graph-based features that describe the position and relationships of each pipe within the network. Two complementary feature types were derived: (i) unsupervised graph embeddings, which provide vector representations of the local network topology, and (ii) a graph scalar feature, which summarizes the historical failure environment in the local neighbourhood of each pipe.

2.2.1. Unsupervised graph embeddings

Graph embeddings provide a way to numerically represent each pipe's network context in the WDN. Embeddings map each node in the line graph to a feature vector. Here, embedding is a function $f: V_L \rightarrow \mathbb{R}^d$ where V_L is the set of nodes in the line graph and d is the embedding dimension.

The mapping f was learned in two stages. First, the topology of the line graph was sampled using random walks to generate training sequences of nodes (each node corresponding to a pipe in the original WDN). The sampling approach follows the node2vec framework (Grover and Leskovec, 2016), which performs second-order biased random walks. In the second stage, the sampled walk sequences were used to learn the embedding vectors using the skip-gram training approach

(Word2Vec) as adopted in node2vec (Grover and Leskovec, 2016; Mikolov et al., 2013).

This procedure produced one embedding vector $z_u \in \mathbb{R}^d$ for each node, where d is the embedding dimension. Nodes that repeatedly occurred close to each other in the sampled sequences were treated as having similar network context and were therefore encouraged to have similar feature vectors. This enabled the embeddings to capture higher-order structural patterns beyond immediate neighbours (Qiu et al., 2018). Stacking the vectors for all nodes formed the embedding matrix $Z \in \mathbb{R}^{|V_L| \times d}$, which was used as input to the predictive model. Fig. 2 summarizes the workflow for graph-based feature generation from the WDN.

2.2.2. Scalar feature (k-hop failure score)

A scalar graph feature was computed to summarize the historical failure conditions around each pipe. Since nodes in the line graph correspond to pipes in the original WDN, neighbouring nodes represent pipes connected through shared junctions.

A k-hop failure score was computed by aggregating historical failures within the local neighbourhood of each pipe. Failures among directly connected pipes (1-hop neighbours) were fully counted, while failures among second-hop neighbours were down-weighted to reflect weaker topological influence, as defined in Eq. (1):

$$S_{k=2}(u) = \sum_{a \in N_1(u)} I(a) + \lambda \sum_{b \in N_2(u)} I(b) \quad (1)$$

where $u \in V_L$ is a node in the line graph, $N_1(u)$ and $N_2(u)$ denote the sets of 1-hop and 2-hop neighbours of u, respectively. $I(a)$ and $I(b)$ are indicator variables equal to 1 if the corresponding neighbour pipe has experienced at least one recorded failure, and 0 otherwise. For this study, $k = 2$ and λ was set to 0.5 for 2-hop neighbours. Higher values of $S_{k=2}(u)$ indicate that a pipe is surrounded by historically failed pipes. Sensitivity analyses for alternative k and λ are provided in Supplementary S1.3.

2.3. Predictive model description

2.3.1. Survival analysis

Survival analysis models time-to-event outcomes, where observations that do not experience the event within the observation period are treated as right-censored. The survival function represents the probability that an individual remains event-free beyond a given time, while the cumulative event probability represents the probability that the event occurs by that time (Klein and Moeschberger, 2005).

2.3.2. Gap-time formulation for recurrent pipe failures

In this study, the event of interest is a pipe failure. Since water distribution pipes can experience multiple recurrent failures over their service life, historical records may contain multiple events for the same pipe (Le Gat, 2014). Recurrent failures were, therefore, modelled using a gap-time time scale (Lin et al., 1999; Soutinho and Meira-Machado, 2023), where the time-to-event was measured either from the previous recorded failure/repair or, for pipes with no prior recorded failures, from the start of observation. Each pipe's failure history was expanded into a sequence of gap-time intervals, where each interval represents one at-risk period between two failures or between the last failure and the end of observation. As a result, a pipe with n recorded failures was transformed into $n + 1$ intervals: one interval before the first failure, $n - 1$ intervals between the n consecutive failures, and a final right-censored interval.

This interval-based representation increases the number of observations available for model training, since pipe failures are rare relative to the total number of pipes in the dataset. It also enables the use of failure-history features such as the pipe's age, number of previous failures and years since last failure at interval start, which are relevant because earlier failures may influence the likelihood of future failures (Tscheikner-Gratl et al., 2016). Such interval expansion has been

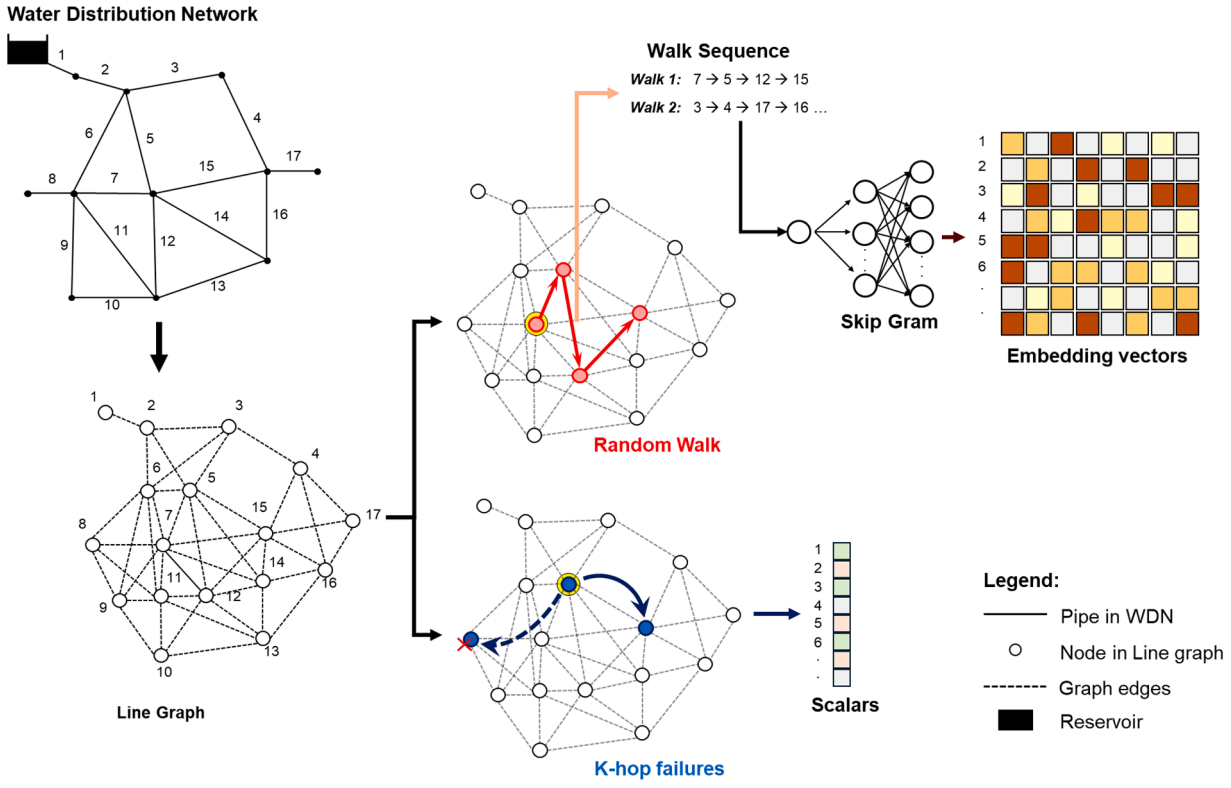


Fig. 2. Workflow for graph feature generation from the WDN. The original WDN is converted into a line graph. Random walks are used to capture network connectivity and learn embedding vectors, while a neighbourhood-based scalar feature (k-hop failure score) summarizes local failure history.

successfully applied in prior pipe-failure prediction studies (Daulat et al., 2024; Winkler et al., 2018). Fig. 3 illustrates the gap-time interval construction used for model training.

To prevent pipes with multiple recurrent intervals from contributing disproportionately to model training, interval-level weights were applied. For pipe i with m_i gap-time intervals, the weight assigned to interval j was calculated as Eq. (2):

$$w_{ij} = \frac{1}{m_i} \tag{2}$$

where m_i is the total number of gap-time intervals generated for pipe i.

2.3.3. Random survival forest

RSF, used as the main prediction model used in this study, is an ensemble ML method that extends the random forest algorithm (Breiman, 2001) to time-to-event outcomes, where observations may be right-censored. We trained the RSF on the gap-time intervals described in Section 2.3.2, where each interval constituted one training instance with a duration and event indicator. For prediction, the RSF used aggregated bootstrap survival trees to estimate the ensemble cumulative hazard function, from which the corresponding survival function was derived (Ishwaran et al., 2008). Tree construction considered random subsets of input variables at each node, and splits were selected using the log-rank criterion (Mantel, 1966). Model discrimination was evaluated using Harrell's concordance index (C-index), which was calculated for the predicted risk scores while accounting for right-censored observations (Harrell et al., 1996). A C-index of 1 indicates perfect agreement between predicted risk ordering and observed failure ordering, whereas a value of 0.5 corresponds to random ranking.

2.4. Modelling framework implementation

2.4.1. Data description and preprocessing

This study uses pipe inventory and failure datasets obtained from a Swedish water utility. The dataset consists of pipes currently in service, together with recorded pipe failure events. Each record is treated as an observed failure-repair-record event, meaning that it corresponds to a physical pipe failure event that was repaired and subsequently logged in the dataset. Only operational pipes are included in the dataset. Pipes that were decommissioned or replaced are not present. Therefore, the dataset represents only pipes that remained in service throughout the observation period. Each pipe is represented as a unique segment, defined as a continuous section of pipe for which the installation year, material, and diameter remain constant.

The dataset comprised 585 km of pipelines with diameters ranging from 20 mm to 1000 mm. Historical failure records from 2007–2024 indicated an overall failure rate of 11.8 failures/100 km/year. Fig. 4 summarizes the distribution of installed pipe length and recorded failures across installation decades and material classes, showing that recorded failures are concentrated mainly in cast iron pipes installed during the 1950s and 1960s, whereas newer polyethylene pipes represent a substantial share of installed length but show fewer recorded failures.

2.4.2. Explanatory variables from the utility dataset

The utility dataset provides explanatory variables (model input features) that describe pipe characteristics, environmental conditions, and operational context (Table 1). In this study, the input features consisted of two categories: (i) static pipe and location variables, which remain constant for a given pipe segment, and (ii) failure-history variables, which evolve over time and are updated at the start of each risk interval under the gap-time formulation (Section 2.3.2).

2.4.3. Graph-based features from the utility dataset

A GIS representation of the WDN was used to derive graph-based features. Graph embeddings were generated using biased random walks and skip-gram training, producing a 64-dimensional representation for each pipe (Section 2.2).

Because the survival model was trained under the gap-time formulation described in Section 2.3.2, graph-based features were assigned according to the network structure available at the start of each interval. Since the WDN evolves over time as utilities expand, separate embeddings were generated for each yearly network configuration to ensure that the graph-based features reflect the network structure available at that time. The k-hop failure score was similarly calculated using only the failure history available at the start of each interval.

Embeddings trained independently for different years are not directly comparable because the embedding space may rotate or flip between runs. Yearly embedding matrices were therefore aligned to the 2019 reference space using an orthogonal Procrustes transformation (Schönemann, 1966). For each yearly snapshot, pipes shared with the 2019 reference snapshot were used as anchor pipes. The transformation estimates an orthogonal matrix (a rotation/reflection matrix) that preserves distances and angles within the embedding space. This matrix was then applied to the yearly embeddings.

Node2vec requires training hyperparameters, including the return and in-out parameters (p, q), walk length, number of walks per node, and embedding dimension. The full embedding configuration is

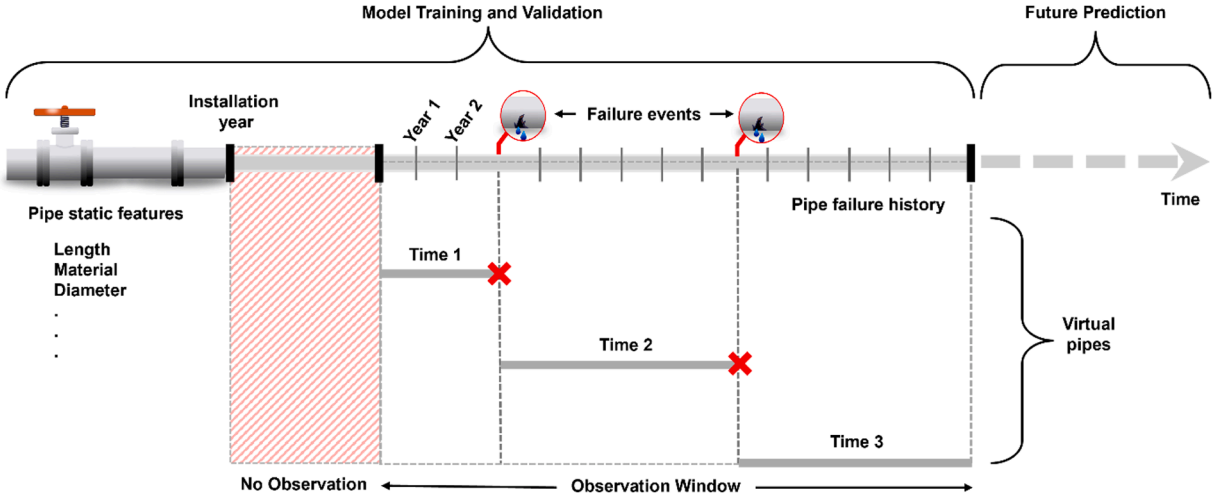


Fig. 3. Gap-time interval construction for recurrent pipe failures. Each pipe's lifetime is divided into intervals between successive failure events, with time reset after each failure. The final interval is censored if no additional failure is observed.

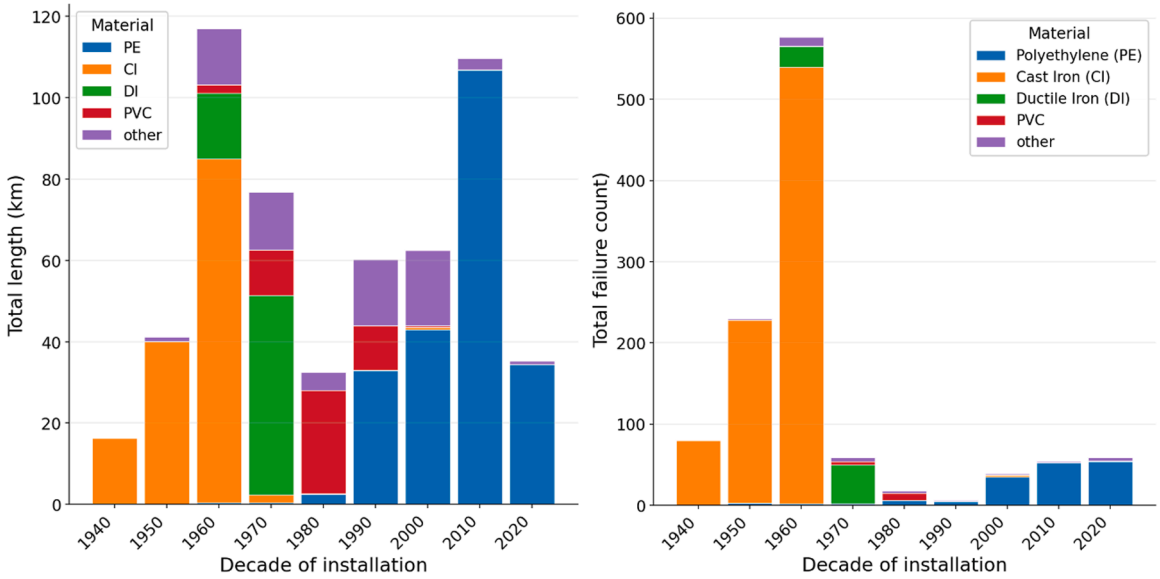


Fig. 4. Distribution of pipe assets by installation decade and material: (a) Total installed pipe length (km), (b) total recorded failure count aggregated by installation decade.

Table 1
Summary of explanatory variables from the utility dataset.

Explanatory variable	Variable type	Description
Length	Numerical	Pipe length (m)
Diameter	Numerical	Pipe diameter (mm)
Soil	Categorical	Dominant soil type at pipe location
Material	Categorical	Pipe material class
Elevation	Numerical	Ground elevation at pipe location (m)
Land use	Categorical	Dominant land-use class at pipe location
Road intersections	Categorical	Indicator of whether pipe is located near road intersection
Near railways	Categorical	Indicator of whether pipe is located near a railway line
GWT	Numerical	Groundwater table depth at pipe location (m)
Service connections	Numerical	Number of service connections linked to pipe segment
Fire hydrants	Numerical	Number of fire hydrants connected to pipe segment
Outlets	Numerical	Number of outlets connected to pipe segment
Air release valves	Numerical	Number of air release valves connected to pipe segment
Age start	Numerical	Pipe age at the start of the modelling interval (years)
No. previous failures	Numerical	Total previous recorded failures at interval start
Years since last failure	Numerical	Number of years since last recorded failure

provided in Table S1 in Supplementary S1.1. In this study, embeddings were generated using a local-context setting ($p = 1$, $q = 2$, 40 walks per node, walk length = 10, embedding dimension = 64). This configuration emphasizes near-neighbour relationships, consistent with the node2vec interpretation that $q > 1$ promotes more local graph exploration, while keeping the embedding dimension moderate. These choices were motivated by the tendency of pipe failures to exhibit localized clustering, as reported in literature (Chen and Guikema, 2020), and by the possibility that neighbouring pipes may share similar operational and environmental exposure. Accordingly, the graph embeddings were used as unsupervised model input features to encode structural neighbourhood context, while the RSF determined whether this information provided additional predictive value beyond the conventional pipe attributes. The embedding dimension was selected based on cross-validation performance in a sensitivity analysis of dimensions 8–256 (Supplementary S1.2).

2.4.4. Data cleaning and feature encoding

Pipes with missing core inventory attributes (installation year, diameter, or length) were removed prior to modelling. Inconsistent observations (e.g., failures recorded before installation) and extreme outliers were also filtered. Similar data cleaning steps are commonly used in water-pipe failure prediction studies (Snider and McBean, 2020). For modelling, categorical variables (e.g., material, soil, and land-use indicators) were encoded using one-hot representations (Bishop and Nasrabadi, 2006), while numerical variables were standard-scaled for consistent preprocessing, although scaling is not required for RSF models.

2.5. Model evaluation and practical prediction

2.5.1. Model performance evaluation setup

Model performance was evaluated using a temporal prediction setup, in which the RSF was trained on recurrent-event gap-time intervals and tested using a fixed prediction start year $t_0 = 2020$. To avoid data leakage, the training, validation, and test data were split at the pipe level before interval construction, so that all intervals from the same pipe remained in the same subset. The dataset was divided into an 80% training and validation set and a 20% holdout test set. The holdout test set was used to evaluate model performance using the C-index (Fig. 5). The C-index was calculated using the standard scalar RSF risk score, defined as the sum of the ensemble cumulative hazard function over the distinct event times in the training data. This scalar risk score is independent of each pipe's observed event or censoring time. Although the RSF was trained using recurrent-event gap-time intervals, the final prediction was performed per pipe by constructing one forward-looking episode starting at t_0 for each test pipe. The fitted RSF produced a yearly pipe-specific survival curve from 2020 onward, which was used to derive the five-year failure risk for 2020–2024.

Three feature configurations were evaluated: (i) baseline RSF, using only utility-derived input features; (ii) combined RSF, using both baseline features and graph-based features; and (iii) topology-only RSF, using only graph-based features. Each configuration was trained and evaluated using 20 random seeds, and the performance was reported as the mean across runs. Differences between the baseline and combined RSF were assessed using paired runs with identical random seeds and

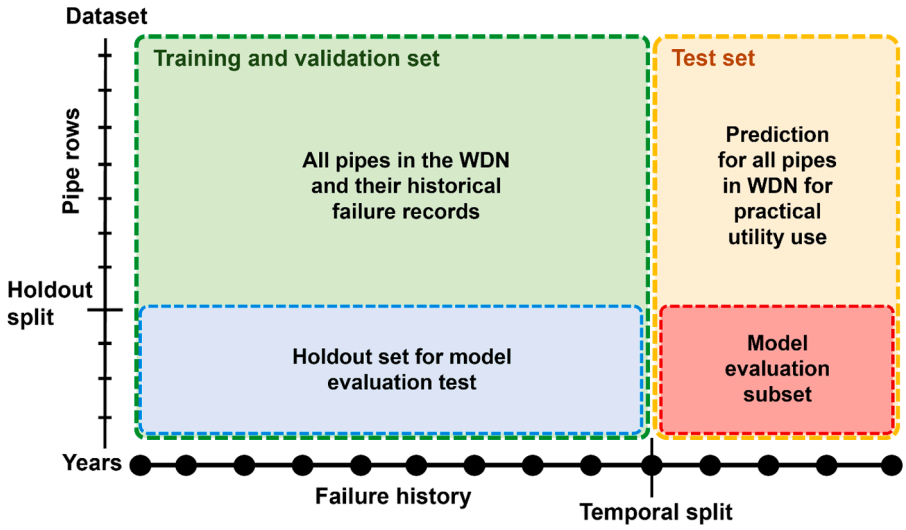


Fig. 5. Data splitting strategy used for model evaluation and practical utility application. Evaluation used a pipe-level 80/20 holdout split with temporal prediction from t_0=2020, while practical predictions were generated for all pipes.

pipe-level splits. Mean paired differences were reported with 95% bootstrap confidence intervals, and Wilcoxon signed-rank tests were used as a complementary significance test at $\alpha = 0.05$. Furthermore, to support interpretation of the RSF results, permutation-based variable importance was computed using the implementation provided in scikit-survival (Pölsterl, 2020). This analysis highlights the features that most strongly influence RSF predictions and supports comparison of influential drivers across feature sets.

In addition to the pipe-level holdout, a geographical holdout evaluation was conducted to assess model generalization to spatially distinct parts of the WDN. The network was divided into geographical districts, and a different district was randomly selected as the holdout area in each run. Pipes within the selected district were excluded from model training and used exclusively for testing. The RSF configurations were evaluated using the same temporal prediction period as the main holdout evaluation. To avoid spatial information leakage, graph-based feature construction was performed without using information from the geographical test district during model training.

2.5.2. Hyperparameter tuning

Hyperparameter tuning was performed within the training-validation set using 10-fold group cross-validation. For each candidate hyperparameter combination, the C-index was calculated on the validation fold using the same scalar RSF risk score and then averaged across the 10 folds. The complete modelling workflow was repeated across 20 random seeds, with each run including cross-validation-based tuning, final model training, and holdout evaluation. Table 2 summarizes the hyperparameter search ranges used for tuning. After tuning, the selected hyperparameters were used to train the final model for each run.

2.5.3. Additional model comparisons

To examine whether the predictive contribution of the topology-derived features was specific to the RSF model, the baseline and combined feature configurations were additionally evaluated using two alternative survival models: Cox proportional hazards (Cox PH) (Cox, 1972) and a Prentice–Williams–Peterson (PWP) (Prentice et al., 1981) gap-time model. The same pipe-level splits and evaluation period were used for direct comparison. Implementation details are provided in Supplementary S3.

Table 2
RSF hyperparameter search ranges used for tuning.

Hyperparameter	Search range
Number of trees	30–500
Maximum tree depth	3–20
Min. samples to split	10–80
Min. samples per leaf	5–40
Max features	sqrt, log2, 0.20, 0.35, 0.50
Bootstrap sampling	True

2.5.4. Full-network practical predictions

While the holdout test set was used to evaluate model performance on independent data, the operational objective was to support future pipe prioritization across the full WDN. Therefore, after model development and evaluation, the RSF was applied to all pipes in the WDN using the same temporal prediction setup (Fig. 5), with fixed prediction start year $t_0 = 2020$. For each pipe, one forward-looking prediction episode was constructed using utility-derived input features, graph-based features, and failure-history available up to 2019. The model then produced survival curves from 2020 onward for all pipes in the WDN.

3. Results

3.1. Prediction results and model interpretation

The three RSF configurations showed different levels of performance on the pipe-level holdout test set (Fig. 6). The baseline RSF achieved a mean C-index of 0.880 ± 0.016, while the combined RSF achieved 0.891 ± 0.016 (Fig. 6 (a)), corresponding to a mean improvement of 0.0105 (95% CI: 0.005–0.016). The combined RSF achieved a higher C-index in 16 of the 20 paired runs. A Wilcoxon signed-rank test additionally indicated a difference between the paired model performances ($p = 0.0017$). The topology-only RSF yielded a mean C-index of 0.746 ± 0.034, indicating that graph-based features contained a predictive signal. Full ablation results, including Brier scores, are provided in Supplementary S2.

In addition to ranking-based discrimination, performance was evaluated in a prioritization setting by examining the top 20% predicted highest-risk pipes (Table 3). Since evaluation was conducted on a 20% holdout test set, this corresponds to prioritizing about 4% of the full pipe population over the 5-year horizon (0.8% per year), comparable to typical utility intervention rates (often 1% of the network per year). Under this setup, the baseline RSF captured 74.8 ± 4.5% of observed failures, while the combined RSF captured 80.6 ± 5.4%, a mean improvement of 5.8 percentage points (95% CI: 4.3–7.3). The combined

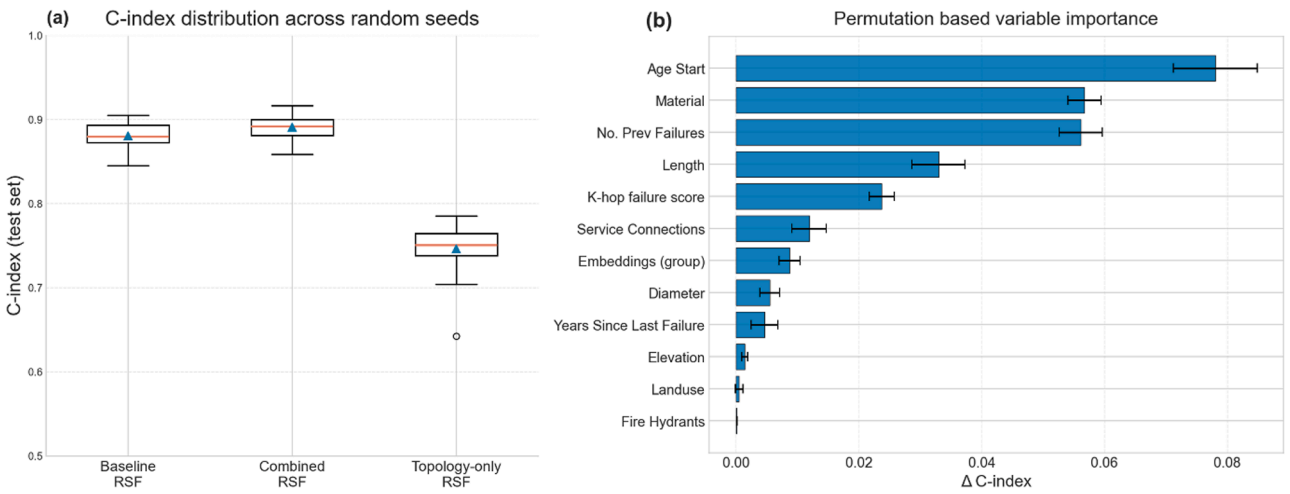


Fig. 6. Model performance and feature contribution on the 20% pipe-level holdout test set. In (a), the C-index distribution across multiple runs is shown; the y-axis is truncated and starts at 0.5, corresponding to random performance. In (b), the permutation-based variable importance for the Combined RSF model is shown.

Table 3
Predictive performance under random pipe-level and geographical holdout evaluation.

Evaluation setup	Model configuration	C-index, mean ± SD	Top 20% failure capture, mean ± SD
Random pipe-level holdout	Baseline RSF	0.880 ± 0.016	0.748 ± 0.045
	Combined RSF	0.891 ± 0.016	0.806 ± 0.054
	Topology-only RSF	0.746 ± 0.034	0.539 ± 0.064
Geographical holdout	Baseline RSF	0.866 ± 0.018	0.687 ± 0.029
	Combined RSF	0.872 ± 0.022	0.753 ± 0.058
	Topology-only RSF	0.709 ± 0.028	0.502 ± 0.025

model achieved higher top 20% failure capture in 18 of the 20 paired runs and equal performance in the remaining two. A Wilcoxon signed-rank test additionally indicated a difference between the paired results ($p < 0.001$). Thus, for the same inspection or renewal capacity, the combined model identified a larger share of pipes that actually failed within the test window.

The relative advantage of the combined RSF was also retained under geographical holdout evaluation. The baseline RSF achieved a mean C-index of 0.866 ± 0.018, compared with 0.872 ± 0.022 for the combined RSF. Under the geographical holdout, top 20% failure capture similarly increased from 68.7 ± 2.9% for the baseline RSF to 75.3 ± 5.8% for the combined RSF.

Additional survival-model comparisons showed a similar pattern for Cox PH and PWP. For Cox PH, the mean C-index increased from 0.869 ± 0.020 for the baseline configuration to 0.874 ± 0.018 for the combined configuration, while for PWP it increased from 0.861 ± 0.017 to 0.865 ± 0.015. Although the magnitude of improvement varied between models, the combined topology-informed configuration achieved a higher mean C-index for both models.

Fig. 6 (b) presents permutation importance computed on the holdout test set, measured as the reduction in C-index (ΔC-index) when a feature (or feature group) is randomly permuted while keeping all other variables unchanged. Larger reductions therefore indicate greater importance for model ranking performance. The results indicate that age at interval start, the number of previous failures, and pipe material are the most influential predictors in the RSF model. This aligns with commonly reported findings in pipe deterioration modelling, where pipe age and material are frequently identified as key drivers of failure risk (Taiwo et al., 2023), and previous failures are associated with an increased likelihood of future failures (Fan et al., 2022).

Notably, the graph-based features provided an incremental but meaningful contribution to the model's ranking performance. The *k*-hop failure score showed measurable importance, indicating that failure risk is topologically clustered in the network and that local exposure to historically failed pipes adds information beyond pipe-level inventory and failure-history features. The embeddings group also contributes, suggesting that embeddings capture additional neighbourhood context and provide predictive value when used alongside baseline predictors. Overall, the graph-based features added incremental value by representing neighbourhood structure and local failure clustering effects that are typically absent from purely tabular utility datasets.

3.2. *Practical survival outputs for future prioritization*

The Combined RSF model was used to generate survival curves for all pipes in the WDN under a temporal prediction setup, where the model was trained on features and historical information up to 2019 and then applied to forecast future failure probabilities. To compare the predictions to observed outcomes, Kaplan–Meier (KM) estimates were overlaid for the first 5 years, corresponding to the observed evaluation window (2020–2024).

In Fig. 7 (a) pipes were grouped into three strata based on their predicted 5-year failure probability at the forecast origin (2020): low risk (≤70th percentile), medium risk (70th–90th percentile), and high risk (>90th percentile), using cut points derived from the 2017–2019 validation window. The predicted survival curves show a clear separation between risk groups, with the high-risk group exhibiting substantially faster decline in survival probability than the medium and low-risk groups. Similar separation is observed when survival curves are grouped by material (Fig. 7 (b)), location (neighbourhoods) (Fig. 7 (c)) and installation period (Fig. 7 (d)).

The separation observed in the survival curves indicated that pipe failures were concentrated within a relatively small subset of high-risk pipes. The highest-risk group pipes experienced a failure rate of approximately 29.5%, more than an order of magnitude higher than the lowest-risk group, which had a failure rate of only 0.8%. Overall, these survival outputs translate the model predictions into directly interpretable survival curves that can support utility planning. Within the observed five-year window, the KM overlays follow the same risk-group ordering as the predicted curves, supporting the model's risk stratification and showing that the predicted separation is consistent with the

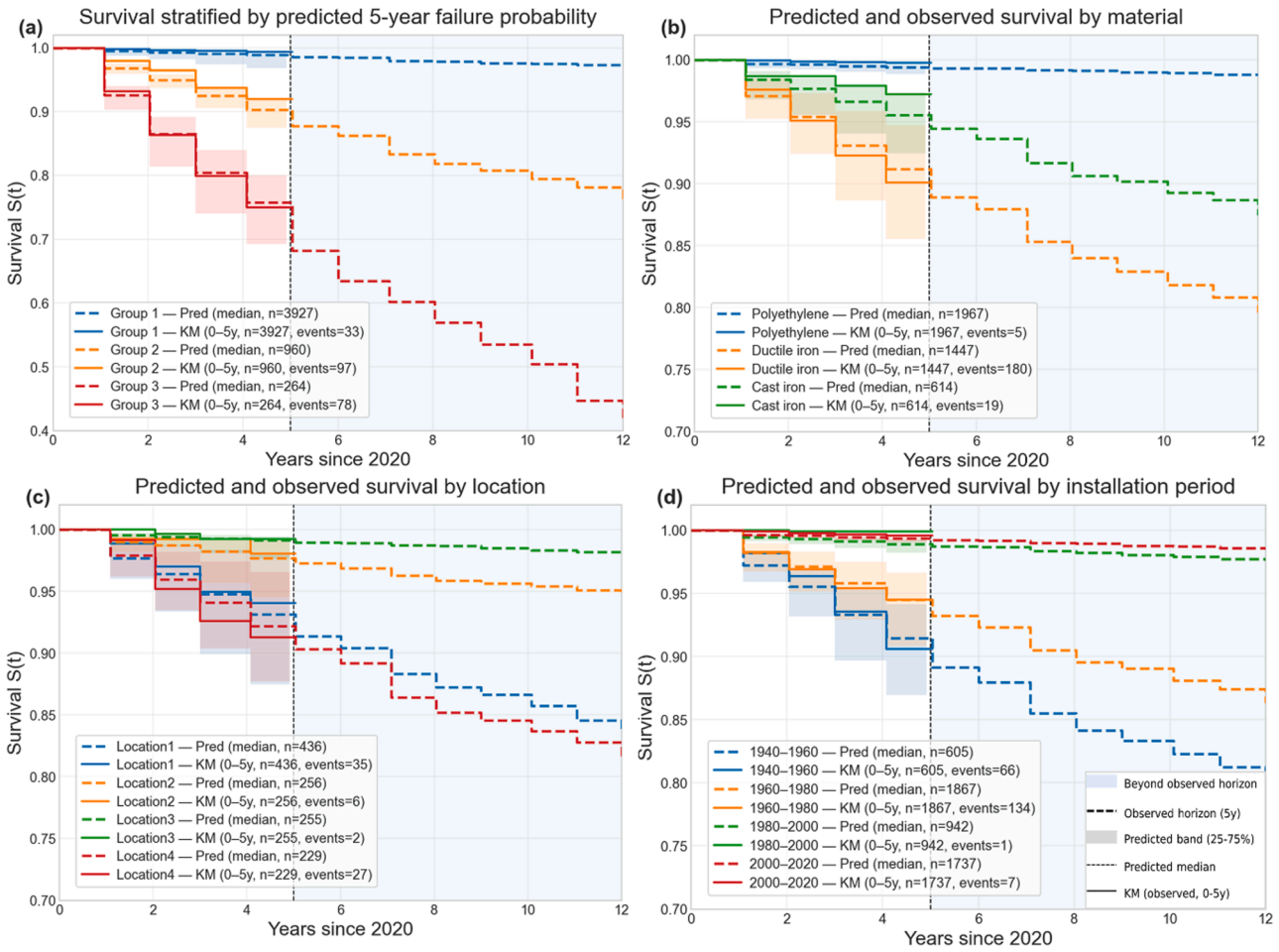


Fig. 7. RSF survival curves (years since 2020), shown as cohort-level summaries stratified by: (a) predicted 5-year failure probability, (b) pipe material, (c) location, and (d) installation-year groups. Dashed lines represent the median predicted survival from the model, with shaded bands indicating the interquartile range (25th-75th percentiles). Solid lines show the observed Kaplan-Meier (KM) survival for the 0-5year period (2020-2024). The vertical dashed line marks the end of the observed horizon.

observed survival patterns. Five-year calibration indicated close agreement between predicted and observed risks, with some overestimation in intermediate-risk groups (Supplementary S6.3). Practical prioritization of the full network was further assessed using precision and failure recall across different intervention capacities (Supplementary S6.2). The results showed that selecting approximately the top 5% of the network over the five-year horizon would have captured almost 40% of observed failures, with a precision of approximately 30%.

4. Discussions

This study demonstrates a topology-informed recurrent survival modelling framework for predicting future pipe failures in WDNs. The framework combines a gap-time formulation for recurrent failures with RSF to estimate multi-year failure risk over future planning horizons, while incorporating WDN topology through graph-based features. The results indicate that features derived from network topology can provide operationally meaningful value when integrated within a pipe failure prediction framework. Moreover, the framework identified pipe groups, locations, materials, and installation cohorts associated with elevated failure risk. This risk stratification was reflected in observed outcomes, where the highest-risk group had a substantially higher failure rate than the lowest-risk group. These findings indicate that a relatively small subset of pipes accounted for a larger share of failures, supporting the practical relevance of risk-ranked outputs for inspection prioritization and renewal planning.

RSF was adopted because it can accommodate high-dimensional input features while capturing nonlinear relationships between pipe characteristics and failure risk. The modelling objective was to identify pipes with elevated future failure risk within the planning horizon to support utility prioritization rather than to estimate the effects of individual input features. Classical recurrent-event regression models, including Andersen–Gill, PWP, and shared frailty formulations, are well established for inferential analysis of recurrent events (Amorim and Cai, 2015). In contrast, the gap-time RSF framework used in this study was selected as a prediction-oriented approach for estimating horizon-based failure risk and ranking pipes for renewal prioritization.

This framework relies on several methodological and data-structural assumptions. Because each pipe contributes multiple inter-failure intervals, observations from the same pipe may not be fully independent. This dependence was partly represented through interval-level input variables, e.g. the number of previous failures, the pipe age at the start of each interval and years since last failure allowing the model to account for the fact that pipes with previous failures may have different subsequent failure risk. The gap-time formulation further assumes that, after each recorded failure, the pipe is repaired and re-enters service, and the risk of the next failure is evaluated from that time onward. To avoid data leakage, all intervals generated from the same pipe were assigned to the same training, validation, or test subset. In addition, pipe-level interval weights were applied during model training to prevent pipes with many failures from dominating the fitted model, although within-pipe dependence and unobserved frailty are not modelled explicitly. Right-

censoring occurred mainly because the observation window ended in 2024; pipes with no further recorded failure before this date were treated as censored at the end of observation. Pipes that were fully replaced or decommissioned were not present in the dataset, i.e., the analysis reflects only the active pipe inventory rather than all historically installed pipes. As a result, the estimated failure risks are conditional on pipes that remained in service during the observation period. This may introduce selection bias if highly deteriorated or failure-prone pipes were previously removed, potentially underrepresenting the highest-risk failure histories. However, since the modelling objective is to estimate failure risk for the active network, the available inventory remains relevant to the prediction task.

Historical failure records in the dataset were available from 1997 onward. For older pipes, failures occurring prior to 1997 were not observed, resulting in incomplete early failure histories, i.e., left truncation. Removing all pipes installed before 1997 would have resulted in substantial data loss and may have introduced bias. Therefore, following a similar approach to Daulat et al. (2024), pipes installed before 1940 were excluded to reduce the influence of the oldest pipes, for which incomplete prior failure histories were expected to be most pronounced. This approach may partially reduce, but does not eliminate, the effects of left truncation. Sensitivity analyses using alternative installation-year cutoffs of 1950 and 1960 showed only modest changes in model performance (Supplementary S5). The excluded pre-1940 pipes accounted for approximately 3% of the raw pipe inventory, corresponding to 18.5 km of network length, suggesting that their exclusion likely had only a minor impact on our results. Furthermore, the available failure records did not consistently distinguish between different failure types or mechanisms. Therefore, all recorded failure-repair events were treated as a single failure outcome. This may limit interpretation because different failure types can arise from different mechanisms and risk factors. More detailed failure categorization could therefore improve both model interpretation and predictive performance.

To ensure that the model outputs capture realistic deployment conditions, graph-based features were computed using only information available at the start of each gap-time interval. Separate embeddings were generated for each yearly network configuration, while the *k*-hop failure score was calculated using the failure history available at the start of each interval. This ensured that future network information did not influence past predictions and allowed the input feature set to reflect the temporal evolution of the WDN. A practical advantage of this approach is that the graph-based features can be recomputed as new network and failure data become available, while remaining separate from the prediction model itself. This allows topology-derived information to be updated and incorporated into existing modelling workflows without reformulating the full predictive model. Unlike spatial representations based primarily on geographic proximity, the graph-based features used here define neighbourhood relationships directly from WDN connectivity, with the *k*-hop failure score providing an explicit measure of local failure clustering among connected pipes. Recent advances in graph-based learning, including GNNs, provide more direct ways to learn from network structure and pipe-level attributes. In this study, topology was incorporated through pre-computed graph-based features rather than end-to-end graph learning, which offered a transparent and model-compatible way to evaluate the added value of network information within a recurrent survival framework. Future work could extend this direction by comparing pre-computed graph-based features from network topology with GNN-based approaches, including graph models adapted for survival or recurrent-event prediction, particularly where richer temporal, hydraulic, or operational data are available.

The predictive contribution of individual graph-based features may depend on how network and spatial context is represented, as further examined through comparisons with simpler descriptors based on line-graph degree, adjacent-pipe failures, and Euclidean-neighbourhood failures (Supplementary S4). Furthermore, as the *k*-hop failure score is based on a weighted sum of failed neighbouring pipes, its magnitude may partly reflect local neighbourhood size. A neighbourhood-normalized *k*-hop score was therefore evaluated, and comparable predictive performance was observed after normalization for neighbourhood size (Supplementary S1.3). Although the embedding configuration adopted in this study was motivated by considerations described in Section 2.4.3, exploratory runs indicated that representations emphasizing local structural context provided the most consistent predictive improvements. Embedding configurations favouring neighbourhood-focused random walks generally performed better than those encouraging extensive global network exploration. This may reflect the tendency of pipe failures to cluster locally within distribution networks, where connected or nearby pipes can share similar environmental and operational conditions. Since features were updated at yearly intervals, network structure was assumed to remain unchanged within each year, meaning structural modifications occurring during an interval were not dynamically represented. However, because large-scale structural changes in WDNs are typically infrequent, this approximation was considered reasonable for the prediction setting.

Predictive performance reported in pipe failure modelling studies varies substantially across the literature (Barton et al., 2022a). WDNs differ in network structure, failure mechanisms, recording practices, observation windows, and available features, all of which can influence both model performance and interpretation of the results (Rifaai et al., 2022). Performance also depends on the prediction task and evaluation design. Artificial settings, such as randomly selected or retrospective evaluation data, may give more optimistic results than temporally separated evaluations. However, temporal prediction may be more relevant for utilities, and recent work has therefore emphasized temporally separated evaluation to better approximate real-world deployment conditions (Forero-Ortiz et al., 2026; Salehinasab and Goshtasbi Rad, 2025). Furthermore, predictive performance is often strongly influenced by data quality, feature design, and the representation of failure history rather than by model choice alone.

Finally, the contribution of graph-based features may depend on how network connectivity is represented and on the extent to which failures cluster within a given WDN. Where utility-derived input features already capture much of the failure signal, the additional improvement from graph-based features may be limited. Feature importance should therefore be interpreted relative to the available input features, modelling strategy, and underlying spatial structure of failures. As with most predictive modelling approaches, the resulting probabilities are estimates subject to uncertainty and should be interpreted as decision-support indicators rather than precise forecasts of individual pipe failures. Nevertheless, because utilities typically renew only a small fraction of the network each year, even moderate improvements in risk ranking is operationally meaningful. The framework presented here provides a practical approach that utilities and researchers can adapt to their own network settings for topology-informed pipe failure prioritization.

5. Conclusion

This study demonstrates a practical approach for topology-informed pipe failure modelling in WDNs. Graph-based features were integrated with conventional pipe input features within a recurrent gap-time survival modelling framework. The gap-time formulation models recurrent pipe failures over time and produces horizon-based failure probabilities that can support inspection prioritization and renewal planning. By using pre-computed network embeddings and graph features, the framework incorporates structural information from the WDN while remaining compatible with conventional prediction models and typical utility data environments. The results show that these graph-based features can add predictive value, particularly for risk ranking and prioritization, by capturing local network context and historical failures among connected pipes.

As the present evaluation is based on a single Swedish WDN, future

work could evaluate the proposed framework on larger, more heterogeneous, and independent WDNs to assess its generalizability and better understand when graph-based features provide the strongest predictive contribution. Incorporating additional hydraulic or operational variables may also help clarify how network topology interacts with operational stresses and different failure mechanisms. Future work could also extend the model comparison to deep survival approaches, such as DeepSurv and DeepHit, to further assess the performance of alternative survival modelling frameworks for pipe failure prediction. Further research could also investigate additional graph-based features or embedding strategies to capture other structural characteristics of the network. The *k*-hop failure score could also be further explored by incorporating failure frequency and recency weighting of neighbouring failures. Improved categorization of failures caused by pipe deterioration versus externally triggered events could further enhance predictive performance and interpretation. Additionally, historical replacement and decommissioning records could be incorporated to account for asset removals and reduce potential survivorship bias.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent to publish

Not applicable.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used Microsoft Copilot in order to improve readability and language. The authors reviewed and edited the output as needed and take full responsibility for the content of the published article.

Funding

This work was financially supported by the Swedish Research Council for Sustainable Development FORMAS project AquaClim: Building a Better Climate with Water Research (project number 2022-01900).

CRediT authorship contribution statement

Uchit Sangroula: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Erik Kristiansson:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Olof Bergstedt:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Thomas J.R. Pettersson:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful to the water utility staff at the municipality for providing information and data for the modelling.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2026.127187.

Data availability

The datasets obtained and analysed during this study are not publicly available. However, all codes, together with a synthetic WDN dataset, are available at: https://doi.org/10.5281/zenodo.22145264.

References

Abokifa, A.A., Sela, L., 2019. Identification of spatial patterns in water distribution pipe failure data using spatial autocorrelation analysis. J. Water. Resour. Plan. Manage. 145, 04019057. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001135.

Amorim, L.D., Cai, J., 2015. Modelling recurrent events: a tutorial for analysis in epidemiology. Int. J. Epidemiol. 44, 324–333. https://doi.org/10.1093/ije/dyu222.

Aslani, B., Mohebbi, S., Axthelm, H., 2021. Predictive analytics for water main breaks using spatiotemporal data. Urban. Water. J. 18, 433–448. https://doi.org/10.1080/1573062X.2021.1893363.

Barton, N.A., Farewell, T.S., Hallett, S.H., Acland, T.F., 2019. Improving pipe failure predictions: factors affecting pipe failure in drinking water networks. Water. Res. 164, 114926. https://doi.org/10.1016/j.watres.2019.114926.

Barton, N.A., Hallett, S.H., Jude, S.R., Tran, T.H., 2022a. An evolution of statistical pipe failure models for drinking water networks: a targeted review. Water Supply 22, 3784–3813. https://doi.org/10.2166/ws.2022.019.

Barton, N.A., Hallett, S.H., Jude, S.R., Tran, T.H., 2022b. Predicting the risk of pipe failure using gradient boosted decision trees and weighted risk analysis. NPJ. Clean. Water. 5, 22. https://doi.org/10.1038/s41545-022-00165-2.

Bishop, C.M., Nasrabadi, N.M., 2006. Pattern Recognition and Machine Learning. Springer.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Chen, R., Wang, Q., Javanmardi, A., 2025. A review of the application of machine learning for pipeline integrity predictive analysis in water distribution networks. Arch. Computat. Methods Eng. https://doi.org/10.1007/s11831-025-10251-6.

Chen, T.Y.-J., Guikema, S.D., 2020. Prediction of water main failures with the spatial clustering of breaks. Reliab. Eng. Syst. Saf. 203, 107108. https://doi.org/10.1016/j.ress.2020.107108.

Cox, D.R., 1972. Regression models and life-tables. J. R. Stat. Soc.: B (Methodol.) 34, 187–202.

Daulat, S., Rokstad, M.M., Bruaset, S., Langeveld, J., Tscheikner-Gratl, F., 2024. Evaluating the generalizability and transferability of water distribution deterioration models. Reliab. Eng. Syst. Saf. 241, 109611. https://doi.org/10.1016/j.ress.2023.109611.

Delnaz, A., Nasiri, F., Li, S.S., 2023. Asset management analytics for urban water mains: a literature review. Environ. Syst. Res. 12, 12. https://doi.org/10.1186/s40068-023-00287-7.

Fan, X., Wang, X., Zhang, X., Yu, X.(B.), 2022. Machine learning based water pipe failure prediction: the effects of engineering, geology, climate and socio-economic factors. Reliab. Eng. Syst. Saf. 219, 108185. https://doi.org/10.1016/j.ress.2021.108185.

Forero-Ortiz, E., Martinez-Gomariz, E., Sanchez-Juny, M., Cardus Gonzalez, J., Cucchietti, F., Baque Viader, F., Sarrias Monton, M., 2023. Models and explanatory variables in modelling failure for drinking water pipes to support asset management: a mixed literature review. Appl. Water. Sci. 13, 210. https://doi.org/10.1007/s13201-023-02013-1.

Forero-Ortiz, E., Sanchez-Juny, M., Martinez-Gomariz, E., Cardus Gonzalez, J., Cucchietti, F., Baque Viader, F., 2026. Near-future prediction of pipe failures in water supply networks: a key determinant for water pipe renewal policies through a machine learning approach. Appl. Water. Sci. 16, 70. https://doi.org/10.1007/s13201-025-02738-1.

Grover, A., Leskovec, J., 2016. node2vec: scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Presented at the KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA. ACM, pp. 855–864. https://doi.org/10.1145/2939672.2939754.

Harrell, F.E., Lee, K.L., Mark, D.B., 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat. Med. 15, 361–387. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4%3C361::AID-SIM168%3E3.0.CO;2-4.

Ishwaran, H., Kogalur, U.B., Blackstone, E.H., Lauer, M.S., 2008. Random survival forests. Ann. Appl. Stat. 2. https://doi.org/10.1214/08-AOAS169.

Klein, J.P., Moeschberger, M.L., 2005. Survival analysis: techniques for censored and truncated data. Corrected Third Printing Statistics for Biology and Health, Second edition. Springer, New York Berlin Heidelberg.

Kleiner, Y., Rajani, B., 2001. Comprehensive review of structural deterioration of water mains: statistical models. Urban Water Ground Water Environ. 3, 131–150. https://doi.org/10.1016/S1462-0758(01)00033-4.

Le Gat, Y., 2014. Extending the yule process to model recurrent pipe failures in water supply networks. Urban. Water. J. 11, 617–630. https://doi.org/10.1080/1573062X.2013.783088.

Liang, S., Li, Z., Liang, B., Ding, Y., Wang, Y., Chen, F., 2021. Failure prediction for large-scale water pipe networks using GNN and temporal failure series. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, CIKM '21. New York, NY, USA. Association for Computing Machinery, pp. 3955–3964. https://doi.org/10.1145/3459637.3481918.

Lin, D., Sun, W., Ying, Z., 1999. Nonparametric estimation of the gap time distribution for serial events with censored data. Biometrika 86, 59–70. https://doi.org/10.1093/biomet/86.1.59.

Makar, J.M., 2000. A preliminary analysis of failures in grey cast iron water pipes. Eng. Fail. Anal. 7, 43–53. https://doi.org/10.1016/S1350-6307(99)00005-9.

Malm, A., Ljunggren, O., Bergstedt, O., Pettersson, T.J.R., Morrison, G.M., 2012. Replacement predictions for drinking water networks through historical data. Water. Res. 46, 2149–2158. https://doi.org/10.1016/j.watres.2012.01.036.

Mantel, N., 1966. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother. Rep. 50, 163–170.

Marlow, D.R., Moglia, M., Cook, S., Beale, D.J., 2013. Towards sustainable urban water management: a critical reassessment. Water Res. Urban Water Manag. Increase Sustain. Cities 47, 7150–7161. https://doi.org/10.1016/j.watres.2013.07.046.

Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. https://doi.org/10.48550/arXiv.1301.3781.

Pölsterl, S., 2020. scikit-survival: a library for time-to-event analysis built on top of scikit-learn. J. Mach. Learn. Res. 21, 1–6.

Prentice, R.L., Williams, B.J., Peterson, A.V., 1981. On the regression analysis of multivariate failure time data. Biometrika 68, 373–379. https://doi.org/10.1093/biomet/68.2.373.

Qiu, J., Dong, Y., Ma, H., Li, J., Wang, K., Tang, J., 2018. Network embedding as matrix factorization: unifying DeepWalk, LINE, PTE, and node2vec. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. Presented at the WSDM 2018: The Eleventh ACM International Conference on Web Search and Data Mining. Marina Del Rey CA USA. ACM, pp. 459–467. https://doi.org/10.1145/3159652.3159706.

Rajani, B., Kleiner, Y., 2001. Comprehensive review of structural deterioration of water mains: physically based models. Urban Water Ground Water Environ. 3, 151–164. https://doi.org/10.1016/S1462-0758(01)00032-2.

Ramos-Salgado, C., Muñuzuri, J., Aparicio-Ruiz, P., Onieva, L., 2022. A comprehensive framework to efficiently plan short and long-term investments in water supply and sewer networks. Reliab. Eng. Syst. Saf. 219, 108248. https://doi.org/10.1016/j.ress.2021.108248.

Raspati, G.S., Bruaset, S., Bosco, C., Mushom, L., Johannessen, B., Ugarelli, R., 2022. A risk-based approach in rehabilitation of water distribution networks. Int. J. Environ. Res. Public Health 19, 1594. https://doi.org/10.3390/ijerph19031594.

Rifaai, T.M., Abokifa, A.A., Sela, L., 2022. Integrated approach for pipe failure prediction and condition scoring in water infrastructure systems. Reliab. Eng. Syst. Saf. 220, 108271. https://doi.org/10.1016/j.ress.2021.108271.

Robles-Velasco, A., Cortés, P., Muñuzuri, J., Onieva, L., 2020. Prediction of pipe failures in water supply networks using logistic regression and support vector classification. Reliab. Eng. Syst. Saf. 196, 106754. https://doi.org/10.1016/j.ress.2019.106754.

Rukmi, A.M., Shiddiqi, A.M., Saikhu, A., Fransiska, M.J., Mazeau, L., 2023. An advanced graph embedding framework with node embedding to design water pipeline networks. In: 2023 14th International Conference on Information & Communication Technology and System (ICTS). Presented at the 2023 14th International Conference on Information & Communication Technology and System (ICTS). Surabaya, Indonesia. IEEE, pp. 243–248. https://doi.org/10.1109/ICTS58770.2023.10330890.

Salehinasab, S., Goshtasbi Rad, E., 2025. Enhancing pipe failure prediction using cluster-specific classification via K-means clustering. Sustain. Resilient. Infrastruct. 10, 591–620. https://doi.org/10.1080/23789689.2025.2470504.

Scheidegger, A., Leitão, J.P., Scholten, L., 2015. Statistical failure models for water distribution pipes – a review from a unified perspective. Water. Res. 83, 237–247. https://doi.org/10.1016/j.watres.2015.06.027.

Scholten, L., Scheidegger, A., Reichert, P., Mauer, M., Lienert, J., 2014. Strategic rehabilitation planning of piped water networks using multi-criteria decision analysis. Water. Res. 49, 124–143. https://doi.org/10.1016/j.watres.2013.11.017.

Schönemann, P.H., 1966. A generalized solution of the orthogonal procrustes problem. Psychometrika 31, 1–10. https://doi.org/10.1007/BF02289451.

Snider, B., McBean, E.A., 2020. Watermain breaks and data: the intricate relationship between data availability and accuracy of predictions. Urban. Water. J. 17, 163–176. https://doi.org/10.1080/1573062X.2020.1748664.

Snider, B., McBean, E.A., 2021. Combining machine learning and survival statistics to predict remaining service life of Watermains. J. Infrastruct. Syst. 27, 04021019. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000629.

Soutinho, G., Meira-Machado, L., 2023. Nonparametric estimation of the distribution of gap times for recurrent events. Stat. Methods Appl. 32, 103–128. https://doi.org/10.1007/s10260-022-00641-6.

Taiwo, R., Shaban, I.A., Zayed, T., 2023. Development of sustainable water infrastructure: a proper understanding of water pipe failure. J. Clean. Prod. 398, 136653. https://doi.org/10.1016/j.jclepro.2023.136653.

Taiwo, R., Zayed, T., Adey, B.T., 2025a. A novel ensemble learning framework for predicting the causes of water pipe failures. Reliab. Eng. Syst. Saf. 264, 111320. https://doi.org/10.1016/j.ress.2025.111320.

Taiwo, R., Zayed, T., Bakhtawar, B., Adey, B.T., 2025b. Explainable deep learning models for predicting water pipe failures. J. Environ. Manage 379, 124738. https://doi.org/10.1016/j.jenvman.2025.124738.

Tscheikner-Gratl, F., Sitzenfrei, R., Rauch, W., Kleidorfer, M., 2016. Integrated rehabilitation planning of urban infrastructure systems using a street section priority model. Urban. Water. J. 13, 28–40. https://doi.org/10.1080/1573062X.2015.1057174.

Wang, C., Zhou, H., Lin, S., Weng, X., Shao, Y., Yu, T., 2025. A hybrid deep survival model for failure modeling of water distribution networks coupling physical survival and data reconstruction. Reliab. Eng. Syst. Saf. 264, 111401. https://doi.org/10.1016/j.ress.2025.111401.

Wéber, R., Huzsvár, T., Hős, C., 2020. Vulnerability analysis of water distribution networks to accidental pipe burst. Water. Res. 184, 116178. https://doi.org/10.1016/j.watres.2020.116178.

Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W., Tscheikner-Gratl, F., 2018. Pipe failure modelling for water distribution networks using boosted decision trees. Struct. Infrastruct. Eng. 14, 1402–1411. https://doi.org/10.1080/15732479.2018.1443145.

Xu, Y., He, Z., 2024. Pipe failure prediction in the water distribution system using a deep graph convolutional network and temporal failure series. ACS EST Eng. 4, 2252–2262. https://doi.org/10.1021/acsestengg.4c00234.

Yu, X., Wu, Y., Meng, F., Zhou, X., Liu, S., Huang, Y., Wu, X., 2024. A review of graph and complex network theory in water distribution networks: mathematical foundation, application and prospects. Water. Res. 253, 121238. https://doi.org/10.1016/j.watres.2024.121238.

Zhou, X., Guo, S., Xin, K., Tang, Z., Chu, X., Fu, G., 2025. Network embedding: the bridge between water distribution network hydraulics and machine learning. Water. Res. 273, 123011. https://doi.org/10.1016/j.watres.2024.123011.