Water Practice & Technology

Water Practice & Technology Vol 18 No 1, 264 doi: 10.2166/wpt.2022.159

Logistic pipe failure prediction models for an urban water distribution network in the developing world: a case study of Kampala water, Uganda

Rose Auma[a], Isaac G. Musaazi[b], Martin D. Tumutungire[a] and Jotham Ivan Sempewo[a,*]

[a] College of Engineering, Design, Art and Technology, School of Engineering, Department of Civil and Environmental Engineering, Makerere University, P.O. Box 7062, Kampala, Uganda
[b] College of Engineering and Architecture, Department of Civil and Environmental Engineering, Howard University, Washington, DC 20059, USA
*Corresponding author. E-mail: jotham.sempewo@mak.ac.ug; jothamsempewo@yahoo.com

JIS, 0000-0002-9897-211X

ABSTRACT

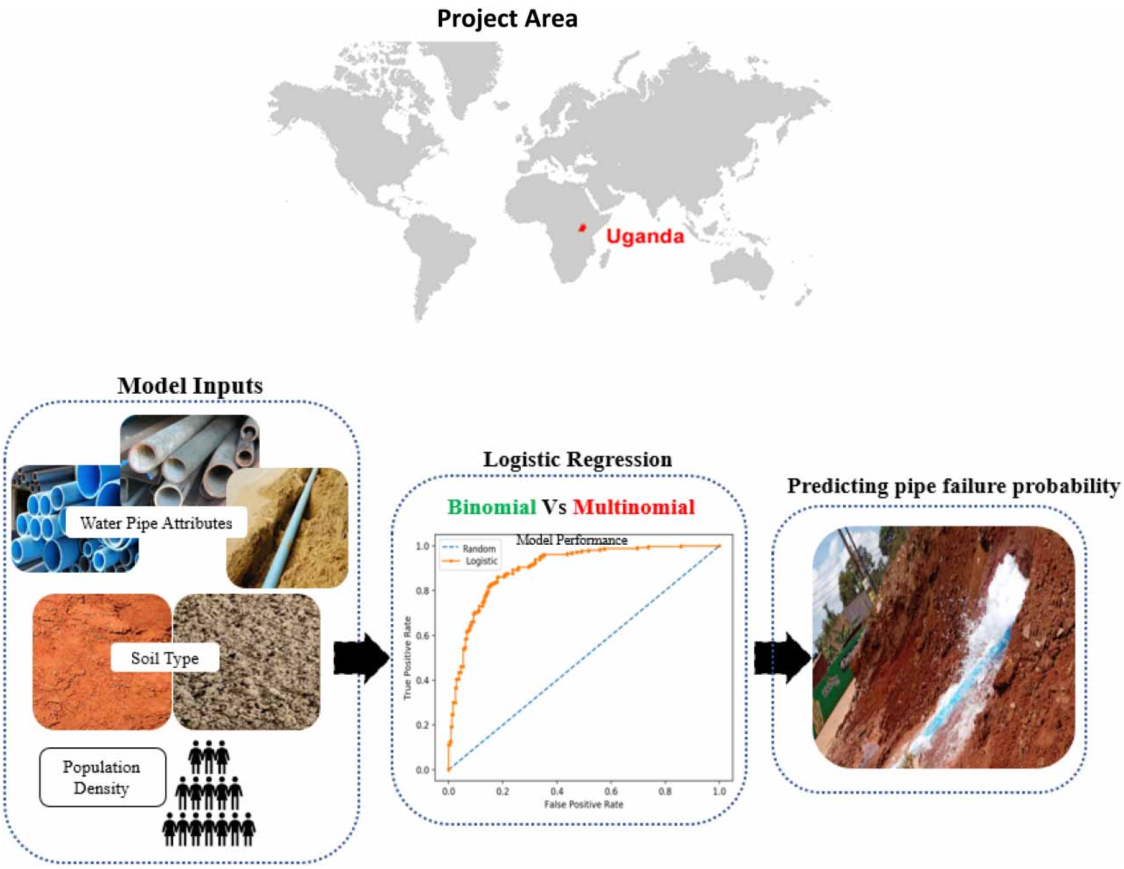
Statistical models can be used as proactive approaches to pipe failure management for the satisfactory and efficient functionality of a water distribution network (WDN). The study aimed to develop two logistic regression models using historical data and evaluated them based on prediction accuracy, receiver operator characteristics (ROC), and area under the curve (AUC). Pipe sizes ranging from 150 mm to 350 mm in the WDN were adequate to prevent pipe failure. However, a 250 mm pipe diameter had the lowest failure probability. Old pipes had a lower failure probability than new pipes. Although it was evident that the installation design of water pipes is changing from steel to unplasticized polyvinyl chloride (uPVC), steel pipes had a lower failure probability than uPVC at the same depth from the soil surface. Pipes buried in gravel with a small diameter had a lower failure probability than those in clay of a bigger diameter. With a median pipe age of 8 years in the WDN and greater class weight on pipe failures, the binomial logistic regression model had better performance (accuracy – 96.9%, AUC – 0.996) than the multinomial logistic model (accuracy – 90.9%, AUC – 0.992), representing reliable model predictions. The models can be used to modify data collection protocols to better identify potential water pipes that require maintenance or replacement.

Key words: failure frequency, logistic regression, model, prediction, water distribution, water mains replacement

HIGHLIGHTS

- Pipe failure is intricate and depends on physical, environmental, and operational pipe attributes.
- 250-mm pipe diameter had the lowest failure probability.
- Old pipes have a lower failure probability than new pipes.
- Pipes with low population density had a higher failure probability than those in densely populated areas.
- Binomial logistic regression model had better performance than the multinomial logistic model.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Water utilities find challenges in maintaining the functionality of water mains that are nearing the end of their design life (Wang *et al.* 2013; Zhou 2018). Significant leakages from the increasing pipe bursts within a water distribution network (WDN) signal a reduction in water pipe integrity. The public health consequences linked to pipe failure mean areas within the WDN that fail consistently need to be prioritized for replacement or repair. However, advanced condition assessment of failing systems requires experts with substantial network knowledge which may sometimes be limited or lacking (Campanella *et al.* 2015; Zhou 2018).

Models can reveal the optimal scale and time of the required rehabilitation efforts in old infrastructure and the factors leading to the failure of water pipes thus contributing to proactively managing failures across WDNs to reduce physical water losses. Among these are physical models developed based on factors such as pipe attributes, quality of the pipe, fittings material, and operational network pressure loads to understand the pipe deterioration process (Wang *et al.* 2013; Gao 2017). Although rehabilitation/replacement programs can be informed from pipe deterioration factors during the design and implementation phases of WDNs. However, simulating pipe failure mechanisms and assessing the pipe capacity to resist failure during WDN design and implementation phases requires a long period of input data (Yamijala *et al.* 2009; Wang *et al.* 2013; Chen *et al.* 2017). In addition, the scarcity of financial resources to conduct regular checks on the whole pipe network to inform pipe maintenance is impossible with pipe repair sought only after breakage of pipes has occurred and been reported, particularly in the developing world. Statistical models can help overcome this challenge based on collected historical data over time to model the status of water distribution pipes and establish the pipes that are more likely to break (Winkler *et al.* 2018).

WDN data can vary across geographical regions which makes reliance on available data collected for model development and appropriate model selection challenging. However, the benefit of using predictive models to identify water infrastructure failure, inform maintenance schedules, and make pipe repair or renewal decisions is widely known. For instance, Sempewo & Kyokaali (2016) and Sempewo & Kyokaali (2019) used the

Markov approach to predicting the condition of the pipes in a water distribution system based on failure history backed up by frequent updates of the data collection protocol at a small geographical service area is not fully representative of the unique failure patterns across the entire WDN. The influence of pipe material on pipe failures showed that pipes with asbestos cement have lower failure rates than ductile cast iron pipes in a cox model developed (Debón *et al.* 2010). However, cox models are not suitable for water utilities in the developing world that have just recently started collecting and archiving pipe information with related causes of failure (Barrantes 2018). Several studies have applied logistic regression to determine pipe failure probability useful to inform replacement decisions and are more accurate than predicting the total number of failures at the pipe level (Yamijala *et al.* 2009; Wang *et al.* 2013; Zhou 2018; Motiee & Ghasemnejad 2019). Although the sample size introduces bias in the approximation of model parameters such as pipe age and diameter, logistic models can capture multiple explanatory variables necessary for improving model reliability and predictive capability (Vladeanu & Koo 2015). Therefore, pipe maintenance guided by logistic regression can be useful to water utilities to plan pipe inspection schedules and will more likely prevent pipe breaks in the future.

2. MATERIALS AND METHODS

2.1. Study area details

This study was conducted in Kampala, the capital city of Uganda with a water supply service covering an area of about 300 km^2 with a total pipe length of approximately 3,587 km of various sizes ranging from DN50 to DN700 (Sakomoto *et al.* 2020). The pipe network was laid in 1929. The pipe materials account for approximately 98% of the Kampala WDN, including steel and unplasticized polyvinyl chloride (PVC). The pipe length was typically 500 m. The data were aggregated to capture localized influences of environmental and operational parameters and allowed for comparison of pipe failure patterns across the different pipelines thus yielding more statistically accurate predictions. Pipe failure models have been limited to predicting the probability of failure. However, the probability of failure is often not sufficient to support the design of pipe inspection regimes, thus our study extends prior efforts and empirically relates failure probability to pipe attributes.

2.2. Data used

Historical pipe information constituting physical, environmental, and operational attributes spanning from the period 1964–2017 was assessed from the National Water and Sewerage Corporation (NWSC) database. The data constituted 58 pipelines with a nominal diameter of 150–300 mm and 84 pipe sections. Among the physical attributes in the database were pipe material, installation year, and the number of failures. The environmental attributes included pipe location, soil type, distance from the road, depth, traffic impact, and population density. Operational attributes included pressure rating, and average system pressure. The pipeline name, pipe size, length, and location defined the water distribution line. The pipe size was extracted from the pipeline name. Descriptions of the attributes in the historical database can be found in the supplemental information (Supplementary material, Table S1). The bathtub curve describes the pipe failure based on three life stages, i.e., early life, useful life, and wear our life (Singh & Adachi 2013). The historical database assessed was not sufficient to provide a complete failure dataset needed to describe all these three stages. Information needed to describe early pipe failure such as poor installation quality and manufacturer flaws could not easily be established. The wear-out life stage defined by conditions when pipes are generally weak with a high probability of failure was also unknown. Therefore, in this study, only the useful life period based on the information in the database was considered.

The frequency of failure per 500 m per year, N_t was determined where N_o represents the total number of failures from the first installation, $t - t_o$ is the age of the pipe since installation date, A is the annual growth coefficient of failure assumed to be 0.05 (Shamir & Howard 1979):

$$N_t = N_o e^{A(t-t_0)} \tag{1}$$

The frequency of failure is better utilized when predicting failure at a network scale while the probability of failure is deemed more appropriate for individual pipes (Barton *et al.* 2022). Therefore, the probability of pipe failure p was determined by ranking the frequency of failure, N_t across the various pipe sections in the historical database, where r is the rank of the frequency of failure within the data, s is the length of unique ranks assigned to the failure probability. Criteria used to assess pipe failure based on failure probability which is of interest to water utilities rather than the number of failures were developed and are presented in Supplementary material, Table S2:

$$p = 1 - \frac{r}{s + 1} \tag{2}$$

ANOVA was used as a basis to compare the significance of each of the categorical variables, i.e., soil type, pipe material, pressure rating, traffic impact, and population density to pipe failure probability by determining the variability between and within groups. The sum of squares (SS) and mean squares (MS) were determined. The *F*-value was determined as a ratio of between and within group MS. *P*-value was estimated from the *F*-value and degree of freedom. The strength and direction of association between the independent variables and pipe failure probability were measured using Spearman's correlation coefficient ($\boldsymbol{\rho}$). The level of significance was defined at 0.05.

2.3. Model development

Logistic regression failure models were developed to predict the probability of pipe failure occurring based on two categorical variables, i.e., soil type, and population density, and five quantitative variables, i.e., pipe age (years), pipe depth (m), average pressure (m), the distance of the pipe from the road and pipe size (mm). The outcome of the estimation was a probability of pipe failure bounded between 0 and 1. A binomial logistic regression had the labels restricted to two conditions which were pipe failure or no pipe failure while the multinomial logistic regression had labels take on more than two values (Hosmer & Lemeshow 2000). A logit transformation was applied on the odds as shown in the binomial and multinomial logistic regression formulae, where Y is the dichotomous dependent variable, α is the intercept term; β_1 and β_{i1} are the regression coefficients and X_1 are the independent variables in Equations (3) and (4):

$$log(Y) = ln\left(\frac{(P(Y = 1|X_1, X_2, \ldots, X_p)}{1 - P(Y = 1|X_1, X_2, \ldots, X_p)}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p \tag{3}$$

$$log(Y) = ln\left(\frac{P(Y = (iX_1, X_2, \ldots, X_p)}{P(Y = K(X_1, X_2, \ldots, X_p)}\right) = \alpha_i + \beta_{i1} X_1 + \beta_{i2} X_2 + \cdots + \beta_{ip} X_p \tag{4}$$

2.4. Logistic regression model training and performance

Both models were trained on 60% of the data and tested on 40% of the data. Before the model was fit to the training set to predict whether pipe failure occurred, model tuning was completed using the *K*-fold cross-validation technique (Pedregosa *et al.* 2011). Here the training set was generally divided into *K* folds (typically five), and then for each iteration of model training, one fold was taken as the test set and the remaining folds were used for training.

Class weights were used depending on the occurrence of pipe failure on the training data. This was important to obtain a more reliable and unbiased classification model for use on the testing set because the usage of unbalanced class weights on pipe failure was responsible for eliminating bias towards the most occurring failure categories in the training data. The models included a regularization step which generally provided a complexity penalty for more extreme model parameters (Hastie *et al.* 2008). The regularization terms added weights to enhance performance on the test set. Grid-searching was completed to configure the most optimal parameters for the two models based on Supplementary material, Table S1 (Details in Supplementary material, Text S2).

The test data were used to represent the unseen data scenario of the model predictions. The actual and predicted values were compared to determine the accuracy of the model based on true-positive (TP), true-negative (TN), false-negative (FN), and false-positive (FP) (Fawcett 2006). Details can be found in Supplementary material, Text S3. Precision measured the relevance of pipe failure (Supplementary material, Equation S1), recall measured the percentage of pipe failures that were classified correctly representing the sensitivity of the model (Supplementary material, Equation S2), and F1 was the weighted average of precision and recall (Supplementary material, Equation S3) of the model. In addition to the above terms, information on how the classification model avoided false classifications was visualized on a receiver operating characteristic (ROC) curve based on the false-positive rate (FPR) (Fawcett 2006). FPR is a measure of how many instances of pipe failure the model predicted from all the pipes that did not fail. The diagonal line $y = x$ represented a model making random guesses about a specific class. The area under the curve (AUC) was a combined measure of the performance of the classification model describing the capability of the model to distinguish between the classes. The AUC measure returned values between zero and one, with random guesses of the class labels producing an area of 0.5 for which no realistic classifier should. The analysis was completed in Python 3.8 (Van Rossum & Drake 2009).

3. RESULTS AND DISCUSSION

3.1. Data analysis

Pipe diameter and material are static parameters that can be determined at the installation phase or changed during the rehabilitation stage to minimize pipe failure. Pipe failure was independent of pipe size ($\rho = -0.08$, $p > 0.05$) (Supplementary material, Table S5), indicating that the WDN range of pipe sizes (150–350 mm) is adequate to prevent pipe failure. Lower pipe sizes (less than 200 mm) and larger pipe sizes (300 mm) (Table 1), and further away from the main road, on average 4 m, were found to be more likely to fail than 250 mm pipes. The lower failure probability of 250 mm pipes cannot be attributed to depth (buried at similar depths), pipe length (all sections had the same length (500 m)), pressure (negligible pressure difference), and pipe material (mostly uPVC). The literature on the influence of pipe diameter on pipe failure is conflicting. Pipes less prone to failure were characterized by large diameters (Debón *et al.* 2010; Motiee & Ghasemnejad 2019). High failure rates in small and large-diameter pipes have been reported elsewhere (Wang *et al.* 2013; Rajeev *et al.* 2014; Bruaset & Sægrov 2018; Giraldo-González & Rodríguez 2020). Smaller diameter pipes are associated with high failure probabilities because of reduced strength and wall thickness (Bruaset & Sægrov 2018). Pipes of size 250 mm were found to be older than other pipe sizes, i.e., greater than the median age (8 years) of the WDN.

Table 1 | Mean, standard deviation, median, minimum, and maximum values for each of the independent variables from the historical database

	Summary statistics				
Variables	**Mean**	**Std.**	**Median**	**Min**	**Max**
Pipeline (mm)	234.76	57.57	250.00	150.00	350.00
Depth (m)	1.04	0.26	1.00	0.50	1.80
Distance from the road (m)	3.87	2.25	4.00	0.00	13.50
Average pressure (m)	47.15	14.51	45.00	25.00	90.00
Age (year)	15.93	16.67	8.00	1.00	54.00
Pipe failure probability	0.52	0.30	0.54	0.01	0.98

Pipe failure had a high negative correlation with pipe age, implying that pipe failure was more likely to happen during the initial installation stage which decreased over time (Supplementary material, Table S4). Age indicates the time a pipe has been laid and exposed to external loads from the surrounding environment. It is expected that older pipes are likely to break and the pipes that have experienced breakage before are likely to be more prone to breaks in the future. However, from a practical view, damages during installation cause fast pipe replacement increasing the probability of failure which biases the older pipes that appear in 'good' condition with lower failure rates (Winkler *et al.* 2018). Our results are in agreement with an analysis of pipe failure records which showed a complex relationship between age and failure with higher failure rates at the beginning that reduce to lower levels before increasing again (Wengström 1993). The need to capture the actual date of installation and pipe condition following installation is needed for better characterization of pipe failure. In addition, more data collection at locations with 250 and 350 mm pipe sizes to better understand the low failure probability. The lower failure probability with a 350 mm pipe diameter is based on an extremely limited dataset ($n = 1$).

Steel pipes installed at about the same depth from the soil surface had a lower failure probability than uPVC. Pipe failure has a very low positive correlation to uPVC pipe material indicating that uPVC is more likely to break than steel (Supplementary material, Table S5). Notably, uPVC was laid further away from the main road than steel pipes meaning other factors are contributing to the high probability of failure than those currently available in the NWSC database. On average, uPVC is relatively new having been installed within the last 10 years reflecting a change in the installation design of water pipes in the Kampala WDN. Information on the strength of the pipe material to define the wear-out life stage conditions across multiple pipe sections can be collected as steel pipes continue to be replaced with uPVC pipes in the current rehabilitation program. Soil corrosion causes the deterioration of steel pipes reducing their ability to resist various internal and external water flow shocks (Tang *et al.* 2019). Perhaps, since many of the steel pipes were earlier installed in the WDN than in uPVC, the low failure probability can be attributed to maintenance efforts to protect against corrosion. Therefore, the database should be updated to differentiate between newly rehabilitated steel pipes and those pending rehabilitation to determine whether maintenance reduces pipe failure probability in steel pipes. In addition, more resolved spatial data at the pipe level is needed to indicate whether the same or different pipe sections experience failure from the time of installation to further our understanding of the relationship between pipe failure and material properties.

The average pressure in the network was ~50 m-head (Table 1). Interestingly, a higher pressure rating had a lower probability of failure than a lower-pressure rating meaning that pipes with higher average pressures were less prone to failure than those pipes experiencing low pressure. No statistically significant difference was found between PN10 and PN16 ($p > 0.05$) (Supplementary material, Table S4). At an average daily pressure scale, higher internal pipe pressure contributed to higher pipe failure rates (Wols *et al.* 2019). The age of pipes with a higher pressure rating was high and might explain the lower failure probability since the distance from the road and pipe diameter was somewhat similar. From an engineering perspective, pressure above nominal operating conditions or sudden fluctuation in pressure both negatively affect the WDN and can contribute to an increase in pipe breakage probability. The limitation in the resolution of pressure data to show the dynamic nature of pressure in this study led to the assumption that the pressure did not vary along the entire pipe section for simplicity. Therefore, pipe failure caused by intermittent and continuous water supply was inconclusive to enable an investigation of the influence of the spatial distribution of pressure differences across given pipe sections. However, from the authors' experience more resolved pressure data is not usually captured in failure records, and not usually included in developing pipe failure models (Winkler *et al.* 2018; Motiee & Ghasemnejad 2019; Wols *et al.* 2019; Kerwin *et al.* 2020) and might require installation of pressure sensors at multiple points within the WDN, unaffordable for many water utilities.

Interestingly, pipe failure had a very low negative correlation with pipe depth. Therefore, pipes buried deeper in the soil were more likely to fail than pipes buried close to the surface. Two explanations could probably explain this observation: (i) the small range between these two depths could have explained the low correlation and (ii) pipes buried deeper in the ground corresponded to large pipe diameter and had a lower failure probability than pipes buried closer to the surface which was smaller in diameter.

Pipe failure across the two soil types was significantly different ($p < 0.05$) (Supplementary material, Table S4). Pipe failure had a significant negative correlation with gravel which indicates that pipe failure is higher under clay soil conditions than in gravel soil (Supplementary material, Table S5). Pipes buried in gravel were smaller in diameter and had a lower failure probability than those buried in clay were predominantly larger in diameter.

Although, pipe failure across locations with low, medium, and high population density was statistically not significantly different ($p > 0.05$) (Supplementary material, Table S4), pipes at locations with lower population density had the highest failure of probability than highly densely populated areas. Notably, highly densely populated areas were served with old and large-diameter pipes. Aggregation of population numbers likely caused this, and better categorization of population groups based on the size of the service area or user consumption characteristics is necessary to capture inter-demand variation accurately to improve the assessment of the impact of population density on pipe failure probability.

Pipe failure across locations with low, medium, and high traffic conditions was significantly different ($p < 0.05$) (Supplementary material, Table S4), pipes at locations with high traffic conditions had the lowest failure of probability than low traffic areas. Areas characterized by high traffic conditions had old and most suitable pipe sizes (250 mm). The contribution of traffic to pipe failure is small and specific road objects need to be defined for traffic information to be deemed useful (Moerman *et al.* 2016). The level of settlement of surrounding soil was also found to be subject to different loadings based on the road construction; however, this was not available at the time of this study. In addition, the database did not provide enough spatial data to determine whether perpendicularity between the water pipes and road network was present at the pipe locations presented. Therefore, corrections need to be applied to this traffic data to be useful for modeling, and with the uncertainty around the grouping of traffic data, traffic impact was dropped from further analysis.

3.2. Logistic pipe failure models and performance

A unit increase in the average system pressure, distance from the road, and pipe size decrease the odds of pipe failure for both binomial and multinomial logistic regression models (Supplementary material, Table S5). Pipe failure is less likely to happen when pipes are buried further in the ground, further away from the road, and have larger pipe diameters. Furthermore, changing soil type from clay to gravel soil is associated with an increase in the odds of pipe failure. The location of the pipe failures, many of which were geographically distant, and the assumption of similar operational characteristics such as pipe length, and pipe material (areas with complete vs incomplete rehabilitation of steel pipes) inevitably affected the model and can potentially hinder the inclusion of model predictions in the decision-making process.

The binomial logistic regression failure model had a better performance in predicting pipe failure than the multinomial model indicated by the high accuracy on the testing set (Table 2) and the better AUC (Figure 1). The binomial model had a slightly better class label identification ability than the multinomial model (Figure 1).

The binomial model provided a larger class weight on failure predictions and a smaller class weight for no pipe failure than the multinomial model. The testing set had more pipe failures than no failures to be predicted, i.e., 63% of the test set consisted of pipe failure conditions since the average probability of failure in the dataset was high (Table 1). With a greater class weight on pipe failures ($w = 0.76$) than the multinomial model ($w = 0.47$), the precision of the binomial model was better given the limited sample size of no failures in the test data (Supplementary material, Table S7). The performance of the multinomial model can be improved by balancing the sample size of class labels (failure/no failure).

Table 2 | Comparison of the performance of the binomial and multinomial models based on the train and test accuracy representing the percentage difference between the pipe status labels (no failure/failure) and the model label prediction for the train and test datasets, respectively

		Binomial				Binomial		
Performance metrics	**Training**	**Precision**	**Recall**	**F1-score**	**Testing**	**Precision**	**Recall**	**F1-score**
No failure		1.00	1.00	1.00		1.00	0.92	0.96
Failure		1.00	1.00	1.00		0.95	1.00	0.98
Accuracy		1.00	1.00	1.00		0.97	0.97	0.97
		Multinomial				**Multinomial**		
No failure		0.96	1.00	0.98		0.86	0.92	0.89
Failure		1.00	0.96	0.98		0.95	0.90	0.92
Accuracy		0.98	0.98	0.98		0.91	0.91	0.91

Note: The accuracy of the logistic models is based on 82 data points. The training and test dataset consisted of 50 and 32 data points, respectively.

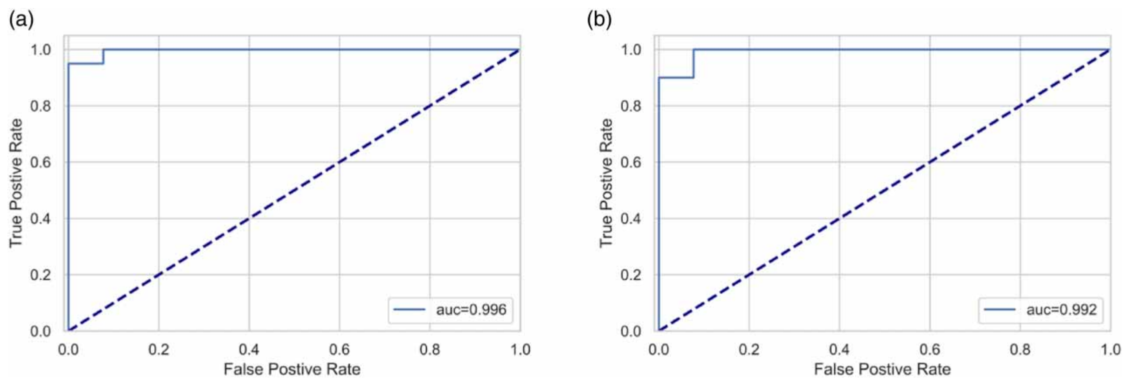


Figure 1 | ROC for binomial (left) and multinomial model (right).

Both models used L2 regularization which forced the weights to be small but did not make them exactly zero. Essentially, the less significant parameters still contributed to the predictions. It is also important to note that L2 regularization is not so robust to outliers. The average pressure (multinomial model), pipe age, and depth of the buried pipes (binomial and multinomial models) had a few outlying values. However, these variables had outlying values in the test set which were embraced in both models that depended on the L2 regularization (Hastie *et al.* 2008). The penalization to prevent overfitting contributed to a deteriorated performance on the test set for the multinomial model than the binomial model, i.e., the binomial model had a higher test accuracy than the multinomial model. The binomial model accuracy is 96.97% representing one incorrectly classified of no pipe failure (index 27, Supplementary material, Table S7). The multinomial model had lower prediction accuracy (90.91%) representing two incorrect classifications of no pipe failure (index 17 and 27, Supplementary material, Table S7) and one incorrect classification of pipe failure (index 30). The model outputs can be useful for planning purposes when identifying pipes that need to be repaired or replaced in locations with a high incidence of failure. The incorrect prediction, however, could lead to a wastage of resources and effort from the early potential replacement of functional water pipes. Our understanding of less significant model parameters and parameters with outlying values on the model predictions will be enhanced if additional models can be explored for example those with L1 regularization to shrink less significant parameters towards zero and exclude outlying values during model development.

Figure 1 shows the ROC curve in the top half of the plot above the $y = x$ diagonal line which represents the lack of the classification model to distinguish between class labels (failure or no failure). In general, an AUC of less than 0.5 is considered unacceptable and >0.9 is considered excellent discrimination (Hosmer & Lemeshow 2000). The top left half of Figure 1 and an AUC value of ~0.99 for both models suggest that the classifiers had an excellent separation ability between the classes and seemed strong to make predictions on failures on the unseen data.

More resolved pressure data and less aggregation of local conditions such as population groups may have proved useful in terms of increasing the accuracy of the model predictions.

4. CONCLUSION AND RECOMMENDATIONS

The focus of this study was to predict pipe failure probabilities using statistical models based on pipe attributes likely to cause failure in pipes. The low correlations of the model inputs, i.e., depth, distance from the road, pipe size, average pressure, population size, and soil type showed that pipe failure is intricate and should require more than one predictor variable. Steel pipes installed at about the same depth from the soil surface had a lower failure probability than uPVC. Smaller pipe sizes (less than 200 mm) and larger pipe sizes (greater than 300 mm) were more likely to fail than 250-mm pipes. Pipes at locations with lower population density had the highest failure of probability than highly densely populated areas characterized by old and large-diameter pipes. The binomial logistic regression failure model had a better performance predicting pipe failure than the multinomial model.

The limitation in the historical database on the status and functionality of the WDN considered in this study posits that practitioners need to understand the benefits of predictive pipe failure models that will encourage the update of data collection protocols. The lack of better resolution of some important variables like average pressure known to accelerate water pipe deterioration, and quality of the information in failure datasets such as the degree of spatial detail at the pipe level indicating whether the same or different pipe sections experienced failure from the time of installation, limited prediction of pipe failures. Improving model accuracy may require tuning the data collection plan to encompass more resolved population density and average pressure readings. However, this data collection plan should consider current challenges and be in concert with the frequency of pipe rehabilitation programs. More invaluable data will potentially improve model predictions and increase reliability in modeling results in making decisions about rehabilitation and replacement efforts.

ACKNOWLEDGEMENTS

This study was undertaken as part of the Building Capacity in Water Engineering for Addressing Sustainable Development Goals in East Africa (CAWESDEA) project which is part of the IDRC-funded programme on Strengthening Engineering Ecosystems in sub-Saharan Africa. CAWESDEA Project is led by Global Water Partnership Tanzania in collaboration with Makerere University Uganda, Moi University Kenya and University of Dar es Salaam Tanzania. We acknowledge the support of the National Water and Sewerage Corporation and

Makerere University for hosting and supporting the project. The anonymous comments from the two reviewers also greatly contributed to framing the content of this paper.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Barrantes, G. 2018 Multi-hazard model for developing countries. *Nat. Hazards* **92**, 1081–1095. https://doi.org/10.1007/s11069-018-3239-6.

Barton, N. A., Hallett, S. H., Jude, S. R. & Tran, T. H. 2022 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.

Bruaset, S. & Sægrov, S. 2018 An analysis of the potential impact of climate change on the structural reliability of drinking water pipes in cold climate regions. *Water (Switzerland)* **10**. https://doi.org/10.3390/w10040411.

Campanella, K., Andreasen, C., Diba, A., Himmelberger, H., Leighton, J., Santini, J. & Vause, K. 2015 Establishing the Level of Progress in Utility Asset Management. American Water Works Association, Denver Colorado, pp. 1–48.

Chen, T. Y., Beekman, J. A. & Guikema, S. D. 2017 Drinking water distribution systems asset management: statistical modelling of pipe breaks. In: *Condition Assessment, Surveying, and Geomatics*, A. Pridmore & J. Geisbush (eds.) (pp. 173–186). American Society of Civil Engineers, Pheonix, Arizona. 173–186.

Debón, A., Carrión, A., Cabrera, E. & Solano, H. 2010 Comparing risk of failure models in water supply networks using ROC curves. *Reliab. Eng. Syst. Saf.* **95**, 43–48. https://doi.org/10.1016/j.ress.2009.07.004.

Fawcett, T. 2006 An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.

Gao, Y. 2017 *Systematic Review for Water Network Failure Models and Cases*. University of Arkansas, Fayetteville, Arkansas.

Giraldo-González, M. M. & Rodríguez, J. P. 2020 Comparison of statistical and machine learning models for pipe failure modeling in water distribution networks. *Water (Switzerland)* **12**. https://doi.org/10.3390/W12041153.

Hastie, T., Tibshirani, R. & Friedman, J. 2008 *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edn. Springer US, New York, NY, USA.

Hosmer, D. W. & Lemeshow, S. 2000 *Applied Logistic Regression*, 2nd edn. John Wiley & Sons, Inc. https://doi.org/10.1002/9781118548387.ch1.

Kerwin, S., Garcia de Soto, B., Adey, B., Sampatakaki, K. & Heller, H. 2020 Combining recorded failures and expert opinion in the development of ANN pipe failure prediction models. *Sustainable Resilient Infrastruct.*, 1–23. https://doi.org/10.1080/23789689.2020.1787033.

Moerman, A., Wols, B. A. & Diemel, R. 2016 The effects of traffic loads on drinking water main failure frequencies in The Netherlands. *Water Pract. Technol.* **11**, 524–530. https://doi.org/10.2166/wpt.2016.057.

Motiee, H. & Ghasemnejad, S. 2019 Prediction of pipe failure rate in Tehran water distribution networks by applying regression models. *Water Sci. Technol. Water Supply* **19**, 695–702. https://doi.org/10.2166/ws.2018.137.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R. & Dubourg, V. 2011 Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830.

Rajeev, P., Kodikara, J., Robert, D., Zeman, P. & Rajani, B. 2014 Factors contributing to large diameter water pipe failure. *Water Asset Manage. Int.* **10**, 9–14.

Sakomoto, T., Lutaaya, M. & Abraham, E. 2020 Managing water quality in intermittent supply systems: the case of Mukono Town, Uganda. *Water (Switzerland)* **12**. https://doi.org/10.3390/w12030806.

Sempewo, J. I. & Kyokaali, L. 2016 Prediction of the future condition of a water distribution network using a Markov based approach: a case study of Kampala water. *Procedia Eng.* **154**, 374–383. https://doi.org/10.1016/j.proeng.2016.07.495.

Sempewo, J. I. & Kyokaali, L. 2019 Comparative performance of regression and the Markov based approach in the prediction of the future condition of a water distribution pipe network amidst data scarce situations: A case study of Kampala water, Uganda. *Water Practice and Technology* **14** (4), 946–958. https://doi.org/10.2166/wpt.2019.075.

Shamir, U. & Howard, C. D. D. 1979 An analytic approach to scheduling pipe replacement. *J. Am. Water Works Assoc.* **71**, 248–258. https://doi.org/10.1002/j.1551-8833.1979.tb04345.x.

Singh, A. & Adachi, S. 2013 Bathtub curves and pipe prioritization based on failure rate. *Built Environ. Project Asset Manage.* **3**, 105–122. https://doi.org/10.1108/BEPAM-11-2011-0027.

Tang, K., Parsons, D. J. & Jude, S. 2019 Comparison of automatic and guided learning for Bayesian networks to analyse pipe failures in the water distribution system. *Reliab. Eng. Syst. Saf.* **186**, 24–36. https://doi.org/10.1016/j.ress.2019.02.001.

Van Rossum, G. & Drake, F. L. 2009 *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.

Vladeanu, G. J. & Koo, D. D. 2015 A comparison study of water pipe failure prediction models using Weibull distribution and binary logistic regression. In: *Recent Advances in Underground Pipeline Engineering and Construction* F. Sever & L. Osborn (eds.) (pp. 1590–1601). American Society of Civil Engineering, Baltimore, Maryland, USA, pp. 1590–1601.

Wang, R., Dong, W., Wang, Y., Tang, K. & Yao, X. 2013 Pipe failure prediction: a data mining method. In: *International Conference on Data Engineering*, C. S. Jensen, C. Jermaine, J. Lu, E. Tanin, & X. Zhou (eds.) (pp. 1208–1218). Institute of Electrical and Electronics Engineers, Brisbane, Australia, pp. 1208–1218. https://doi.org/10.1109/ICDE.2013.6544910.

Wengström, T. 1993 *Comparative Analysis of Pipe Break Rates: A Literature Review*. Chalmers Univ. Technol. Gothenburg, Sweden.

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.

Wols, B. A., Vogelaar, A., Moerman, A. & Raterman, B. 2019 Effects of weather conditions on drinking water distribution pipe failures in The Netherlands. *Water Sci. Technol. Water Supply* **19**, 404–416. https://doi.org/10.2166/ws.2018.085.

Yamijala, S., Guikema, S. D. & Brumbelow, K. 2009 Statistical models for the analysis of water distribution system pipe break data. *Reliab. Eng. Syst. Saf.* **94**, 282–293. https://doi.org/10.1016/j.ress.2008.03.011.

Zhou, Y. 2018 Deterioration and optimal rehabilitation modelling for urban water distribution systems. *Deterioration Optimal Rehabil. Modell. Urban Water Distrib. Syst.* https://doi.org/10.1201/9780429451799.

First received 11 November 2021; accepted in revised form 2 December 2022. Available online 19 December 2022