

Predicting the Future Failures of Urban Water Systems: Integrating Climate Change and Machine Learning Prediction Models [†]

Melica Khashei ^{*}, Fatemeh Boloukasli ahmadgourabi  and Rebecca Dziedzic 

Building, Civil and Environmental Engineering Department, Concordia University, Montreal, QC H3G 2W1, Canada; fatemeh.boloukasliahmadgourabi@mail.concordia.ca (F.B.a.); rebecca.dziedzic@concordia.ca (R.D.)

^{*} Correspondence: m_khashe@live.concordia.ca

[†] Presented at the 3rd International Joint Conference on Water Distribution Systems Analysis & Computing and Control for the Water Industry (WDSA/CCWI 2024), Ferrara, Italy, 1–4 July 2024.

Abstract: The state of watermain systems is intrinsically linked to climate factors such as fluctuations in temperature and variations in rainfall. However, the integration of these climate-related factors into watermain failure prediction models, with a specific focus on climate change impacts, remains insufficiently explored. In response to these challenges, this research incorporates the potential effects of climate change on the frequency of watermain breaks by utilizing machine learning techniques, including K-Nearest Neighbours, Random Forest, Artificial Neural Network, and Extreme Gradient Boosting. By leveraging projected climate trends, the models provide actionable intelligence that can inform the development of more robust maintenance and rehabilitation strategies.

Keywords: watermain failure; climate change; machine learning



Citation: Khashei, M.; Boloukasli ahmadgourabi, F.; Dziedzic, R. Predicting the Future Failures of Urban Water Systems: Integrating Climate Change and Machine Learning Prediction Models. *Eng. Proc.* **2024**, *69*, 35. <https://doi.org/10.3390/engproc2024069035>

Academic Editors: Stefano Alvisi, Marco Franchini, Valentina Marsili and Filippo Mazzoni

Published: 3 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Water utilities ensure a consistent supply of clean water to customers, with pipe infrastructure playing a crucial role in maintaining the security and quality of the water supply. These utilities possess a significant number of aging pipe assets that are nearing or have exceeded their intended lifespan [1]. At the same time, climate change is impacting all forms of infrastructure, including watermains, through changes in temperature fluctuations, freeze–thaw cycles, and rainfall patterns [2]. These challenges necessitate the urgent development of watermain break prediction models that incorporate climate factors, ensuring more effective and proactive watermain management.

Over the past decades, a variety of models for predicting pipe breaks have been developed. The evolution of Machine Learning (ML) algorithms stands out as one of the key technological breakthroughs of the 21st century. Lately, experts in water research are turning to these ML techniques to tackle the issue of pipe failure, a critical challenge for the security of urban water distribution systems. Commonly used models in the literature include Artificial Neural Networks (ANN), Support Vector Machines (SVM), Evolutionary Polynomial Regression (EPR), and, more recently, tree models. These models integrate different features to predict watermain breaks, including pipe intrinsic, historical, and environmental data [3].

Incorporating climate data into the analysis of historical failure records allows the model to uncover patterns and connections that might remain hidden when examining failure data in isolation. Watermain breaks influenced by climatic conditions primarily arise from changes in temperature, rainfall, and wind speed [4]. Climatic covariates identified in previous studies are applied herein to explore the impact of climate change.

Climate change presents challenges for utilities in developing sustainable management and rehabilitation plans [4]. It could modify rainfall patterns, potentially leading to

prolonged droughts that reduce groundwater levels. This drop in groundwater can lead to soil compaction, which might increase differential soil settlement. Consequently, such changes in the ground could pose a risk to the integrity of buried water infrastructure [5]. Climate change-driven behavioral shifts, like varying heating and cooling requirements, can also impact patterns of water usage. Elevated demands for water can lead to a rise in internal pressure within watermains, which, in turn, may contribute to the likelihood of pipe failures [6].

However, only a few studies have investigated the effect of climatic variations on watermain break prediction. This study seeks to bridge this research gap by developing ML models designed to forecast watermain breaks, with a focus on incorporating climate-related covariates. Eventually, this model would be deployed with the purpose of identifying potential watermain breaks that may be attributed to the effects of different climate change scenarios. Such predictive capability would enable utilities to adopt proactive maintenance and repair tactics, thereby minimizing the risk of disruptions in the water supply.

2. Materials and Methods

This section outlines the data collection, processing, and modeling methodology used in this study. System-related data and historical break records were obtained from Kitchener's watermain network [7]. The first dataset encompasses the watermain inventory, recording the characteristics of pipes, covering aspects such as their length, diameter, material, and more. The second dataset records watermain breaks between 1985 and 2018. As the present research investigates the impact of climate change on watermain breaks, historical weather data including minimum, maximum, and mean temperatures, and rainfall are also collected from the Environment and Climate Change Canada (ECCC) [8]. To clean the data, initial steps involved removing records with missing values, inconsistencies, or outliers to ensure data reliability. Furthermore, categorical components were encoded using one-hot encoding as necessary. The cleansed datasets, including inventory, break records, and climate data, were then merged through the unique ID of each pipe and time period.

After the data were cleaned and prepared, it was divided into training and testing sets to evaluate the predictive performance of the ML models by randomly assigning 30% of the data to the test set, and the remaining 70% was used for training.

This study seeks to predict the future status of water pipes, either broken or unbroken ones. Since compiling a dataset with yearly records for each pipe is cumbersome and leads to extreme data imbalance, decade-long time intervals were used. This approach facilitates a comprehensive evaluation of the influence of time-dependent variables, including cumulative failures, pipe age, and climate-related factors, within each defined interval. Furthermore, the available data for this study exhibits a high level of imbalance, meaning that one class has significantly more observations than the other. To address the issue of imbalance data, this study focuses on Cast Iron (CI) pipes, given their higher number of breaks compared to other types.

To examine the impact of climate change on watermain failures, various climate-related variables were considered, including min, max, and mean temperatures, air temperature changes, intensities of air temperature changes, variation in temperature, freezing and thawing index, cumulative cold, hot, and thawing days, and total rain.

Four ML models were compared: Random Forest (RF), K-nearest neighbour (KNN), Artificial Neural Network (ANN), and Extreme Gradient Boosting (XGBoost). Random Forest is an ensemble learning method that utilizes multiple decision trees to improve prediction accuracy and reduce the risk of overfitting. XGBoost as a robust and effective algorithm that uses a collection of decision trees to create accurate predictions. KNN, a non-parametric method, uses the distance between data points to classify new data. The underlying assumption is that data points in close proximity to one another are probably members of the same class. The Artificial Neural Network (ANN), a type of feedforward

neural network, uses the backpropagation algorithm to construct the predictive model. The configuration, including the number of neurons in each layer, was optimized through a trial-and-error method. And the hyperparameters of each ML algorithm were optimized using Randomized Search CV optimization. For evaluating the models, the following evaluation metrics were employed: accuracy, precision, and F1 score.

To evaluate the effect of climate change, future projection of temperature and precipitation for three scenarios (SSP1, SSP2, and SSP5) were taken from Environment and Climate Change Canada simulations [8]. The first scenario, SSP1, embodies a sustainable, low-emission future, potentially keeping warming below 2 °C. SSP2 represents a moderate path, with uneven development and a mid-century emissions peak, leading to moderate warming. In contrast, SSP5 depicts a high-emission, fossil-fuel-reliant world with significant warming [9].

3. Results

The performance of the ML models is shown in Table 1. Among the four compared ML models, RF was found to perform the best, especially in terms of F1 Score. KNN showed excellent precision; however, its recall is the lowest, suggesting it might miss the prediction of a significant number of broken pipes. Conversely, for ANN, its recall rate stands out, but lower precision leads to a higher false-positive rate, implying a less-reliable result. While XGBoost demonstrates a balanced result between precision and recall, its performance metrics are slightly lower than RF. Therefore, as RF offers a balanced classification with strong performance across all metrics, it was used to make future predictions under three different climate change scenarios.

Table 1. Performance of each ML model.

	KNN	RF	ANN	XGBoost
F1 Score	0.632	0.642	0.628	0.627
Recall	0.531	0.568	0.808	0.560
Accuracy	0.930	0.928	0.892	0.925
precision	0.780	0.738	0.513	0.713

The predicted number of broken pipes for SSP1, SSP2, and SSP3 are 484, 352, and 456, respectively. The highest predicted number of broken pipes under the SSP1 scenario indicates that this is the most challenging or deteriorative condition, suggesting greater vulnerability of CI pipes in colder climates. Conversely, the SSP2 scenario, having the fewest predicted broken pipes and being associated with moderate future warming, implies that the pipes are less susceptible to breaks in moderately warm climate conditions. The higher incidence of broken pipes observed under the SSP5 scenario, relative to SSP2, suggests that more extreme warming accelerates the deterioration of water pipes in comparison to moderate warming. This may stem from various reasons, like longer dry periods affecting soil settlements, and increase in water demand, which leads to higher internal pressure in pipes [4].

4. Discussion

The variation in predicting numbers of broken pipes across scenarios demonstrates how sensitive infrastructure responses may be to diverse future situations. It is, therefore, important to understand how different scenarios would influence the future state of pipe infrastructure for purposes of adequate planning and allocation of resources. For decision makers, such findings will be able to help them decide on the investments and interventions so as to lower the likelihood of pipe failures, as well as enhance the resilience and reliability of water distribution networks.

Author Contributions: This study was collaboratively conducted. Conceptualization, M.K., F.B.a. and R.D.; methodology, M.K. and F.B.a.; validation, M.K., F.B.a. and R.D.; formal analysis, M.K. and F.B.a.; investigation, M.K., F.B.a. and R.D. data curation, M.K., F.B.a. and R.D.; writing—original draft preparation, M.K. and F.B.a.; writing—review and editing, R.D.; visualization, M.K. and F.B.a.; supervision, R.D.; funding acquisition, R.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Sciences & Engineering Research Council, grant number RGPIN-2022-04664, and by the National Research Council, grant number A-0048967.

Data Availability Statement: The original data presented in this study are openly available in [8,9].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rifaai, T.M.; Abokifa, A.A.; Sela, L. Integrated approach for pipe failure prediction and condition scoring in water infrastructure systems. *Reliab. Eng. Syst. Saf.* **2022**, *220*, 108271. [CrossRef]
2. Lawrence, J.; Blackett, P.; Cradock-Henry, N.A. Cascading climate change impacts and implications. *Clim. Risk Manag.* **2020**, *29*, 100234. [CrossRef]
3. Almheiri, Z.; Meguid, M.; Zayed, T. Review of Critical Factors Affecting the Failure of Water Pipeline Infrastructure. *J. Water Resour. Plan. Manag.* **2023**, *149*, 03123001. [CrossRef]
4. Ahmad, T.; Shaban, I.A.; Zayed, T. A review of climatic impacts on water main deterioration. *Urban Clim.* **2023**, *49*, 101552. [CrossRef]
5. Wols, B.A.; Van Thienen, P. Modelling the effect of climate change induced soil settling on drinking water distribution pipes. *Comput. Geotech.* **2014**, *55*, 240–247. [CrossRef]
6. Wols, B.A.; Vogelaar, A.; Moerman, A.; Raterman, B. Effects of weather conditions on drinking water distribution pipe failures in the Netherlands. *Water Supply* **2018**, *19*, 404–416. [CrossRef]
7. Kitchener GeoHub. Available online: <https://open-kitchenergis.opendata.arcgis.com/search?q=kitchener> (accessed on 23 March 2024).
8. Environment and Climate Change Canada. Available online: <https://climate.weather.gc.ca/> (accessed on 23 March 2024).
9. Eyring, V.; Bony, S.; Meehl, G.A.; Senior, C.A.; Stevens, B.; Stouffer, R.J.; Taylor, K.E. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model. Dev.* **2016**, *9*, 1937–1958. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.