INTEGRATED INTELLIGENT MODELS FOR UNDERSTANDING AND PREDICTING PIPE FAILURES IN WATER DISTRIBUTION NETWORKS

TAIWO RIDWAN ADEMOLA

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

Department of Building and Real Estate

Integrated Intelligent Models for Understanding And Predicting Pipe Failures In Water Distribution Networks

TAIWO Ridwan Ademola

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

August 2024

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

(Signed)

  TAIWO Ridwan Ademola   (Name of student)

DEDICATION

I dedicate this thesis to my Creator – ALLAH – SubhanaHu wa ta'ala for granting me the opportunity to complete this work. I also dedicate this to my parents – late Alhaji Jamiu Taiwo and Alhaja Ramotullahi Taiwo, my siblings and my partner.

ABSTRACT

Water distribution networks (WDNs) are critical infrastructure systems facing increasing challenges due to frequent failures, which have significant environmental, social, and economic consequences. The failure of water pipes and the subsequent breakdown of WDNs pose significant obstacles to the sustainability and functionality of these vital systems. To mitigate this challenge, it is crucial to gain a comprehensive understanding of the factors contributing to pipe failure and develop predictive models capable of forecasting the probability of failure (POF) of water pipes and their associated causes. Therefore, the primary aim of this study is to improve the current understanding of water pipe failure factors and develop predictive models to enhance the management of WDNs considering four objectives: 1) identify failure factors and failure modes of water pipes, 2) model, rank, and investigate the relationship between water pipe failure factors and failure modes, 3) develop and automate optimized models to predict the POF, probability of leak (POL), and probability of burst (POB) of water pipes, and 4) develop and automate an optimized model to predict the causes of water pipe failure (COF).

To address these objectives, this study employs a rigorous multi-method approach, combining innovative techniques in data analysis and machine learning. A scientometric and systematic review identifies failure factors and modes, while Fault Tree Logic (FTL) establishes a structured framework for analyzing complex factor relationships. The application of Partial Least Squares Structural Equation Modeling (PLS-SEM) investigates the relationship between failure factors and modes using global expert survey data. For predictive modeling, the study introduces novel advanced machine learning techniques. A synergistic integration of Logistic Regression and Genetic Algorithm optimizes POF prediction. For POL and POB prediction, state-of-the-art deep learning algorithms (Deep Neural Networks, Convolutional Neural Networks, and TabNet) are enhanced through Bayesian Optimization, representing a pioneering approach in WDN management. The COF prediction model employs cutting-edge

ensemble learning algorithms (AdaBoost, Random Forest, XGBoost, LightGBM, and CatBoost), optimized using the Tree-structured Parzen Estimator (TPE) algorithm. This comprehensive methodological framework demonstrates the study's rigor and innovative approach to addressing complex WDN challenges.

The systematic review identifies 30 failure factors, categorized into pipe-related, operation-related, soil-related, and external-related factors, along with five distinct failure modes of water pipes. The PLS-SEM model reveals 19 critical failure factors and confirms the hypothesis that factors influencing water pipe failure significantly impact failure modes, as evidenced by p-values less than 0.05 and a path coefficient (β) of 0.567. Using historical data from the Hong Kong (HK) WDN, the POF prediction model achieves remarkable accuracy, with an F1 score of 0.868 and an AUC of 0.944 The POL model achieves high accuracy (0.994) and F1 score (0.924), while the POB model shows similarly impressive results with accuracy of 0.999 and F1 score of 0.872. Both models exhibit strong performance in precision, recall, Matthews Correlation Coefficient (MCC), and Cohen's Kappa, indicating their robust predictive capabilities for leak and burst probabilities in WDNs. The COF model, optimized using TPE, shows significant improvements with macro F1 scores increasing by up to 13%. The optimized XGBoost model achieves the highest accuracy (0.82) and macro precision (0.65), while LightGBM excels in AUC (0.87) and computational efficiency. SHapley Additive exPlanations and feature importance analyses identify water type, material, age, and diameter as key predictive factors for water pipe failure causes.

This research contributes both theoretically and practically to the field of WDNs, providing valuable insights for sustainable management alongside web-based applications for implementing the POF, POL, POB, and COF models developed in this study. By understanding the underlying failure factors, accurately predicting failure probabilities, and forecasting causes of water pipe failure, stakeholders and decision-makers can effectively allocate resources,

prioritize inspections, and implement preventive measures. Ultimately, this study contributes to the performance, reliability, and sustainability of WDNs, ensuring the consistent delivery of clean water to communities.

LIST OF RESEARCH PUBLICATIONS

Journal Articles Relating to This Research (Published)

1. **Taiwo, R.,** Zayed, T. & Ben Seghier, M. E. A. (2024). " Integrated intelligent models for predicting water pipe failure probability". Alexandria Engineering Journal (IF = 6.8, Q1), 86, 243-257, https://doi.org/10.1016/j.aej.2023.11.047

2. **Taiwo, R.,** Yussif, A., Ben Seghier, M. E. A., & Zayed, T. (2024). "Explainable Ensemble Models for Predicting Wall Thickness Loss of Water Pipes". Ain Shams Engineering Journal (IF = 6.0, Q1), https://doi.org/10.1016/j.asej.2024.102630

3. **Taiwo, R.,** Shaban, I. A., & Zayed, T. (2023). "Development of sustainable water infrastructure: A proper understanding of water pipe failure". Journal of Cleaner Production (IF = 11.07, Q1), 398: 136653 https://doi.org/https://doi.org/10.1016/j.jclepro.2023

4. **Taiwo, R.,** Ben Seghier, M. E. A., & Zayed, T. (2023). "Towards sustainable water infrastructure: The state-of-the-art for modeling the failure probability of water pipes". Water Resources Research (IF = 6.16, Q1). e2022WR033256. https://doi.org/10.1029/2022WR033256

5. Farh, H.M.H, Ben Seghier, M. E. A, **Taiwo, R.,** & Zayed, T. (2023). "Analysis and Ranking of Corrosion Causes for Water Pipelines: A Critical Review." npj Clean Water (IF = 11.4, Q1), 6, 65 https://doi.org/10.1038/s41545-023-00275-5

Journal Articles Relating to This Research (Under review)

1. **Taiwo, R.,** Zayed, T. & Adey, B.T. "Explainable deep learning models for predicting water pipe failure." Journal of Environmental Management (IF = 8.7, Q1). (Under review – 1st cycle)

2. **Taiwo, R.,** Zayed, T. & Adey, B.T. "Interpretable ensemble models for predicting causes of water pipe failure." Reliability Engineering and System Safety (IF = 8.1, Q1). (Under review – 1st cycle)

3. **Taiwo, R.,** Zayed, T. Elshaboury, N., & Abdelkader, E. M. "Promoting Sustainable Water Distribution Networks: Modeling of Water Pipe Failure Factors and Modes." Cleaner Engineering and Technology (IF = 5.3, Q1). (Under review – 2nd cycle)

Journal Articles Relating to Infrastructure and Construction Management (Published)

1. **Taiwo, R.,** Bello, I. T., Abdulai, S. F., Yussif, A.-M., Salami, B. A., Saka, A., Ben Seghier, M. E. A., & Zayed, T. (2025). "Generative Artificial Intelligence in Construction: A Delphi Approach, Framework, and Case Study." Alexandria Engineering Journal (IF = 6.2, Q1, top 95th percentile), 116, 672-698 https://doi.org/10.1016/j.aej.2024.12.079

2. Yussif, A., **Taiwo, R.,** Zayed, T., Mohandes, S.R., & Abdulai, S. F. (2025). "Enhancing Sustainable Urban Mobility: Computer Vision Applications in Pedestrian Path Studies". Journal of Traffic and Transportation Engineering (IF = 7.4, Q1, top 97th percentile), Accepted.

3. Muddassir, M., Zayed, T, **Taiwo, R.,** & Ben Seghier, M. E. A. (2024). "Advancing the analysis of water pipe failures: a probabilistic framework for identifying significant factors". Scientific Reports (IF = 3.8, Q1), 14, 19218 https://doi.org/10.1038/s41598-024-69855-w

4. Abdulai, S. F., Nani, G., **Taiwo, R.,** Antwi-Afari, P., Zayed, T., & Sojobi, A. O. (2024). "Modelling the relationship between circular economy barriers and drivers for sustainable construction industry." Building and Environment (IF = 7.4, Q1), 254, 111388 https://doi.org/10.1016/j.buildenv.2024.111388

5. Yussif, A.-M., Zayed, T., **Taiwo, R.,** & Fares, A. (2024). "Promoting sustainable urban mobility via automated sidewalk defect detection." Sustainable Development (IF = 12.5, Q1), [1-21] https://doi.org/10.1002/sd.2999

6. Saka, A., **Taiwo, R.,** Saka, N., Salami, B. A., Ajayi, S., Akande, K., & Kazemi, H. (2024). "GPT models in construction industry: Opportunities, limitations, and a use case validation." Developments in the Built Environment (IF = 8.2, Q1), 17, 100300. https://doi.org/10.1016/j.dibe.2023.100300

7. Akindele, N., **Taiwo, R.,** Sarvari, H., Oluleye, B. I., Awodele, I. A., & Olaniran, T. O. (2024). "A state-of-the-art analysis of virtual reality applications in construction health and safety". Results in Engineering (Q1, IF = 6.0, top 95th percentile), 23, 102382 https://doi.org/10.1016/j.rineng.2024.102382

8. Awodele, I.A., Chan, A.P.C., Mewomo, M.C., Darko, A., Municio, A.M.G., **Taiwo, R.,** Eze, E.C., Awodele, O.A., & Olatunde, N.A. (2024). "Awareness, adoption readiness and challenges of railway 4.0 technologies in a developing economy." Heliyon (IF = 4.0, Q2), 10, e25934 https://doi.org/10.1016/j.heliyon.2024.e25934

9. Abdelkader, E. M., Zayed, T., Elshaboury, N., & **Taiwo, R.** (2024). "A hybrid Bayesian optimization-based deep learning model for modeling the condition of saltwater pipes in Hong Kong." International Journal of Construction Management (IF = 3.9, Q2), https://doi.org/10.1080/15623599.2024.2304392

10. Bello, I. T., **Taiwo, R.,** Esan, O. C., Adegoke, A. H., Ijaola, A. O., Li, Z., Zhao, S., Wang, C., Shao, Z., & Ni, M. (2024). "AI-enabled materials discovery for advanced ceramic electrochemical cells." Energy and AI (CiteScore = 11.8, Q1), 15, 100317. https://doi.org/10.1016/j.egyai.2023.100317

11. Adedara, M. L., Taiwo, R., & Bork, H.-R. (2023). "Municipal Solid Waste Collection and Coverage Rates in Sub-Saharan African Countries: A Comprehensive Systematic Review and Meta-Analysis." Waste, 1(2), 389-413. https://doi.org/10.3390/waste1020024

12. **Taiwo, R.,** Hussein, M., & Zayed, T. (2022). "An Integrated Approach of Simulation and Regression Analysis for Assessing Productivity in Modular Integrated Construction Projects." Buildings (IF = 3.8, Q2), 12: 2018. https://doi.org/10.3390/buildings12112018

Journal Articles Relating to Infrastructure and Construction Management (Under review)

1. Saka, A., **Taiwo, R.,** Saka, N., Oluleye, B., Dauda, J., & Akanbi, L. "Integrated BIM and Machine Learning System for Circularity Prediction of Construction Demolition Waste." Sustainable Cities and Society. (IF = 10.5, Q1), (Under review – 1st cycle)

2. Abdulai, S. F., Nani, G., **Taiwo, R.,** Zayed, T., & Jimoh, A. "A Nexus of Barriers and Strategies for Circular Economy Implementation in the Construction Industry: A Structural Equation Modeling Approach." Cleaner Engineering and Technology (IF = 5.3, Q1), (Under review – 1st cycle).

Refereed Conference Papers

1. **Taiwo, R.,** Ben Seghier, M. E. A., & Zayed, T. (2023). "Predicting Wall Thickness Loss in Water Pipes Using Machine Learning Techniques". In EuroStruct Conference (pp. 1-10). Boku, Austria. September 25-29. https://doi.org/10.1002/cepa.2075

2. **Taiwo, R.,** Wang, K.C.; Olanrewaju, O.I.; Tariq, S.; Abimbola, O.T.; Mehmood, I.; Zayed, T. (2022)." An Analysis of Employee Motivation in the Construction Industry: The Case of Hong Kong" Engineering Proceedings (Cite Score: 0.7) 22: 11 https://doi.org/10.3390/engproc2022022011

ACKNOWLEDGMENTS

I begin with the deepest gratitude to Allah, the Most Merciful and the Giver of all strength, for sustaining me throughout my life and the course of this research. His endless mercy and blessings have been my constant support, without which this thesis could not have been accomplished. As the Holy Quran says, "*And whatever blessing you have, it is from Allah*" (Quran 16:53), I am ever mindful that every opportunity and success I have achieved is by His grace.

I extend my heartfelt thanks to my chief supervisor, Professor Tarek Zayed, who also serves as the Associate Head for Research. His expert guidance and unwavering support have been instrumental in shaping my academic and research journey. Professor Zayed not only guided me meticulously through my academic inquiries but also encouraged me to explore my interests deeply. His supportive presence in every meeting and discussion has been invaluable to my growth and confidence as a researcher. Beyond his role as a mentor, his approachable and encouraging nature has greatly enhanced my experience, making my scholarly pursuit both enjoyable and profoundly rewarding. I also appreciate the Hong Kong Water Supply Department for providing valuable data for my study.

My gratitude also goes to the Department of Building and Real Estate (BRE) at the Hong Kong Polytechnic University and the Innovation and Technology Fund (grant: ITS/033/20FP) for funding my PhD program. I am immensely thankful for the supportive environment provided by BRE staff, particularly Ms. Chloe Sing, Ms. Irene Pang, Mr. K.J. Cheong, and Mr. Ken Chui, whose administrative assistance was crucial throughout my studies. Additionally, I would like to acknowledge all past and present members of the Smart Infrastructure Management Systems (SIMS) research lab. Special thanks to Dr. Salman Tariq, Dr. Sherif Abdelkhalek, Dr. Mamdouh Hussein, Dr. Eslam Abdelkadir, Dr. Nehal Elshaboury, Dr. Eslam Saleh, Dr.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

AADT   : Annual average daily traffic

AC   : Asbestos

AdaBoost   : Adaptive Boosting

AHP   : Analytical Hierarchy Process

ANFIS   : Adaptive network-based fuzzy inference system

ANP   : Analytic Network Process

ANN   : Artificial Neural Network

ARIMAX   : AutoRegressive Integrated Moving Average with Explanatory Variable

AUC   : Area under curve

AVE   : Average variance extracted

BO   : Bayesian Optimization

CatBoost   : Categorical boosting

CFA   : Confirmatory factor analysis

CI   : Cast iron

C-Index   : Concordance Index

CMB   : Common method bias

CML   : Cement mortar lining

CNN   : Convolutional Neural Networks

COF   : Cause of failure

CP   : Cathodic protection

DI   : Ductile iron

DNN   : Deep Neural Networks

EL   : Ensemble Learning

ELM   : Extreme learning machine

FAHP   : Fuzzy Analytic Hierarchy Process

FANP   : Fuzzy Analytic Network Process

FORM   : First Order Reliability Method

FOSM   : First Order Second Moment

FW   : Freshwater

GBM	: Gradient boosting machine
GI	: Galvanized Iron
HBO-DL	: Hybrid Bayesian optimization-based deep learning
HTMT	: Heterotrait-monotrait ratio of correlations
IPMA	: Importance-performance map analysis
LSF	: Limit state function
LSTM	: Long Short-Term Memory
MAE	: Mean Absolute Error
MAPE	: Mean Absolute Percentage Error
MCC	: Matthews Correlation Coefficient
MCS	: Monte Carlo simulation
MIC	: Microbiologically induced corrosion
ML	: Machine Learning
MLP	: Multilayer perceptron
MSE	: Mean Squared Error
OTE	: Ordered Target Encoding
PCCP	: Prestressed concrete cylindrical pipes
PE	: Polyethylene
PLS	: Partial least squares
POB	: Probability of burst
POF	: Probability of failure
POL	: Probability of leak
PVC	: Polyvinyl chloride
RF	: Random forest
RSF	: Random survival forest
SEM	: Structural equation modeling
SHAP	: SHapley Additive exPlanations
SMBO	: Sequential model-based optimization
SVM	: Support vector machine
SW	: Saltwater
TPE	: Tree-Structured Parzen Estimator

UR	: Uncertainty range
UPVC	: Unplasticized Polyvinyl Chloride
VIF	: Variance inflation factor
WDN	: Water distribution networks
WSD	: Water Supplies Department
XAI	: Explainable artificial intelligence
XGBoost	: Extreme Gradient Boosting
σ_y	: yield stress (kPa)
σ_b	: bending stress (kPa)
l_d	: length of corrosion defect (mm)
d_d	: depth of corrosion defect (mm)
p_f	: failure pressure (kPa)
σ_f	: tensile strength (kPa)
σ_o	: initial tensile strength (kPa)
δ	: maximum corrosion rate
b_o	: initial pipe wall thickness (mm)
σ_{uts}	: ultimate tensile strength
v_d	: radial corrosion rate (mm. a^{-1})
P_e	: soil load (kPa)
v_l	: axial corrosion rate (mm. a^{-1})
P_s	: surface load (kPa)
c_d	: cold days (days)
D	: pipe diameter (mm)
h_d	: hot days (days)
t	: pipe wall thickness (mm)
I_l	: interval to last break (years)
$d(T)$	: time-dependent depth of the defect
L	: pipe length
A	: pipe age
T_a	: allowable wall thrust (MPa)

T_{cr} : critical wall thrust (MPa)

ΔX : ring deflection (mm)

σ_m : maximum bending stress (kPa)

P_f : failure probability

Chapter 1

INTRODUCTION

1.1. INTRODUCTION

This chapter introduces the current study. The research background is discussed, followed by the research scope and problem statement. Subsequently, the research objectives and methodologies to fulfill them are expatiated. Finally, the significance of the study and the structure of the thesis are highlighted.

1.2. BACKGROUND

Water is an essential natural resource, which is transmitted from the sources to the point of consumption through different types of pipelines (L. Chen et al., 2021; Gurung et al., 2015). Despite the advancement in the material, manufacturing, and installation methods of these pipelines over time, water distribution networks (WDNs) still face the challenge of failure (Mahmoodian & Li, 2016; K. Pietrucha-Urbanik & Tchórzewska-Cieślak, 2017). Globally, water losses in most WDNs are more than 30%, which has a negative impact on the infrastructure's resilience and financial aspects (Prieto et al., 2015; Tariq et al., 2021). Similarly, a study conducted by Folkman (2018) showed that the failure rate of water pipes increased from 11.0 to 14.0 breaks per 100 miles per year in the USA and Canada within six years (2012 -2018). In the UK, almost 22% of potable water is lost yearly due to the failure of the water pipe system (Farewell et al., 2012). In 2017, more than 2.2 billion m^3 of water was lost in China (Liao et al., 2021). Furthermore, the failure rate in Australia was estimated to be 20 bursts per

100 km per year. In Hong Kong, it was reported that 4953 bursts occurred in private water pipes between 2009 – 2013 (Ombudsman, 2014). In South Korea, 52.5% of the water pipes will require rehabilitation by 2024, as indicated in a study by Seo et al. (2015). In the case of Colombia, a developing country, approximately 50% of water is lost due to water pipe failures (Giraldo-González & Rodríguez, 2020). Additionally, it was reported that about 780 million people have no access to potable water due to the lack of an appropriate water infrastructure system, including the failure of water pipes, and yearly, more than 3.41 million individuals die globally as a result of water-related diseases (Paradkar, 2012). These figures show that water pipe failure is a global issue that needs to be tackled. Therefore, it is not only important to provide water to humanity, but it is equally essential that the water is safe, clean, and consistent in terms of supply.

In addition, the consequences of water pipe failure are numerous. Some of them are property loss, repair and remediation costs, deterioration of human health, damage to the environment, and customer dissatisfaction (Dawood et al., 2021; Fares & Zayed, 2010; Katarzyna Pietrucha-Urbanik & Tchórzewska-Cieślak, 2020). Moreover, the failure of the WDN and its associated system affects global economic conditions. For instance, in 2006, although the original need to maintain and rehabilitate the WDNs in the US was $6 billion, only about $1.2 billion was spent because of the unavailability of funds (Paradkar, 2012). According to the American Water Work Association (AWWA), the USA needs to invest around $1 trillion in replacing and repairing the deteriorating components of their WDNs for the next few decades (Fan et al., 2022). In Australia, the estimated cost of repairing and maintaining WDNs is about AUD 1.4 billion (Weeraddana et al., 2020).

Therefore, a WDN should be developed in such a way that several technical requirements, i.e., functionality, serviceability, and durability (Farshad, 2006), are achieved. To function optimally and sustainably, the material properties of components of a WDN in relation to

mechanical, thermal, and durability, which determine the serviceability state of the network, are highly important. Water pipe failure is a complex phenomenon that depends on various factors. Properly understanding these factors and developing predictive models are crucial to improving the status of WDNs.

1.3.   RESEARCH SCOPE AND PROBLEM STATEMENT

Water Distribution Networks (WDNs) play a critical role in delivering both fresh and saltwater for various human needs, making them essential infrastructure systems. However, these networks often encounter frequent pipe failures, which significantly challenge their efficient operation and maintenance. Understanding the factors that contribute to water pipe failures is paramount for effective management and mitigation strategies.

Various researchers have investigated these factors in the past (Barton et al., 2019; Fares & Zayed, 2010; Shaban et al., 2023). Some researchers have classified the factors into two: internal and external factors (Rajani & Kleiner, 2001; Rostum, 2000), while others have classified them into three: physical/pipe-intrinsic, environmental, and operational materials (Al-Barqawi & Zayed, 2008; Barton et al., 2019). However, scholarly literature lacks comprehensive investigations into the multitude of factors influencing water pipe failure, including their relative importance and intricate interrelationships.

The study of water pipe failure is a complex and multifaceted subject, requiring a holistic approach that considers numerous variables. Each of these factors can significantly impact the structural integrity and longevity of water pipes within a distribution network. Despite their undeniable relevance, a comprehensive understanding of how these factors interact and contribute to pipe failures is still limited, creating a significant knowledge gap in the field. To effectively address the challenges associated with water pipe failures, it is essential to identify the comprehensive list of the influencing factors and comprehend the relative importance and

interrelationships among these contributing factors. By doing so, water utility managers, engineers, and policymakers can develop informed decision-making frameworks, implement targeted maintenance programs, and allocate resources more effectively.

Furthermore, developing optimized predictive models is important to mitigate incessant water pipe failures. Studies have been conducted on establishing predictive models to forecast the number of failures and failure rates in a network (Konstantinou & Stoianov, 2020; Ogutu et al., 2017; Yamijala et al., 2009). However, the existing literature lacks a fully optimized model to predict the probability of water pipe failure. Within the context of managing WDNs, accurately predicting the probability of pipe failure holds greater significance and practical implications compared to simply quantifying the number of failures or failure rates. Such a predictive model would enable water utility managers and decision-makers to move beyond reactive measures and adopt proactive strategies to prevent failures and optimize network performance. By having reliable estimates of failure probabilities, utility managers can make informed decisions regarding asset maintenance and replacement schedules. Instead of relying solely on age-based or reactive replacement approaches, they can prioritize replacements based on the predicted probabilities, focusing on pipes with higher risk levels. This approach maximizes the utilization of existing assets, prolongs their service life, and optimizes operational costs by avoiding premature replacements.

Another dimension in the management of WDNs that has rarely been investigated is the establishment of predictive models to forecast the causes of water pipe failure. Predictive models for pipe failure causes allow for early detection and prevention of potential failures. By analyzing various influential factors, these models can identify patterns and indicators that precede failures. This early detection enables utilities to implement preventive measures, such as targeted inspections, maintenance, and rehabilitation, before the failures occur, reducing the likelihood of service disruptions and minimizing the associated costs. Predictive models can

enable improvements in the design and construction of WDNs. By identifying and predicting the causes of pipe failures, utilities can incorporate this knowledge into the design phase of new network infrastructure or expansion projects. This includes selecting appropriate pipe materials, considering environmental and soil conditions, and implementing construction techniques that mitigate the identified causes of failure. As a result, the network can be designed and built to withstand potential challenges, enhancing its durability and reducing the likelihood of failures in the future.

1.4. RESEARCH OBJECTIVES

In connection with the problem statement highlighted in the previous section and the limitations in the extant literature, the main aim of this study is to improve the current understanding of water pipe failure factors and develop predictive models to enhance the management of WDNs. To fulfill this goal, the specific objectives of this study are identified as follows:

1.   **Identify comprehensive failure factors and failure modes of water pipes**

This objective will answer the following questions: How many factors influence the failure of water pipes? What are these factors? What category does each factor belong to? What are the common water pipe failure modes?

2.   **Model, rank, and investigate the relationship between water pipe failure factors and failure modes**

This objective will address the following research questions: What is the significance of each failure factor based on their categories? What is the interrelationship between the failure factors categories? What is the influence of the failure factors on water pipe failure modes?

3.   **Develop and automate optimized models to predict the probability of failure, leaks, and bursts of water pipes**

This research objective will answer the following questions: How to model the failure probability of water pipes? How to model the probability of leak, burst, and no leak/burst? What are the most significant failure factors to include in the model? What is the contribution of each failure factor to the predictive model?

4. Develop and automate an optimized model to predict the causes of water pipe failure

This research objective will address the following questions: How to model the causes of water pipe failure? What are the most significant failure factors to include in the model? What is the contribution of each failure factor to the predictive model?

1.5. RESEARCH METHODOLOGY

In order to achieve the four objectives stated in the previous section, the following methodology is adopted (see Figure 1.1):

1. A systematic literature review is conducted to identify a comprehensive list of failure factors influencing water pipe failure. Subsequently, water pipe failure modes are reviewed and discussed. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is adopted as a protocol for conducting the review (Objective 1).

2. An analytical model using the structural equation modeling (SEM) technique is developed to rank the failure factors, investigate the relationship between their categories, and examine the influence of the failure factors on water pipe failure modes. The identified factors and failure modes from objective one are used to design a questionnaire survey. The survey is distributed to academics and professionals in the field of WDN, and the data formed the basis of the SEM model (Objective 2).

3. Logistic regression (LR) and Genetic Algorithm (GA) are fused together to develop an optimized model to predict the probability of failure (POF) of water pipes. The GA is

employed to optimize the LR model's hyperparameters and features (i.e., input factors). The contribution of each factor to the optimized predictive model is explained using the SHapley Additive exPlanations (SHAP) framework (Objective 3a). Subsequently, deep learning models are integrated with Bayesian Optimization (BO) algorithm to develop a model to predict the probability of leak, burst, and no leak/burst (Objective 3b).

4. Ensemble state-of-the-art algorithms such as extreme gradient boosting (XGBoost) are combined with Tree-Structured Parzen Estimator (TPE) to develop predictive models for forecasting the causes of water pipe failure (COF) (Objective 4).

1.6. SIGNIFICANCE OF THE STUDY

The findings of this study will provide valuable insights into the failure factors and failure modes of water pipes. By identifying comprehensive failure factors and categorizing them, this study will enhance the understanding of the underlying causes of pipe failure. The developed predictive models will enable WDN managers and professionals to assess the probability of failure, leak, burst, and potential causes, facilitating proactive maintenance and targeted interventions. This, in turn, will lead to improved decision-making processes, efficient resource allocation, and enhanced overall management of WDNs.

1.7. STRUCTURE OF THE REPORT

This thesis is structured into seven chapters. Chapter one introduces the research background, scope, problem statement, research objectives, a summary of the research methodology, and the significance of the study. Chapter two presents a comprehensive literature review of the research area, including previous studies relating to failure factors, failure modes, and predictive models for estimating the probability and causes of water pipe failure. Subsequently, chapter three discusses the research methodology adopted to fulfill the research objectives.

Chapter four presents the data collection for the study. Chapter five shows and discusses the research findings. Chapter six presents the automated applications developed to facilitate easy usage of the research output, while chapter seven gives a summary of the main findings, research limitations, and recommendations for future work.

Objectives	Methods	Outcomes
Identify comprehensive failure factors and failure modes of water pipes	Systematic review + Fault tree logic	Checklist of water pipe failure factors and failure modes
Model, rank, and investigate the relationship between water pipe failure factors and failure modes	Structural equation modeling	1. Significant failure factors and modes; 2. Influence of failure factors on failure modes
Develop and automate an optimized model to predict failure probability of water pipes	LR + GA + SHAP + + DNN + CNN + TabNet + BO	1. An automated model to predict POF; 2. An automated model to predict probability of leak, burst and no leak/burst
Develop and automate an optimized model to predict the causes of water pipe failure	Ensemble learning algorithms	An automated optimized model in form of a web page to predict COF

Figure 1. 1: Research objectives, methods, and outcomes

Chapter 2

LITERATURE REVIEW[1]

2.1. INTRODUCTION

This chapter presents a comprehensive systematic review of previous research relating to the failure of water pipes in WDNs. The chapter starts by reviewing the factors influencing water pipe failure following a systematic approach. Subsequently, the common failure modes of water pipes are reviewed. After that, the different techniques that have been employed to predict the probability of failure of water pipes in the extant literature are reviewed. At the end of the chapter, the research gaps in the literature are highlighted.

2.2. FACTORS INFLUENCING WATER PIPE FAILURE

This section describes the literature retrieval process and methodology employed for conducting the systematic review. A scientometric analysis of previous studies is conducted. Subsequently, the factors influencing the failure of water pipes are systematically reviewed and

---

[1] [1] This chapter is largely based upon:

Taiwo, R., Zayed, T. & Ben Seghier, M. E. A. (2024). " Integrated intelligent models for predicting water pipe failure probability". Alexandria Engineering Journal, 86, 243-257, https://doi.org/10.1016/j.aej.2023.11.047

Taiwo, R., Yussif, A., Ben Seghier, M. E. A., & Zayed, T. (2024). "Explainable Ensemble Models for Predicting Wall Thickness Loss of Water Pipes". Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2024.102630

Taiwo, R., Shaban, I. A., & Zayed, T. (2023). "Development of sustainable water infrastructure: A proper understanding of water pipe failure". Journal of Cleaner Production, 398: 136653 https://doi.org/https://doi.org/10.1016/j.clepro.2023

Taiwo, R., Ben Seghier, M. E. A., & Zayed, T. (2023). "Towards sustainable water infrastructure: The state-of-the-art for modeling the failure probability of water pipes". Water Resources Research. e2022WR033256. https://doi.org/10.1029/2022WR033256

Farh, H.M.H, Ben Seghier, M. E. A, **Taiwo, R.,** & Zayed, T. (2023). "Analysis and Ranking of Corrosion Causes for Water Pipelines: A Critical Review." npj Clean Water, 6, 65 https://doi.org/10.1038/s41545-023-00275-5

categorized into four: pipe-related, operation-related, external-related, and soil-related factors based on the literature synthesis.

2.2.1. Search strategy and framework

2.2.1.1 Literature Retrieval Procedures

The complete literature retrieval and selection process is presented in Figure 2.1. Firstly, a pilot search on the website of WoS was performed to validate the proposed topic. Afterward, to have comprehensive search results (Tawfik et al., 2019), higher content coverage with better availability of various scientometric data, and strict content indexing criteria & inclusion policies (Pranckutė, 2021), the WoS and Scopus databases were selected. Different search strings were used and improved upon until one was finally chosen. The results (i.e., articles) were filtered by subject areas, downloaded, and the duplicates were removed.

The downloaded articles were further subjected to screening by reading their title and abstracts. This is an important step to discard non-relevant and out-of-scope papers. Subsequently, the full text of the articles was evaluated. Inclusion and exclusion criteria were set at this stage. These criteria are important to avoid personal biases (Tariq et al., 2021). The inclusion criteria for this study include research papers focusing on physical, environmental, or operational causes of water pipe failure or factors contributing to the failure of water pipes. Research studies from non-relevant areas, written in a language other than English, and papers with no full-text availability were excluded. As the search strings may not be absolutely perfect, the screened texts were subjected to forward and backward snowballing. The former refers to looking for papers that cite the paper being examined, while the latter refers to checking the reference list of the examined paper to find another relevant research paper(s). After the forward and backward snowballing, 105 articles were selected. The framework for the literature review is depicted in Figure 2.2.
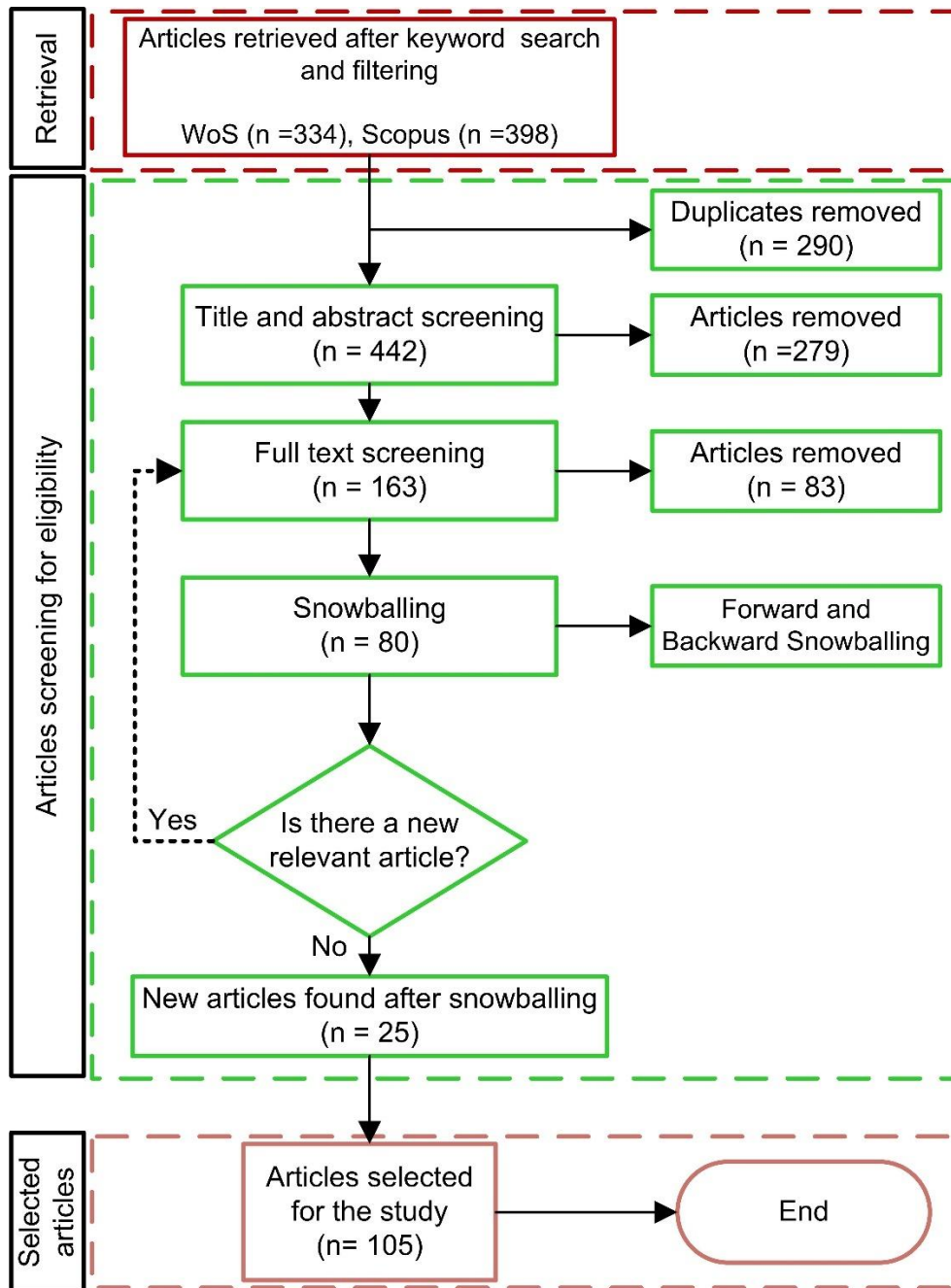
Figure 2. 1: Complete procedures for literature retrieval for investigating failure factors of water pipes
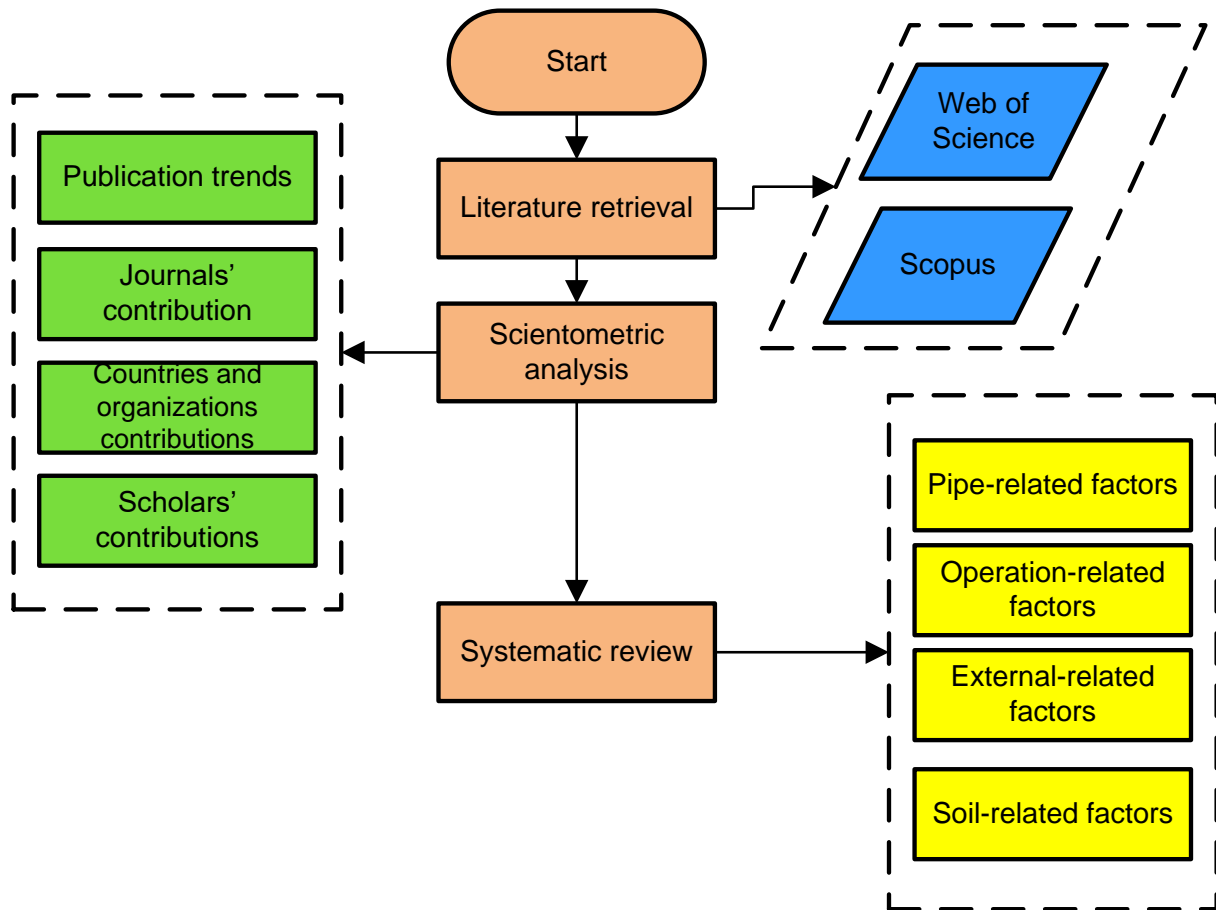
Figure 2. 2: Framework of the literature review regarding factors influencing water pipe failure

2.2.1.2.   **Scientometric analysis**

A scientometric analysis is an approach of objectively mapping a scientific domain with the help of bibliographical data to critically assess the evolution and development of the scientific domain through various indexes (Andriamamonjy et al., 2019; Olawumi & Chan, 2018; G. Wang et al., 2020). VOSviewer was chosen for this study due to its unique features, including easiness in visualizing large maps with the zoom-in feature and well-labeled algorithms (Hussein & Zayed, 2021). The procedures employed for the analysis in this study are discussed in the subsequent section.

2.2.1.3.   **Qualitative analysis**

The qualitative analysis involves a systematic discussion of the factors influencing water pipe failure. After retrieving the articles relevant to the study's objective (i.e., objective 1), each

article was read carefully to understand the causes of water pipe failure. For each article, data regarding the failure factors of water pipe failure, type of pipe material, techniques adopted for developing failure prediction models (where available), and failure modes were extracted and documented in an Excel sheet. Research gaps within the articles were identified and formed the basis for other objectives of the study.

2.2.2. Scientometric review of factors influencing water pipe failure

2.2.2.1. Publication Trends

Figure 2.2 shows the number of publications made annually for the 105 retrieved articles. The publication years ranged from 1982 to 2022. First, a bar chart (Figure 2.3.) shows the number of publications per year. Second, Figure 2.4. represents the number of publications grouped by decade. From these analyses, it is noticed that the study on the causes of water pipe failure did not attract the attention of researchers until the 1980s, despite the availability and usage of water pipes since the 1900s (Clair & Sinha, 2014). Additionally, as one may observe from Figure 3b, studies on the causes of water pipe failure have gained momentum from one decade to another, as 2, 4, 12, and 66% of the papers were published in 1982-1990, 1991-2000, 2001-2010, and 2011-2020, respectively. Although only 12% of the articles were published within 2021-2022, there is the possibility that the number of articles will exceed that of the previous decade by 2030, based on the publication trend.

2.2.2.4. Keyword cluster analysis

Keywords play an important role in a research paper, as they show the paper's main focus, thereby giving potential readers an insight into what the paper entails (Rahman et al., 2020). Out of the 1106 keywords used in the 105 articles, only 122 of them met the VOSviewer criteria of occurring at least three times.
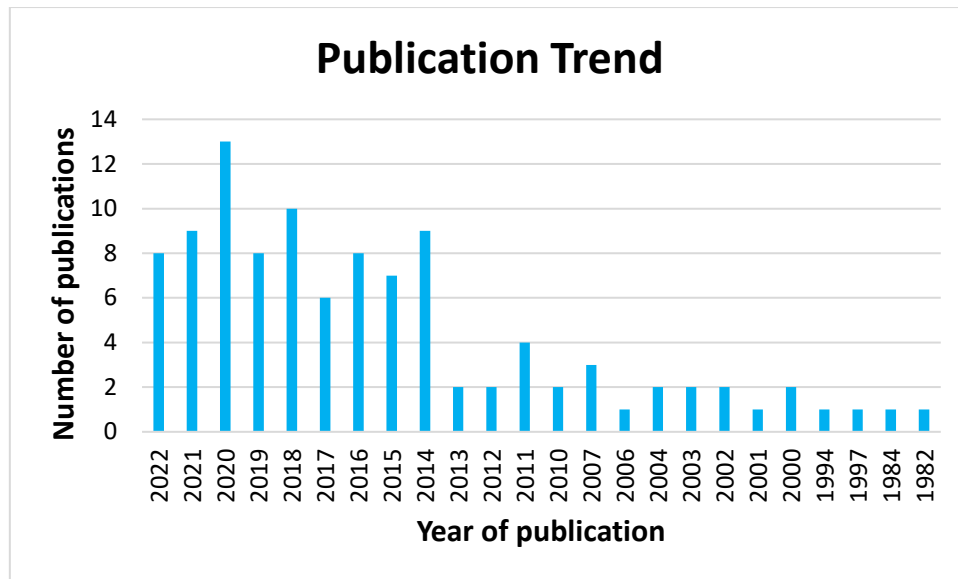
Figure 2. 3: Number of publications per year
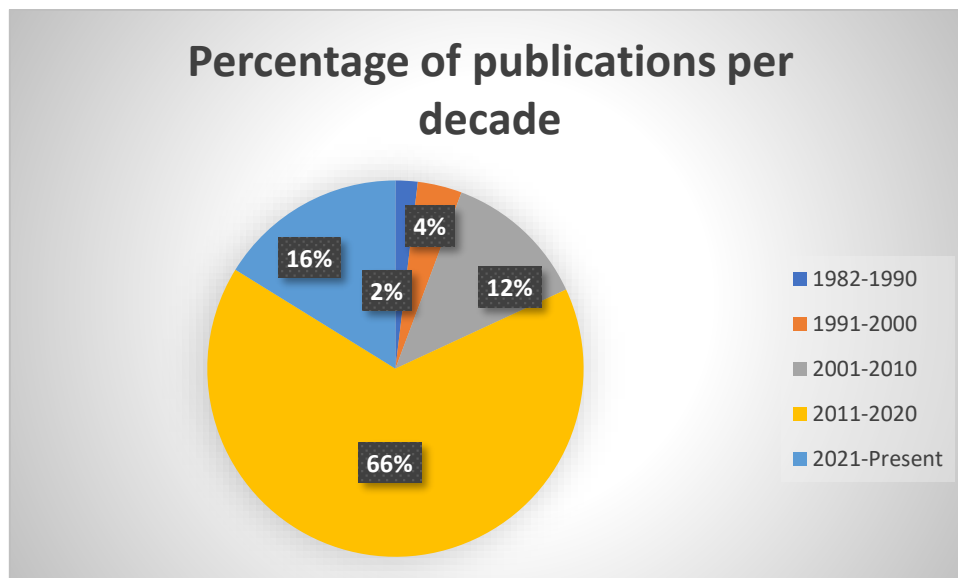
Figure 2. 4: Percentage of publications per decade

The top 20 keywords are shown in Table 2.1. Moreover, some keywords with the same meaning but different wordings were noticed. These keywords were merged. For instance, keywords such as "pipeline failures" were replaced with "pipe failures." A .csv file containing the original keywords and the keywords that replaced them was later imported into the VOSviewer. This approach ensures that the appropriate node sizes for the relevant keywords are generated, which indicates the number of occurrences of each keyword.

From Figure 2.5., it is observed that "corrosion" has the biggest node size in terms of the keywords that depict the factors contributing to the failure of water pipes, which implies that corrosion is the most common focus of the retrieved articles. Additionally, "soils" could be deemed a causal factor of "corrosion" as they are clustered together. However, "soils" have a smaller node, which means that the soil as a cause of corrosion has not been fully studied. It can also be inferred that the most common type of pipe that has been studied is "cast iron pipe."
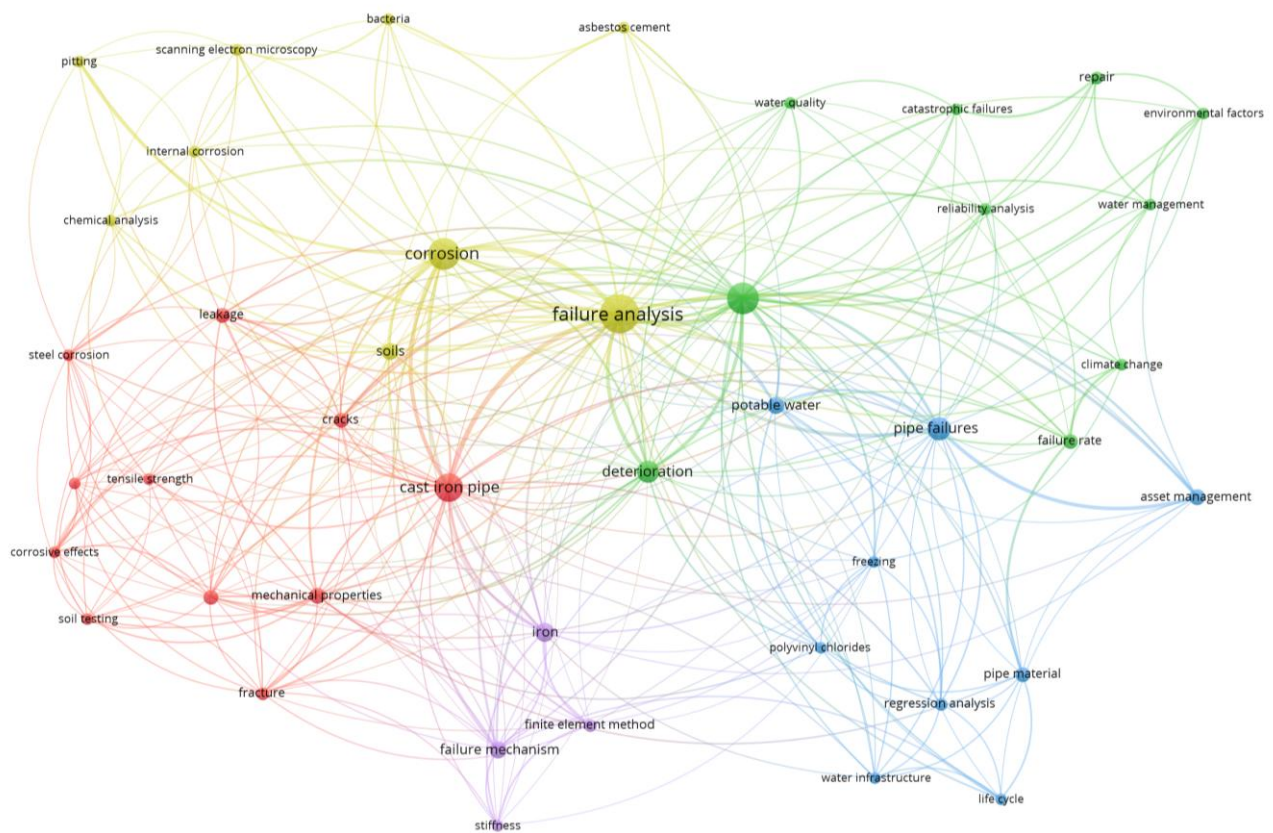
Figure 2. 5: Co-occurrence of keywords network map

According to Table 2.1, the keywords with the lowest occurrence and total strength link are "pipe material," "fracture," "bacteria," and "climate change," indicating a low level of research in these sub-domains of knowledge, as related to causes of water pipe failure.

Table 2. 1: Keyword occurrences

Keyword	Occurrences	Total link strength
Failure analysis	42	39
Water distribution networks	30	26
Corrosion	27	24
Cast iron pipe	19	19
Pipe failures	16	18
Deterioration	14	15
Iron	11	12
Failure mechanism	9	11
Potable water	8	11
Soils	7	7
Asset management	7	6
Mechanical properties	6	6
Cracks	5	5
Failure rate	5	5
Leakage	5	5
Pipe material	5	4
Environmental factors	5	4
Fracture	4	4
Bacteria	4	4
Climate change	3	3

2.2.2.3. Contributions of journals

The contributions of the research outlets are discussed in this section. This is important to show

the most productive research outlets in terms of the different considered criteria and can guide

researchers in making decisions based on these criteria. Using the VOSviewer software, the

minimum number of documents for a source was set to 1, while the minimum number of

citations was set to 25. There is no limit on this threshold setting in scientometric analysis

(Tariq et al., 2021). Out of the 49 sources, 14 met this threshold, as shown in Table 2.2.

In terms of the number of citations, "Canadian Journal of Civil Engineering" is ranked as the

most productive journal outlet, despite the fact that the outlet only published four articles used

in this study compared to "Engineering Failure Analysis," which has 10 articles. The same trend

is noticed in the total link strength, as presented in Table 2.2. This means that the nature of

studies published by the "Canadian Journal of Civil Engineering" has the highest impact and significance in the domain, as they have received the highest recognition from researchers in terms of citations.

Table 2. 2: Contribution of research outlets

Name of research outlets	No. of documents	Citations	Total link strength
Canadian Journal of Civil Engineering	4	292	12
Engineering Failure Analysis	10	284	8
Journal of Water Supply: Research and Technology	3	256	6
Advances in Water Resources	1	230	2
European Journal of Operational Research	1	222	6
Journal of Infrastructure Systems	2	215	6
Journal of Pipeline Systems Engineering And Practice	6	209	5
Canadian Geotechnical Journal	1	197	6
Journal of Materials In Civil Engineering	2	134	4
Procedia Engineering	2	102	4
Computer-Aided Civil And Infrastructure Engineering	1	61	2
Arabian Journal for Science and Engineering	1	41	0
Journal of American Water Works Association	2	39	0
Journal of Hydroinformatics	1	37	2

2.2.2.4.   Contributions of countries and organizations

The contributions of countries and organizations in the domain of the factors influencing water pipe failure are discussed in this section. In terms of the number of publications and total link strengths, Canada is the most productive, with 24 documents and 38 link strengths (see Table 2.3), followed by the United States of America and Australia; while the United Kingdom and Turkey are the least productive countries with 2 documents each.

Consequently, the productivity of Canada in the research domain was also confirmed by the analysis of the organizations' contributions, as four organizations from Canada are among the top 5 organizations (see Table 2.4).

Table 2. 3: Productivity of countries

Country	No. of documents	Citations	Total link strength
Canada	24	925	38
United States	20	239	17
Australia	10	122	15
China	8	104	11
Poland	7	93	4
United Kingdom	2	88	4
Turkey	2	54	6

This implies that Canadian researchers and the government paid more attention to the failure of water pipes than other countries and that Canada has the highest potential for future collaboration with other countries in related research. Additionally, the enormous effort of Canada in this domain can be traced to annual financial loss estimated at one billion dollars, generated by water loss due to failed pipeline systems in Canada (Renzetti & Dupont, 2013).

Table 2. 4: Organizations' contributions

Organization	Documents	Citations	Total Link Strength
National Res. Council, Canada	3	323	9
University of British Columbia, Canada	2	246	9
University of Johannesburg, South Africa	1	175	0
Concordia University, Montreal, Canada	1	102	5
University of Toronto, Canada	1	96	2

2.2.2.5.   Science mapping of scholars

Citation and co-citation analyses of researchers were conducted. These analyses indicate scholars that are actively working in the scientific domain: factors/causes of water pipe failure. The citation analysis indicates four clusters, typifying Rajani B., Sadiq R., and Kodikara J. as the most productive researchers in terms of the number of publications. In terms of citations, Rajani B., Sadiq R., and Tesfamariam are the most cited scholars, with 497, 389, and 260

citations, respectively (see Table 2.5). This analysis also indicates the existence of a strong relationship between the two top scholars – Rajani B. and Sadiq R. – as their nodes are close to each other, suggesting that they often collaborate. Thus, collaboration can increase the visibility of researchers' output.

A co-citation analysis of the authors was also conducted. In this context, co-citation refers to two authors that are cited together by a third author (Surwase et al., 2011). Rajani B. has the highest total link strength, at 1645, which could indicate how often the researcher collaborates with other researchers. As shown in Table 2.6, other top scholars include Kleiner Y., Sadiq R., Tesfamariam S., and Kodikara J., with a total link strength of 1203, 1148, 952, and 634, respectively. These analyses will assist individuals or organizations seeking consultations on the causes of water pipe failure to identify the top researchers in the domain.

Table 2. 5: Citation analysis of scholars

Author	Documents	Citations	Total link strength
Rajani B.	4	497	29
Sadiq R.	4	389	26
Tesfamariam S.	3	260	21
Francisque A.	2	224	17
Seica M.V.	3	201	28

Table 2. 6: Co-citation analysis of scholars

Author	Citations	Total link strength
Rajani, B.	273	1645
Kleiner, Y.	188	1203
Sadiq, R.	81	1148
Tesfamariam, S	72	952
Kodikara, J.	59	634

2.2.3. **Systematic review of factors influencing water pipe failure**

As highlighted earlier, the factors influencing water pipe failure are grouped into four main classes: pipe-related, operation-related, external-related, and soil-related factors. This section provides a comprehensive discussion of these factors.

2.2.3.1 **Pipe-related factors**

(I) Pipe age

Pipe age is a significant factor influencing water pipe failure, but its relationship with failure rates is complex and not always linear. While many studies show that older pipes generally have higher failure rates, some research indicates inverse relationships or no correlation (Ellison & Spencer, 2016; Fares & Zayed, 2010; Jun et al., 2020). For instance, it was noticed that the pipe failure rate increases as the pipe's age increases until they reach 50 –60 years. The pipes in this age bracket (50 –60 years) exhibited the highest failure rate. Even pipes that had been in service for 60–130 years had a significantly lower failure rate (Hekmati et al., 2020). This suggests that age alone isn't determinative; other factors like material quality and installation techniques also play crucial roles. The general conclusion is that while pipes typically deteriorate over time, the impact of age on failure rates can vary depending on other contributing factors.

(II) Pipe diameter

Pipe diameter influences water pipe failure rates, with most studies indicating an inverse relationship between diameter and failure frequency (Bruaset & Sægrov, 2018; Hekmati et al., 2020; Kutyłowska & Hotloś, 2013; Kutyłowska & Orłowska-Szostak, 2016; Katarzyna Pietrucha-Urbanik, 2015; Rajeev et al., 2014; Rezaei et al., 2015; Singh, 2011; Zywiec et al., 2019). Larger diameter pipes (>300 mm) generally exhibit lower failure rates due to their thicker walls, which provide greater structural integrity and resistance to pressure fluctuations.

For instance, Rezaei et al. (2015) found that pipes with diameters over 600 mm have shown up to 87.5% fewer bursts compared to those 100 mm or less. However, some research suggests that very large pipes (e.g., 1050 mm) might have slightly lower safety factors than moderately large ones (e.g., 750 mm) due to increased hoop and axial stresses (D. Wilson et al., 2017). Despite this, the overall trend indicates that larger diameter pipes are less prone to failure than smaller ones.

(III) Pipe length

Pipe length has been shown to influence water pipe failure rates, though the relationship is not always straightforward. Several studies, including Vipulanandan et al. (2012) and Zamenian et al. (2017), have found a positive correlation between pipe length and failure rate, particularly for smaller diameter pipes. This is often attributed to the increased number of joints and service connections in longer pipes, which can be points of vulnerability. However, Andreou et al. (1987) argued against a simple linear relationship, suggesting that localized factors like corrosion might have more impact on failure rates than overall length.

(IV) Pipe wall thickness

The failure of water pipe is impacted by its wall thickness, with thicker walls generally associated with lower failure frequencies (Bruaset & Sægrov, 2018; Gao, 2017; K. Liyanage & Dhar, 2018). Wilson et al. (2017) found that safety factors increase with wall thickness, independent of pipe diameter. For instance, a 1500 mm thick wall had a safety factor of 9.8, compared to 2.0 for a 500 mm wall. Chang et al. (2021) corroborated this, showing that thinner walls led to increased settlement and breakage. These findings indicate that thicker pipe walls provide better resistance to stresses, thereby reducing failure likelihood in WDNs.

(V) Pipe material and manufacturing flaw

Pipe material is a critical factor that contributes to pipe failure, as highlighted by numerous studies (Almheiri et al., 2020; Fares & Zayed, 2010; García et al., 2018; Pouri & Heidarimozaffar, 2022; Rezaei et al., 2015; Zamenian et al., 2017). Cast iron (CI) pipes typically exhibit higher failure rates due to corrosion susceptibility, accounting for up to 40% of total failures in some networks (Folkman et al., 2012; Rezaei et al., 2015). While ductile iron (DI) and steel pipes offer improved strength, they still require protection against corrosion. Asbestos cement (AC) and concrete pipes provide good pressure resistance but are prone to degradation through lime leaching (Al-Adeeb & Matti, 1984).

Plastic pipes, primarily polyethylene (PE) and polyvinyl chloride (PVC), generally demonstrate lower breakage rates due to their corrosion resistance, flexibility, and ability to absorb impact loads (Clair & Sinha, 2014; Rezaei et al., 2015). However, they can be vulnerable to degradation by certain organic substances and fatigue failure. Manufacturing flaws such as porosity, cold shuts, and inclusions can significantly reduce pipe strength across all materials. For example, Makar (2000) found that inclusions can lower CI pipe tensile strength from 130 to 33 MPa. When selecting pipe materials, it's crucial to consider local conditions, installation methods, and potential chemical interactions to minimize failure risks in water distribution networks.

(VI) Protection efficiency and third-party damage

Pipe lining and coating, introduced in the 1920s, have proven effective in reducing failures, with Lee (2011) finding that unlined cast iron pipes failed four times more frequently than lined ones. Other protective methods include cathodic protection and polyethylene wrapping (Kleiner et al., 2004; Rajani & Kleiner, 2003). Third-party damage, such as accidental damage during construction or careless handling during transportation, can also lead to pipe failures (Farshad, 2006). For plastic pipes, prolonged exposure to ultraviolet light can increase

brittleness and susceptibility to failure (Tran et al., 2021). These factors underscore the importance of proper protection, handling, and installation practices in minimizing water pipe failures.

(VII) Buried depth

Studies consistently show an inverse relationship between buried depth and breakage rate of water pipes (Jun et al., 2020; Trickey et al., 2016; Daniel Wilson et al., 2015). Shallower pipes experience higher soil pressures, leading to increased stress and failure risk. Trickey et al. (2016) found that reducing the burial depth of cast iron pipes from 2 m to 1.5 m increased the longitudinal bending moment by over 140%, resulting in circumferential fractures. Wilson et al. (2015) observed that pipes at 0.5 m depth failed nearly ten times more frequently than those at 1.5 m, even with identical diameters.

2.2.3.2.   **Operation-related factors**

(I) Internal water pressure

Numerous studies have established a direct relationship between pressure and breakage frequency (Grigg, 2017; Jiang et al., 2019; Kabir et al., 2015; Poojitha & Jothiprakash, 2022; Tang et al., 2019a). Pressure fluctuations create additional hoop stress, leading to longitudinal or circumferential failures (Rezaei et al., 2015). Ellison & Spencer (2016) found that high-pressure asbestos cement pipes failed 2.5 times more often than low-pressure ones. However, the impact of pressure can vary depending on network design and local conditions. For instance, Martínez García et al. (2020) observed strong correlations between high pressure and failure rates in mountainous districts but weaker or negative correlations in non-mountainous areas, suggesting that other factors may dominate in certain regions.

(II) Water quality

Water quality, encompassing aesthetical, microbiological, and physicochemical properties (Kabir et al., 2015; Lee, 2011), impacts both water potability and pipe failure rates. As part of water quality-relating factors, corrosive substances, water temperature, and velocity are discussed in this sub-section.

Corrosive substances: Corrosive substances in water, such as chlorine, dissolved salts, and minerals, can degrade pipe materials and lead to failure (Rehan Sadiq, 2010). Kanakoudis (2004) observed that high iron concentrations in a Greek WDN caused water discoloration and accelerated pipe corrosion, ultimately resulting in pipe failures.

Water temperature: Due to low viscosity, an increase in the temperature of water leads to an increase in its diffusivity, thereby activating the transfer of electrons, which can accelerate the corrosion process. The study by Jun et al. (2020) implies that an increase in the water temperature tends to increase the corrosivity of the water, negatively impacting the pipe.

Water velocity: Water velocity impacts pipe failure rates, particularly in small-diameter pipes. Lower velocities in these pipes can lead to sediment accumulation and bacterial growth, increasing the likelihood of failure (Dao et al., 2021; Gao, 2017). Hence, the velocity of water should be properly controlled.

(III) Water pH

The pH of water has also been found to be a cause of water pipe failure (Arriba-Rodriguez et al., 2018). This sub-section discusses water acidity and alkalinity as they relate to water pipe failure.

Water acidity: Acidic water (pH below 7) can be especially detrimental to pipe integrity (Shull, 2021). Water with pH values between 6.5 and 9.5 has been referred to as safe drinking water (United Utilities Water Limited, 2019). Acidic and soft water attacks can leach calcium silicate

hydrate and lime from concrete and asbestos cement pipes, weakening their structure (Hu & Hubble, 2007). Zraick et al. (2019) observed higher corrosion rates (20-40 μm/year) in water pipes when pH values fell below 7.5. These findings highlight the importance of maintaining appropriate pH levels in water distribution systems to minimize internal corrosion and subsequent pipe failures.

Water alkalinity: Jun et al. (2020) found a positive correlation between water alkalinity and pipe failure rates. Alkaline conditions can promote corrosion by facilitating dissolved oxygen reactions (Arriba-Rodriguez et al., 2018) and providing a favorable environment for sulfate-reducing bacteria, which can lead to microbiologically influenced corrosion (MIC) (Doyle et al., 2003).

(IV) Number of leaks

The Leak-Before-Break (LBB) concept suggests that pipe leakage typically precedes catastrophic failure (IAEA, 1993; Wilkowski, 2000). Rathnayaka et al. (2017) confirmed this in laboratory tests on large-diameter cast iron pipes with simulated corrosion patches. Studies have shown that the frequency of pipe breakage increases with the number of leaks (Ma et al., 2022; Pouri & Heidarimozaffar, 2022). Additionally, leaking pipes can erode surrounding soil, disturbing bedding conditions and potentially inducing excessive stress on pipe walls.

(V) Water hammer

Water hammer, caused by sudden changes in water flow velocity, creates pressure shockwaves in pipes that can result in leaks, cracks, bursts, and fitting separations. The severity depends on pipe characteristics and flow conditions (Roy & Pijush, 2014). In the USA and Canada, 25% of all the annual failures of water pipes are associated with water hammers (Leishear, 2019).

(VI) Installation and pump operation

Poor installation practices and improper pump operation contribute significantly to water pipe failures. Installation issues include not following guidelines, inadequate support, and faulty joints, with 16% of PVC pipe failures attributed to joint defects (Burn et al., 2005). Improper pump operation, such as overloading, excessive pressure, using incorrectly sized pumps, and inadequate maintenance, can also damage pipes and lead to failures (Barton et al., 2019).

(VII) Maintenance practices

Another operation-driven problem in WDN is the lack of effective maintenance practices. Neglecting regular upkeep can lead to various issues, causing pipe failures. These include failing to replace worn components, ignoring water quality problems, insufficient pressure adjustments, lack of regular inspections, and poor repair procedures (Barton et al., 2019).

2.2.3.3. **External-related factors**

(I) Climate-related factors

The link between climate-related factors such as temperature, frost, precipitation, and water pipe failure is discussed in this section. This includes discussion relating to the performance of water pipes in various seasons, especially the winter and summer seasons.

Temperature: Temperature influences water pipe failure, with numerous studies establishing an inverse relationship between temperature and failure frequency(Makar et al., 2001; Tran et al., 2021; X. Wang et al., 2019; Zamenian et al., 2017). As temperatures drop, pipes contract, potentially leading to tensile stress and circumferential breaks if the pipe's strength is exceeded (Habibian, 1994). A study in Norway found an 86% increase in failure rates when temperatures decreased from 23°C to -15°C (Bruaset & Sægrov, 2018).

Winter seasons generally see higher failure rates due to increased earth loads from freezing and water expansion (García et al., 2018; Hekmati et al., 2020; Kutyłowska & Orłowska-Szostak,

2016). However, summer can also present challenges, with some UK regions reporting high breakage rates due to dried and shrinking soil (Gao, 2017). It's important to note that temperature alone isn't the sole indicator of pipe failure, as the relationship between temperature drops and breakage rates isn't strictly linear Habibian (1994). The accumulation of damage over time creates weak points in pipes, which can fail at varying temperature thresholds, leading to cyclical patterns of failure as temperatures fluctuate.

Precipitation: The amount of precipitation in a season can directly affect the amount of moisture in the soil. High rainfall periods increase clay soil swelling, causing additional stress on pipes. Studies in Australia found the highest pipe failure rates during peak rainfall months (July-September), with similar correlations observed elsewhere (Hekmati et al., 2020; Rak et al., 2021).

Frost action: When frost penetrates the soil or ground, it causes soil movement, which induces additional stress on water pipes (Tang et al., 2019a; Zywiec et al., 2019). As water freezes, its volume expansion is constrained by pipe walls, generating ice expansion pressure. The extent of damage depends on factors such as pipe stiffness, diameter-to-wall thickness ratio, water solidification rate, and pipe material characteristics. This process can lead to elastic-plastic deformation of pipes, with repeated freeze-thaw cycles potentially causing cumulative damage. However, some materials like Polypropylene Random Copolymer (PPR) plastic pipes can withstand these cycles without permanent damage, provided aging and fatigue factors remain constant (Zheng et al., 2020).

(II) Biological and chemical-related causes

Biological and chemical activities also influence the corrosion of water pipes. Microbiologically induced corrosion (MIC), lime leaching, and concentration of soluble salts or chemical substances as related to water pipe failure are explained in this section.

Microbiologically induced corrosion (MIC)*:* It occurs predominantly with partial oxygen supply and is driven by sulphate-reducing bacteria (SRB) Alkaline environments can promote MIC by facilitating anaerobic bacterial growth (Doyle et al., 2003; Seica et al., 2002). San et al. (2012) carried out an experimental study to examine the effects of two bacteria: Aeromonas salmonicida and Delftia acidovorans, on the corrosion of water pipes. The inoculation of these bacteria on the water pipe was observed to decrease the mass of the steel water pipe by 1.86 and 2.01 ug, respectively, due to the formation of corrosion products on the pipe. In asbestos cement pipes, various bacteria types can form biofilms, reducing structural strength and promoting pitting corrosion (D. Wang & Cullimore, 2010).

Lime leaching: Lime leaching is a process affecting asbestos cement (AC) and concrete water pipes. It involves the loss of free lime (portlandite) from the pipe material, which can be detected by pH changes. Portlandite protects calcium silicate hydrate, a key component for concrete strength. While carbonation can initially increase pipe strength by filling pores with calcium carbonate, continued leaching eventually degrades the pipe structure. Studies have shown that used AC pipes have significantly lower free lime content than new pipes, indicating ongoing leaching in buried pipes. Factors influencing this process include pipe permeability, water aggressiveness, and soil carbon dioxide levels. Autoclaved AC pipes generally show better chemical stability than water-cured pipes (Al-Adeeb & Matti, 1984; Gong et al., 2016).

Chemical substances: High concentrations of sulphides and chlorides are primary corrosion causes for metallic pipes (Rezaei et al., 2015; Vipulanandan et al., 2012). Concrete pipes are vulnerable to sulphate attacks, both conventional and thaumatic. While PVC and PE pipes resist electrochemical corrosion, they can degrade when exposed to aggressive chemicals like chlorine, which leaches out protective antioxidants (Mikdam et al., 2017). Soil composition also plays a role, with substances like pyrite accelerating corrosion processes (Pękala & Pietrucha-Urbanik, 2018).

(III) Location-based causes

Traffic: Traffic load can cause the buried water pipes to shear or collapse due to the differential loading from passing vehicles (Farewell et al., 2012; H. Nguyen et al., 2022). Moerman et al. (2016) found higher failure rates at road crossing sections compared to bump sections, though pipe age and diameter were more influential factors. Mackey et al. (2014) noted that traffic loading can induce cyclic fatigue effects, contributing to pipe failure.

Land use: Land use such as commercial, industrial, or residential use of land can link with pipe failure as some dynamic loads can be transmitted to the pipes from these structures. Andreou et al. (1987) identified commercial land use as a contributor to pipe failures. However, the relationship between land use and pipe failure is currently understudied, suggesting a need for further research.

2.2.3.4.   **Soil-related factors**

Soil corrosivity, which entails soil resistivity, aeration, moisture content, pH, and soluble salts present in the soil, is arguably one of the most critical factors alongside bedding conditions that influence water pipe failure (Demissie et al., 2016). The complex and heterogeneous nature of soil makes corrosivity difficult to assess accurately (Pritchard et al., 2013). While the Ductile Iron Pipe Research Association (DIPRA) developed a 10-point scoring method to classify soil corrosivity, it has limitations in representing the true severity of corrosion (Ductile Iron Pipe Research Association, 2017; Najjaran et al., 2006a), Metallic pipes primarily experience two types of corrosion: graphitization (mainly in cast iron pipes) and corrosion pitting (in all metallic pipes). Corrosion occurs when two metallic components contact each other in a corrosive environment, leading to electron exchange and subsequent degradation of the metal with lower electric potential (Lee, 2011).

Soil resistivity: Generally, as seen in previous studies (Najjaran et al., 2006a; Pritchard et al., 2013), corrosivity increases with a decrease in soil resistivity. Studies show that soil samples linked to iron-based pipe failures typically have resistivity values between 1000-3000 ohm-cm (Seica et al., 2002). Although resistivity is an important factor that determines soil corrosivity, it cannot be solely used as an absolute index for corrosion (Arriba-Rodriguez et al., 2018).

Soil type: Clay soils pose problems due to shrink-swell behavior, increasing external loading and creating non-uniform support (Mackey et al., 2014). Soft organic soils are unsuitable for bedding due to high moisture retention and low bearing capacity (Pritchard et al., 2013). Sandy soils generally contribute less to pipe failure, with better drainage and higher resistivity (Doyle et al., 2003), though they may pose erosion risks in certain locations.

Soil pH: Some studies have shown a correlation between water pipe failure and pH, while others have shown otherwise (Pritchard et al., 2013). Hou et al.'s (2016) experiment demonstrated higher corrosion rates in more acidic soil solutions. However, Rajani & Makar (2001) found no clear relationship between them. This suggests that while pH can influence pipe corrosion, it's not a standalone indicator of soil corrosivity. Other factors may dominate in different situations, emphasizing the need for comprehensive soil analysis when assessing corrosion risk in water distribution systems.

Soil moisture content: Soil corrosivity increases with an increase in moisture until an optimum moisture content is reached, which then declines afterward (Murray & Moran, 1989). Noor & Al-Moubaraki (2014) found that corrosion rates in Saudi Arabian soils increased with moisture content up to 10%, then decreased. Another study reported a higher optimum range of 50-65% for maximum corrosion rates (Pritchard et al., 2013). These varying results suggest that the optimal moisture content for corrosion differs across locations, depending on other environmental factors.

Soil aeration: Soil aeration, the oxygen exchange between the atmosphere and soil, can influence water pipe corrosion. Differential oxygen concentrations create cells that contribute to corrosion (Wasim et al., 2018). Excavation can increase aeration (Arriba-Rodriguez et al., 2018), while redox potential measures aeration levels, with low values indicating low oxygen (Fiedler et al., 2007). However, the direct relationship between soil aeration and pipe failure remains understudied, highlighting a need for further research.

Bedding conditions and backfilling: Materials that have a low bearing capacity, such as organic soil, and a high potential for shrinking/swelling, and water content retaining capacity, such as clay, have been found inadequate for bedding materials (Hu & Hubble, 2007; Pritchard et al., 2013). In the study of Affolter et al. (2018), it was stated that a 12-year-old pipe failed due to its poor bedding condition and ageing. For example, replacing frozen native soil with granular backfill can significantly increase pipe deflection and moment. Proper selection of bedding and backfill materials, considering soil properties and compatibility, is essential for ensuring long-term water pipe performance and preventing failures in WDNs.

2.3. WATER PIPE FAILURE MODES

This section presents a review of the common failure modes of water pipes in the existing literature. The failure modes reported by the 105 articles are studied and discussed in this section. Failure mode describes the exact manner in which the pipe fails rather than the cause of its failure. This largely depends on the pipe material and its diameter (Makar, 2000; Rajeev et al., 2014). The most common forms of water pipe failure are discussed in this section. Table 2.6 shows the schematic representation of the failure modes with their common causes and typical pipe material and sizes that exhibit the failure modes.

Circumferential cracking*: This occurs when the pipe fails at its circumference. The circumferential mode of failure occurs due to the development of bending moments on the

pipe, which may be attributed to external forces (Makar, 2000). Inadequate bedding conditions, frost penetration, and backfill, which may be sources of external forces, play important roles in this particular mode of failure (Grigg, 2017; K. T. H. Liyanage & Dhar, 2017; Trickey et al., 2016).

This mode of failure is the most common mode in grey cast iron (CI) pipes (Makar, 2000). Due to the development of tensile hoop stress, CI pipes with large diameters usually have circumferential cracking as their failure mode. The axial stress developed in large-diameter pipes is usually significantly lower than the hoop stress (Daniel Wilson et al., 2017). The failure mode is also common in asbestos (AC) and metallic pipes with a diameter of less than 200 mm (Barton et al., 2019).

Longitudinal Cracking: This is the failure mode that describes the failure of a pipe at its longest side. The failure mode occurs when a longitudinal crack appears on the pipe. This could result from compressive forces acting along the pipe or internal water pressure (Makar et al., 2001). In addition, longitudinal cracking can occur on pipes due to traffic loads, causing the development of cross-sectional tension; radial tension caused by water pressure (Farewell et al., 2012); expansion due to frozen water, causing the development of cross-sectional loads; and ground loads, causing the development of cross-sectional loads (Mora-Rodríguez et al., 2014).

This failure mode is common in small and large-diameter CI pipes. For instance, the data presented in the study by Ji et al. (2020) showed longitudinal fracture as the prominent failure mode of small-diameter pipes. Longitudinal, circumferential, piece blown out, and fitting failures were responsible for 53, 21, 14, and 5% of the failures, respectively (Ji et al., 2020). On the other hand, out of four failure modes examined by Rajeev et al. (2014), the authors found that large-diameter pipes failed mostly via longitudinal cracking and bell splitting.

Bell Splitting*:* This is a failure mode that occurs when the pipe's bell (at the joint) splits due to differences in the coefficient of thermal expansivity of the pipe and the joint material (Makar et al., 2001; Makar, 2000). It should be noted that this mode of failure is also common in small-diameter pipes and is different from longitudinal cracking in that its (bell-splitting) cracking ends just below the bell. The common joint material in metallic pipes, leadite, has a coefficient of thermal expansion that differs from that of metallic pipes. This difference could cause stresses in the pipe, which could contribute to the pipe's failure (Makar, 2000).

Corrosion pitting: The localized form of corrosion that occurs on the inner surface of a water pipe, creating pits or small holes, is referred to as corrosion pitting (Rajeev et al., 2014). The formation of corrosion pits weakens the pipe and thus facilitates water loss (Rajani & Tesfamariam, 2005). Various factors, including poor water quality, the presence of corrosive substances such as high levels of chlorides and sulphates in water, and the presence of micro-organisms such as fungi and bacteria, are responsible for corrosion pitting in water pipes (Sadiq et al., 2004).

Pitting corrosion is a common type of failure mode in both large and small-diameter metallic pipes. A study conducted by Pękala & Pietrucha-Urbanik (2018) indicated that corrosion pitting, fitting failure, and mechanical damage accounted for 51%, 26%, and 6% of all the water pipe failures, respectively, in a network.

Blown-out hole: This form of failure mode is the occurrence of a large hole or rupture on a water pipe due to the sudden release of pressure. The impact of the water pressure becomes more severe on the pipe, particularly when the pipe has been compromised by other faults such as corrosion pitting, mechanical damage, and manufacturing defects, among others (Ji et al., 2020; Tang et al., 2019a). In the case of corrosion pitting, the localized corrosion weakens the pipe, and the pressure blows out the remaining thin pipe wall. In 2014, Rajeev et al. (2014)

found that a blown-out hole is the major failure mode in five Australian utilities. Previous studies, such as the ones conducted by Makar et al. (2001) & Makar (2000), also confirm this type of failure mode in water pipes. This failure mode has been found to be dominant in CI and steel pipes (Rajeev et al., 2014).
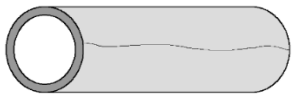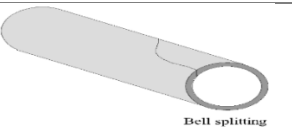
2.4. TECHNIQUES FOR PREDICTING THE PROBABILITY OF FAILURE OF WATER PIPES

This section describes the literature retrieval process and methodology employed for conducting the systematic review. A scientometric analysis of previous studies is conducted. Subsequently, the techniques used for predicting the probability of failure in extant literature are systematically reviewed. Nvivo software is utilized for easy data synthesis during the systematic review. The framework for the review is shown in Figure 2.7.

2.4.1. Search strategy and framework

The framework for the review is shown in Figure 2.7. The research approach used for conducting the review is a hybrid of quantitative and qualitative methodologies. The scientometric analysis and systematic reviews represent the quantitative and qualitative methods, respectively. The hybrid methods are adopted to overcome the limitation associated with the "mono review method," such as bias in selecting research papers and a lack of holistic view of a research domain (Khodabandelu & Park, 2021; Pluye & Hong, 2014). As a result, the study involves three main stages, as shown in Figure 2.7. The first stage entails data acquisition, in which the processes adopted for retrieving the relevant literature for this study are illustrated in Figure 2.8. Firstly, the topic validation was conducted by carrying out a preliminary search through the "Scopus" and "Web of Science" databases to ascertain the necessity of this review. Subsequently, "Scopus" was selected for the literature retrieval since it is the largest academic database with recent publication coverage (Debrah et al., 2022).

Table 2. 6: Common failure modes of water pipes

Failure modes	Schematic	Descriptions	Common causes	Typical pipe materials	Typical pipe diameter	References
Circumferential cracking	Circumferential cracking	Failure of the pipe at its circumference	• Inadequate bedding condition • Inappropriate backfilling • Frost penetration	• Cast iron • Ductile iron • Steel • Asbestos cement • Copper	It can occur in all pipe sizes but is common in pipes with <200 mm diameter	(Grigg, 2017; K. T. H. Liyanage & Dhar, 2017; Rajeev et al., 2014; Trickey et al., 2016)
Longitudinal cracking	Longitudinal cracking	Failure of the pipe at its longer side	• Traffic loads • Internal water pressure • Frost penetration	• Cast iron • Polyvinyl chloride • Polyethylene • Asbestos cement	It can occur in all pipe sizes but is common in pipes with >300 mm diameter	(Barton et al., 2019; Farewell et al., 2012; Ji et al., 2020; Mora-Rodríguez et al., 2014)
Bell splitting	Bell splitting	Splitting of the bell at the joint of the pipe	• Temperature • Internal water pressure	• Cast iron • Ductile Iron	It can occur in all pipe sizes but is common in pipes with >300 mm diameter	(Makar et al., 2001; Makar, 2000)
Corrosion pitting	Corrosion pitting	Loss of metal at a localised region on the pipe	• Soil acidity • Bacteria • Soil resistivity • Water quality	• Cast iron • Ductile iron • Steel	It can occur in all pipe sizes but is common in pipes with <200 mm diameter	(Grigg, 2017; Pękala & Pietrucha-Urbanik, 2018; Rajeev et al., 2014)
Blown-out hole	Blow-out hole	Creation of large hole(s) on the pipe	• Corrosion • Internal water pressure	• Cast iron • Ductile iron • Steel • Polyvinyl chloride • Polyethylene	All pipe sizes	(Ji et al., 2020; Makar et al., 2001; Makar, 2000; Tang et al., 2019a)

After defining the exclusion criteria, as shown in Figure 2.8, the search string returns 385 documents. These documents were forwarded to the following step, where a preliminary evaluation of abstracts and snowballing were performed. Snowballing is the process of identifying additional papers to include in the review. The two types of snowballing were adopted: forward and backward snowballing. Forward snowballing involves searching for papers that cite the paper being examined, while backward snowballing involves checking the reference list of the examined paper to find other relevant research articles. This process ensures no related paper is omitted. The outcome of these processes produced 76 research papers (see Figure 2.8).

The study's second stage involves a scientometric review to map and visualize the selected research papers, identifying publication trends, keyword analysis, and contributions from research outlets and institutions. The final stage is a systematic review to recognize the contributions of scientific literature and identify knowledge gaps in the field. The systematic review categorizes techniques for modeling the failure probability of water pipes into three groups: physical, statistical-based, and AI-based models.
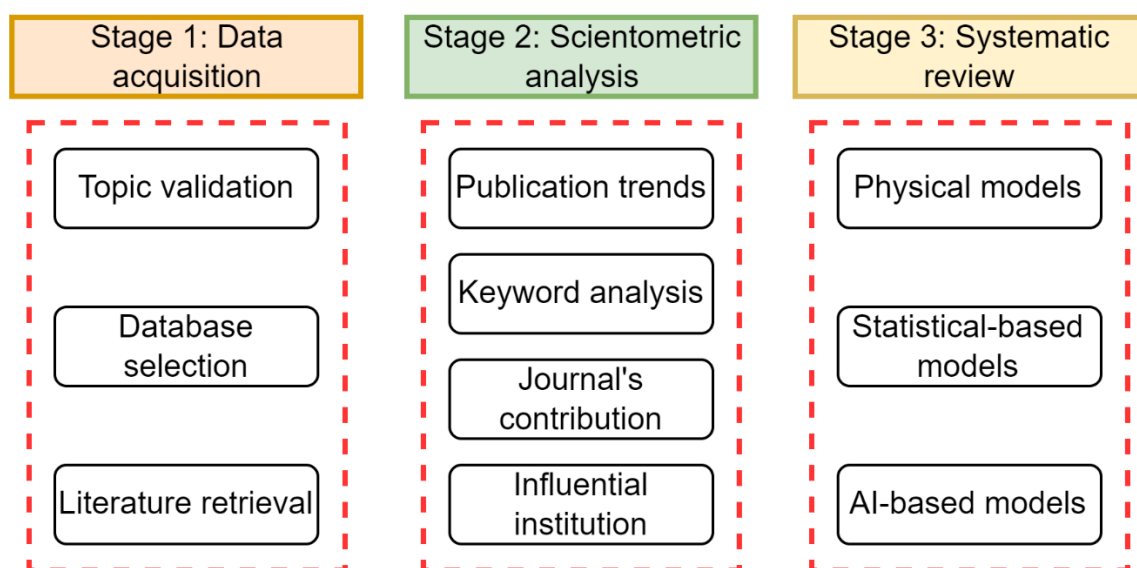
Figure 2. 7: Review framework for techniques used in predicting failure probability of water pipe
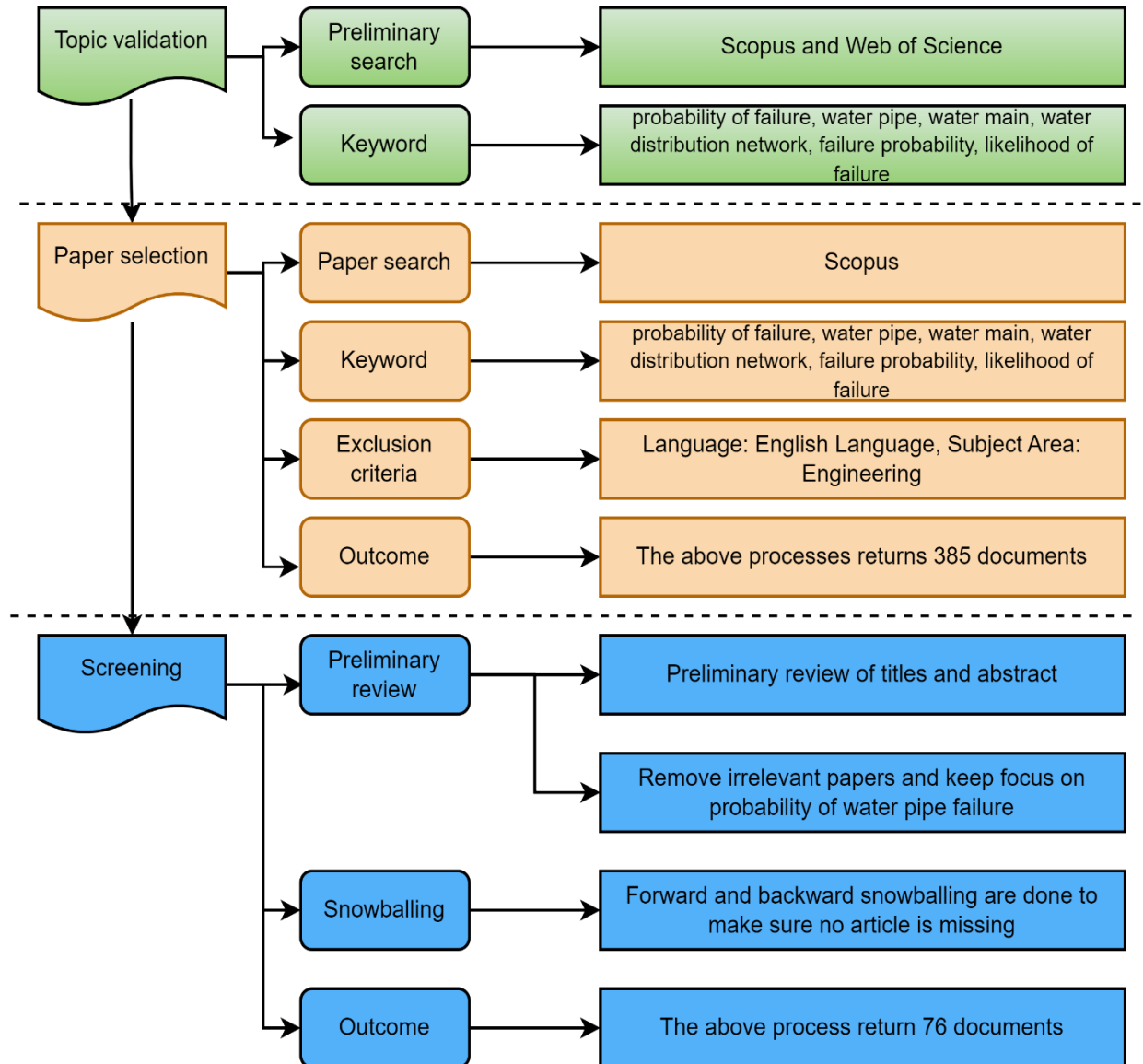
Figure 2. 8: Complete procedures for literature retrieval for investigating failure probability of water pipes

2.4.2. Scientometric review of techniques used for failure probability of water pipes

As previously stated, this review employs scientometric analysis to identify the trend in the annual publication, keyword analysis, and the contribution of research outlets and influential institutions in the domain of water pipe failure probability prediction. Although various tools are available for conducting a bibliometric analysis, VOSviewer was chosen for this research

because it is straightforward and produces strong bibliometric networks (Debrah et al., 2022). Moreover, VOSviewer is open-source software.

2.4.2.1 Publication trends

Figure 2.9 shows the annual publishing trends for scholarly literature in the domain of failure probability of water pipes. As per the articles included in this study, the publication year ranged from 1987 to 2022. It can be seen that only two articles were published before the year 2000. This either shows that minimal efforts were paid to this domain in the last century, or the problem was not significant as the pipes were not aged. This could be attributed to the limited robust technologies and tools for developing predictive models in this era (Aryai et al., 2022). Another reason may be the lack of sophisticated digital tools or platforms to showcase the conducted research in this era (i.e., before 2000). From 2000 to 2010, the overall number of publications was 17, whereas from 2011 to 2020, the total number of publications was 42. This implies that the field is emerging as the number of publications increases from one decade to another. Although only 15 articles have been published in 2021 and 2022, it is expected that the number will exceed that of the previous decade by the end of 2030. This is due to the fact that many utilities are taking proactive measures to construct prediction models for their WDNs (Weeraddana et al., 2020). Furthermore, the emergence of AI-based models may promote further growth in this domain's publishing.

2.4.2.2. Keyword co-occurrence analysis

Keyword co-occurrence analysis is an important aspect of bibliometric/scientometric analysis, as it gives insight into the links between research areas within a specific domain. Using VOSviewer, the "minimum number of occurrences" requirement was set to 5; 38 keywords satisfied this condition (Ness Van Eck & Waltman, 2010).
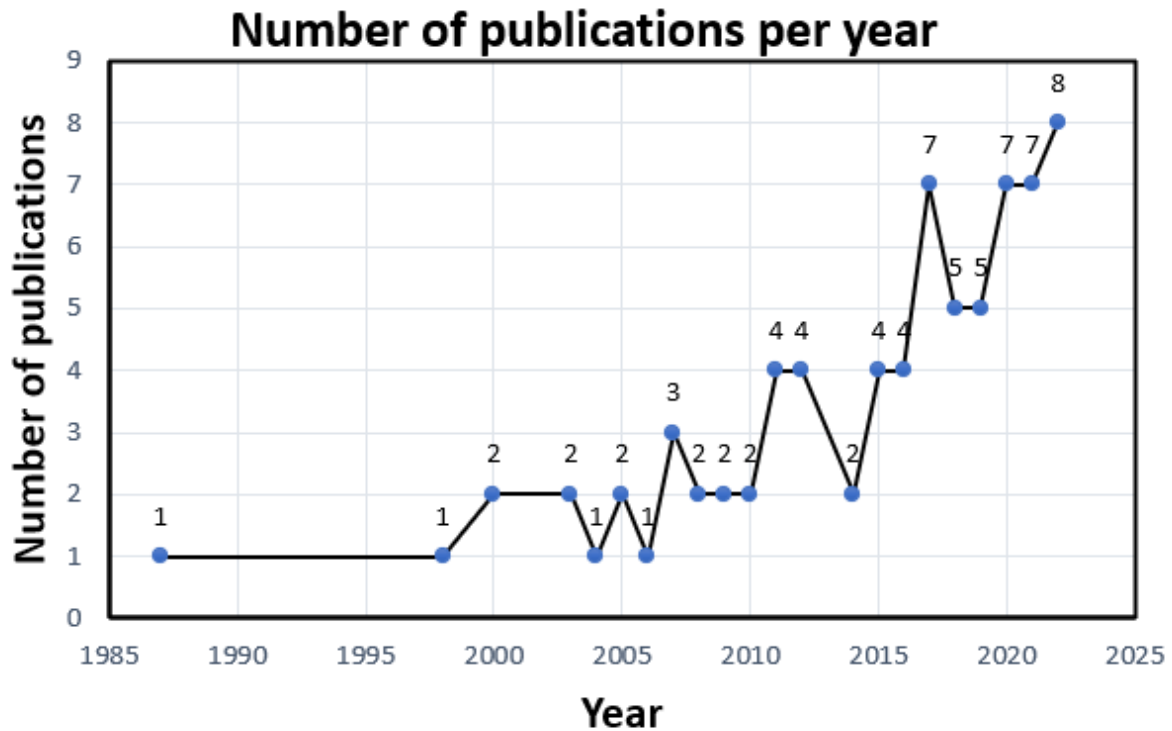
Figure 2. 9: Trends in research publications

However, it was observed that some of these keywords express the same meaning; hence, they were merged using the "Thesaurus file function" of the software. For instance, the keywords "risks" and "risk" were merged. According to Figure 5, three clusters are observed from the mapping network. The red cluster is dominated by "probability of failure," whereas the blue and green clusters are dominated by "water distribution network" and "risk assessment," respectively. The proximity of each node to one another indicates the strength of their relationship. For example, because their nodes are adjacent, the strength between "probability of failure" and "reliability analysis" is significant (see Figure 2.10). Table 2.7 reports the top 30 keyword occurrences and their respective total link strength. This evidences that the selected articles to be reviewed in this study are highly representative of the domain, where "probability of failure" and "water pipe" are the most occurred keywords in the list. According to the frequency of occurrences and total link strength, "physical models" are well-researched in the literature compared to "artificial intelligence" or "machine learning" models. Furthermore, it is also observed that "cast iron pipe" is the most investigated water pipe type. In this context, a

link refers to the co-occurrence of two keywords in a publication. The strength of the link is represented by a positive numerical value, with a higher value indicating a stronger association between the two keywords. As presented by the developers of VOSViewer, the formula for calculating the link strength can be found in Equation 6 in (Nees van Eck & Waltman, 2009). The total link strength denotes the number of publications in which the two keywords appear together.
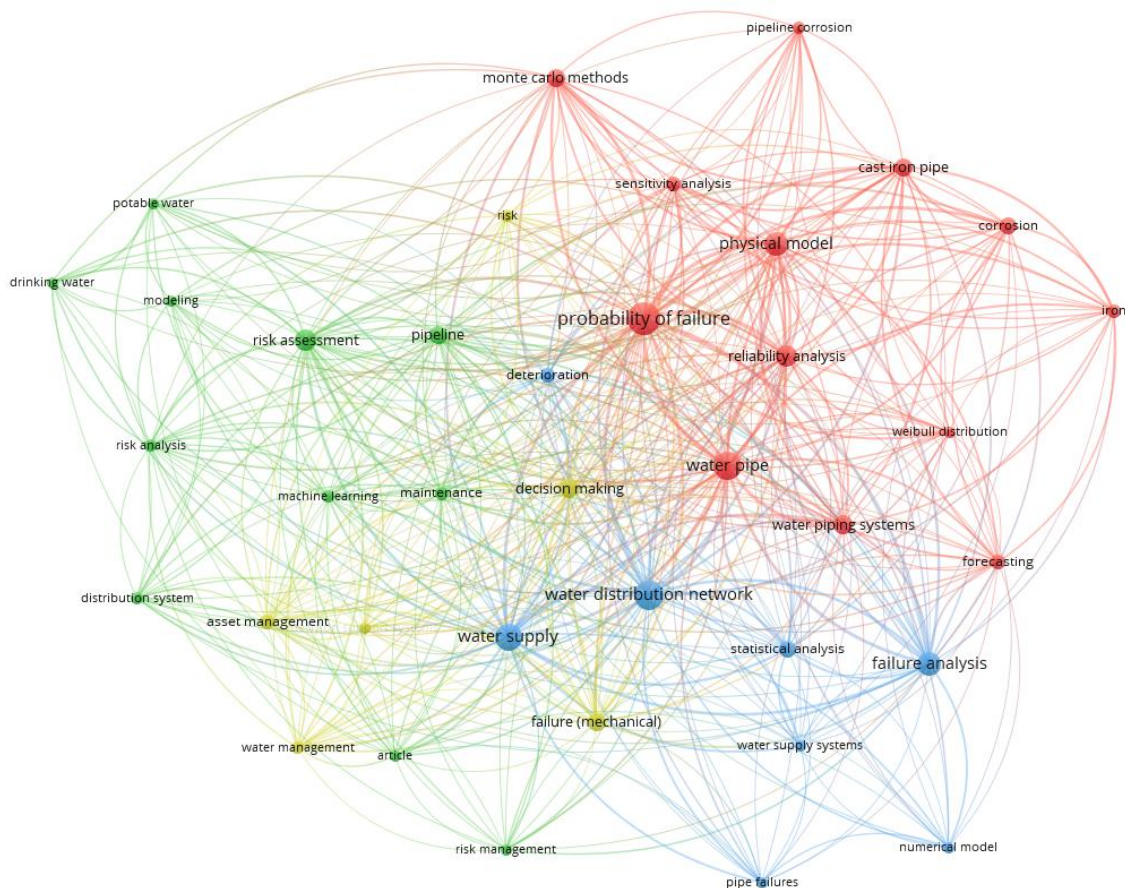
Figure 2. 10**: Keyword co-occurrence network

2.4.2.3. Journals' contributions

Table 2.8 lists the top 10 research outlets as per the scope of this review. This sort of analysis is beneficial for readers who want to know where to source information related to a particular research focus. Furthermore, it can provide insights into how institutional and commercial libraries allocate journal subscription funds based on their research interests.

Table 2. 7: Keyword co-occurrence and total link strength

Keyword	Occurrences	Total link strength
Probability of failure	37	212
Water distribution network	29	165
Water pipe	28	162
Water supply	26	157
Physical model	19	120
Failure analysis	20	113
Reliability analysis	16	105
Risk assessment	16	104
Decision making	13	90
Water piping systems	13	87
Pipeline	12	86
Cast iron pipe	12	82
Failure (mechanical)	12	81
Monte carlo methods	11	78
Corrosion	10	62
Statistical analysis	10	54
Asset management	8	53
Maintenance	7	51
Forecasting	8	50
Iron	7	50
Risk analysis	6	50
Water management	6	49
Sensitivity analysis	8	48
Deterioration	8	43
Weibull distribution	5	43
Machine learning	6	42
Decision support systems	5	41
Pipeline corrosion	6	40
Risk	6	37
Distribution system	6	35

Using VOSviewer, the analysis type was set as "citation," and the unit of analysis was "sources." Even though there is no limit for the "minimum number of documents" and "minimum number of citations," these requirements were set to 2 and 25, respectively, for generating the optimal network after multiple attempts. Besides, this threshold limit was also

in agreement with past studies (Tariq et al., 2021). Table 2.8 shows that "Reliability engineering and system safety," and "Water research" are the most productive journal in terms of the number of citations, documents, and total link strength. The table also reveals that, despite the fact that "Journal of water supply: research and technology - aqua" and "Journal of infrastructure systems" have five and four articles published, the citations of "Journal of hydroinformatics" and "Journal of water resources planning and management" with 3 research papers are higher.

Table 2. 8: Journals contributions

Source	Documents	Citations	Total link strength
Reliability engineering and system safety	8	319	11
Water research	5	303	12
Journal of hydroinformatics	3	204	8
Journal of water resources planning and management	3	150	12
Journal of water supply: research and technology - aqua	5	124	3
Water resources research	2	99	7
Engineering failure analysis	3	47	1
Journal of infrastructure systems	4	36	6
Urban water journal	3	34	8
Journal of pipeline systems engineering and practice	2	31	2

2.4.2.4. **Influential institutions**

The co-authorship of organizations is another necessary sort of bibliometric analysis performed in this review in order to know the most collaborative institutions in the domain of water pipe failure probability predictions. This will also be helpful for individuals or organizations interested in researching the failure probability of water pipes to know who they can collaborate with effectively. The analysis type was set to "co-authorship," and the unit of analysis was set to "organizations." The "minimum number of documents" and the "minimum number of citations" were set to 1 and 25, respectively. 61 out of 149 institutions qualified for these

criteria, and the top 10 are presented in Table 2.9. According to the results, " Université Laval, Canada," "McGill University, Canada," "Eawag:Swiss federal institute of aquatic science and technology, Switzerland," and "University of British Columbia, Canada" are the most collaborative institutions, with each exhibiting a total link strength of 4. Furthermore, the results show that few institutions in Canada, Switzerland, Australia, and Spain have established some collaboration with other institutions. To achieve a high standard in combating the increasing failure rate of water pipes through the development of robust predictive models, institutions across the globe should collaborate with each other so they can benefit from diverse knowledge and experience, as this is currently lacking in the scholarly literature.

Table 2. 9**:** Influential institutions

Organization	Documents	Citations	Total link strength
Université Laval, Québec city, Canada	1	132	4
McGill university, Montreal, Canada	1	35	4
Eawag: Swiss federal institute of aquatic science and technology, Switzerland	3	166	4
University of British, Columbia, Kelowna, Canada	1	32	4
Monash university, Melbourne, Australia	1	55	3
Rmit University, Australia	1	33	3
Swinburne university of technology, Australia	1	33	3
Rajani Consultants inc, Ottawa, Canada	1	33	3
Southwest Petroleum University, China	1	29	3
Universidad politécnica de valencia, valencia, Spain	1	32	2

2.4.3. **Systematic review of techniques used for failure probability of water pipes**

The systematic review findings, which center on the adopted techniques for predicting the failure probability of water pipes, are presented. Figure 2.11 illustrates the proposed classification of the techniques, which involve physical, statistical, and AI-based models.
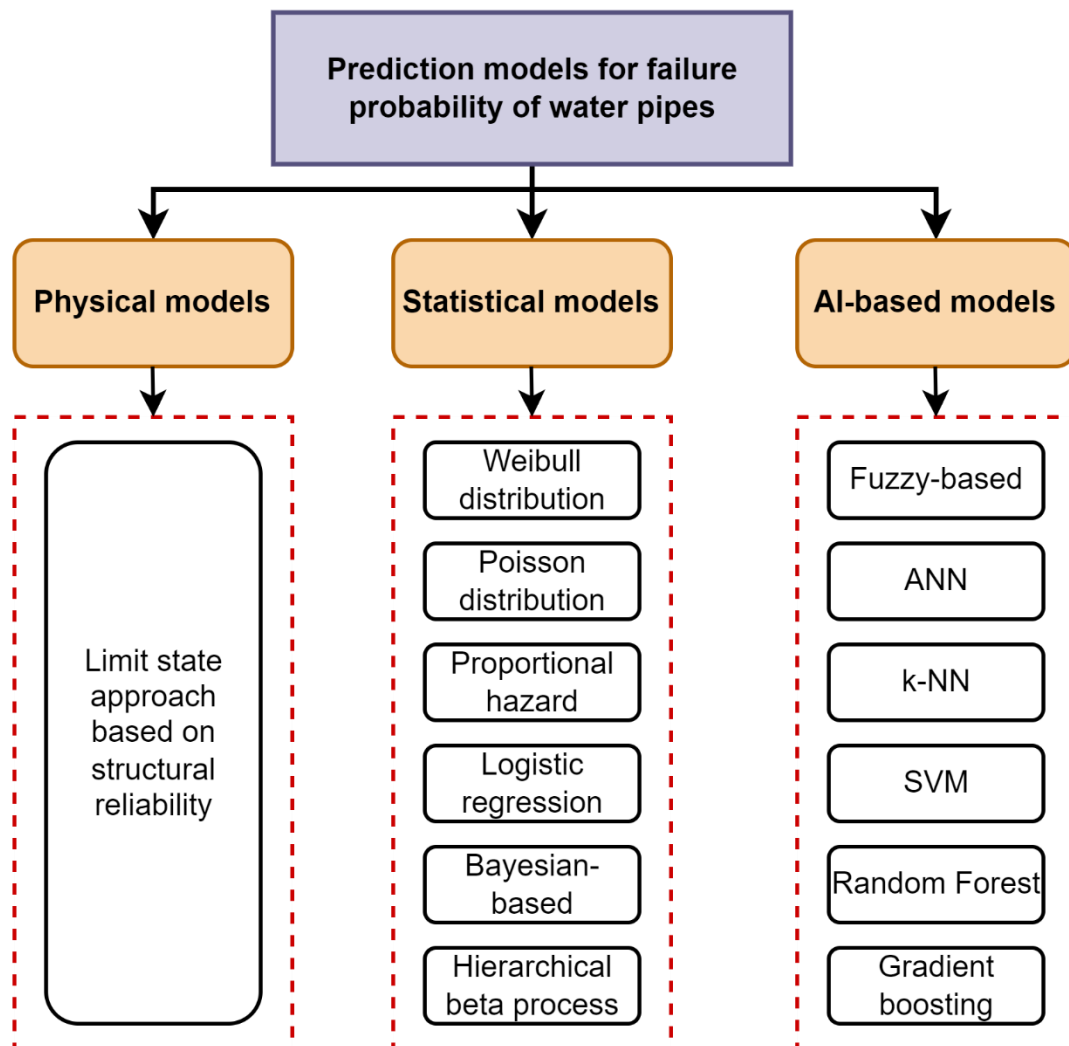
Figure 2. 11: Prediction models for failure probability of water pipes

2.4.3.1. **Physical models**

Researchers in this domain have invested enormous efforts in modeling the failure probability by considering the pipes' physical failure mechanism. Table 2.10 gives a summary of the reviewed physical models, including the authors, the employed methodology, model accuracy

(if available), validation status, failure mode, the most important factors considered, the used data type, and the pipe material. It should be noted that the "most important factors considered" refer to the critical factors affecting the failure probability in each study (where applicable), which are determined from a sensitivity analysis, correlation analysis, or relative importance analysis.

(I) Limit state equations based on structural reliability

During the development of a physical model based on reliability theory, a limit state function (LSF) is defined by comparing the capacity of a pipe with the stress exerted on it. Equation 2.1 states a typical LSF, where LSF violation will result in a pipe's failure. Thus, the probability of water pipe failure can be represented by Equation 2.2 (Mahmoodian & Li, 2018). It should be noted that the LSF can be defined based on different failure modes such as burst, leakage, longitudinal deflection, etc. (Davis et al., 2008; W. Wang et al., 2021). Furthermore, the LSF can be transformed into a factor of safety (FoS) to define the failure criteria of pipes (Sadiq et al., 2004). An FoS of more than 1 indicates that the pipe is safe and has no failure, while an FOS of less than 1 shows that the pipe's capacity has been exceeded and, thus, is considered failed. In terms of quality-based assessment, a partial FoS is derived by comparing the load exerted on the pipe to its resistance.

$$G\ (S, L, T) = L(T) - S(T) \tag{2.1}$$

$$P_f = p\ [\ L(T) - S(T)\ \leq 0\ ] \tag{2.2}$$

where LSF is represented by $G\ (S,\ L, T)$, L denotes the capacity of a pipe, S represents the exerted stresses on the pipe, T refers to the time, p denotes the probability of LSF violation, and P_f denotes the failure probability. Therefore, if G(S,L,T)>0, the pipe is considered safe, while failed if G(S,L,T)≤0. In most studies, the failure probability (i.e., Equation 2.2) is estimated using a structural reliability analysis technique such as simulation approaches (i.e.,

Monte Carlo Simulation, MCS) or analytical approaches (i.e., First Order Reliability Method, FORM). However, MCS is the most adopted approach in scholarly literature due to its ability to deal with complex problems such as the failure of water pipes (Padmanabhan et al., 2006). Further details on MCS can be found in the study of Rubinstein & Kroese (Rubinstein & Kroese, 2008). In this method, the variables formulating the LSF are randomly selected and employed to evaluate its outcome. If the LSF is violated, the pipe will fail; otherwise, no failure event will be recorded. This process requires a large simulation number (e.g., over 10,000 simulations) (Punurai & Davis, 2017; Wilson et al., 2015) to estimate the failure probability by using Equation 2.3.

$$P_f = \frac{n}{N} \qquad (2.3)$$

where n is the number of times the LSF is violated, and N is the total number of simulations.

As seen in Equation 1, the LSF comprises two components: resistance of the pipe to failure and exerted stresses on the pipe. The two components of LSF for previous studies are shown in Table 2.11.

Mahmoodian & Aryai (2017) investigated the failure probability of corroded steel water pipes. The non-linearity of corrosion was taken into consideration based on the Power Law model (Romanoff, 1957)–using Equation 2.4.

$$\Delta = at^b \qquad (2.4)$$

where Δ denotes the corrosion depth at a time "t," "a," and "b" could be determined from the analysis of the inspection data.

Table 2. 10**:** Summary of the physical models

Authors	Methodology	Validated?	Failure mode	Accuracy	The most important factors considered	Type of data	Material type
(Mazumder, Salman, Li, et al., 2021)	Limit state equations + MCS	No	Collapse	-	-	Field data	CI
(W. Wang et al., 2021)	Limit state equations + MCS	No	Leakage, Burst, Deflection, and Bending failure	-	-	Field data	CI
(W. Li et al., 2021)	Finite Element Analysis and MCS	Yes	Burst	-	Pipe thickness, Internal pressure, Traffic	Literature-based + Field data	CI
(Mady, 2021)	Limit state approach + MCS	No	Collapse	-		Field data	Concrete
(Aryai et al., 2020)	Finite element modeling + copula method	Yes	Corrosion pit and collapse	-	Diameter, wall thickness	Field data	CI
(Zhang et al., 2019)	Limit state equations + MCS	No	Corrosion pit, burst, and rupture	-	Radial corrosion rate, axial corrosion rate	Literature-based	Steel
(Mahmoodian & Li, 2018)	Limit state equations + MCS	No	Corrosion pit	-	Wall thickness, corrosion depth	Literature-based + Field data	CI
(Phan et al., 2018)	Limit state equations + MCS + Weibull distribution	No	Corrosion pit and burst	-	Wall thickness, Loads, Corrosion size	Literature-based data	-
(Aryai & Mahmoodian, 2017)	Limit state equations + MCS	Yes	Leakage, circumferential cracking, ring deflection, wall rupture, and buckling	15.6% (prediction error)	Length	Field data + Historical data	CI
(Mahmoodian & Aryai, 2017)	Limit state equations + MCS	No	Bending, wall thrust, ring deflection, longitudinal	-	Corrosion factors	Field data	Steel

			deflection, leakage, and buckling				
(Punurai & Davis, 2017)	Limit state equations + MCS	No	Burst and collapse	-	-	Field data + Literature-based data	AC
(Daniel Wilson et al., 2015)	Factor of safety analysis + MCS	No	Corrosion pitting and collapse	-	Diameter, Buried depth	Field data + Literature-based data	CI
(Qian et al., 2013)	Limit state equations + FITNET FFS Procedure + MCS	Yes	Collapse	-	Tensile strength, Crack depth, Internal pressure	Literature-based	-
(Jallouf et al., 2011)	Factor of safety analysis + MCS	No	Burst	-	-	Field data	CI, Steel, PVC
(Qian et al., 2011)	FITNET FFS + Limit state equations +MCS	No	Burst	-	Depth of defect, Tensile strength, Wall thickness	Field data + Literature-based data	-
(Davis et al., 2008)	Limit state equations + MCS	No	Burst and collapse	-	-	Field data	AC
(De-Silva et al., 2006)	Limit state equations + FOSM	No	Corrosion pit	-	-	Field data	Steel
(De Leon & Macías, 2005)	Limit state equations + FOSM	No	Corrosion pit and burst	-	-	Field data	-
(Davis et al., 2004)	Limit state equations + Survival function + MCS	No	Corrosion pit	-	-	Field data	CI
(Sadiq et al., 2004)	Factor of safety analysis + MCS	No	Corrosion pit and collapse	-	Corrosion depth	Literature-based + Field data	CI

Table 2. 11: Components of limit state functions and driving factors

Failure modes	Resistance	Stress	Driving Factors	Reference
Longitudinal split	$\sigma_{f=} \sigma_o - 120\frac{(Age \times \delta)}{b_o}$	$W = (P_e + P_s)(D + b_o)$	Corrosion	(Davis et al., 2004)
Pitting	P_f $= \frac{2\sigma_{uts} \times t \times 0.5^{\frac{65}{\sigma_{ys}}}}{(D-t)}$ $\times [\frac{1-\frac{d(T)}{t}}{1-\frac{d(T)}{t} \times Q^{-1}}]$	P_{op}	Corrosion	(Qian et al., 2011)
Flexural failure	$M_n = \frac{2D_f E \Delta Y y_o S_{f1}}{D_m^2}$	F_y	Ground movement and external loadings	(Gabriel, 2011)
Ring deflection failure	$\Delta X = \frac{K(D_L W_c + P_s)D_m}{\frac{8EI}{D_m^3} + 0.061E'}$	$\Delta X_{cr} = 0.05\, D_i$	Soil compression	(BS 9295, 2010)
Buckling failure	$P = \frac{1}{S_f}\sqrt{(32R_w B' E_s \frac{EI}{D_m^3})}$	$P_{cr} = R_w \frac{w_c}{D_m} + \frac{p_s}{D_m}$	Elevated temperature	(Moser & Folkman, 2008)
Circumferential failure	σ_y	$p_f = \frac{P_{op} \times D}{2t_w} + \sigma_b$	Internal pressure and bending stress	(Phan et al., 2018)
Leakage failure	$S_{total} = \sum \pi R_k^2$	S_{lim}	Corrosion	(Li et al., 2017)
Wall thrust failure	$T_a = F_y(W_t - \Delta)\emptyset$	$T_{cr} = 1.3\,(1.67P_s C_L + P_W)\frac{D_o}{2}$	Traffic, hydrostatic and soil loads	(Gabriel, 2011)
Collapse pressure	p_{total} $= \frac{\gamma_w}{1728}H_w$ $+ R_w \frac{w_s}{12D_o} + \frac{w_L}{144D_o}$	p_{cr} $= \frac{1.2C_n\,(EI)^{0.33}(\emptyset_s E' k_v)^{0.67} R_H}{S_f r_o}$	External and internal loads	(W. Wang et al., 2021)
Bending failure	$\sigma_{m=} \frac{(w_s + w_L)L^2}{8\pi[d - a(t)]r^2}$	σ_f	External loadings	(W. Wang et al., 2021)

In order to achieve a high level of prediction accuracy, 10,000 simulations were considered, as the sample size has a significant effect on the accuracy of MCS results. More importantly, six failure modes (i.e., limit states) were considered: buckling, ring deflection, longitudinal

deflection, flexural, wall thrust and burst. The flexural and wall thrust limit states were considered as a series system since the occurrence of any of the two limit states can cause failure. The other four states were considered as a parallel system since the occurrence of any of them could not cause the system to fail. Figure 2.12 shows the six modes of failure using the appropriate shapes in fault tree analysis to represent the top event (i.e., failure of water pipes), failure modes (i.e., basic events), and the logic gates. Further reviewed articles relating to the physical based models are presented in Appendix A.
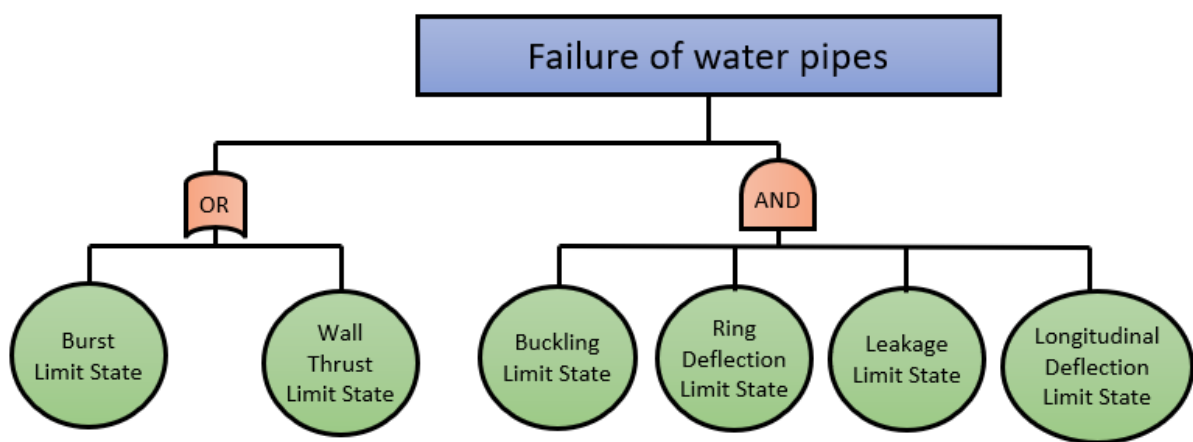
Figure 2. 12: Multiple failure modes

(II) Advantages and limitations of physical models

Physical models are based on interpretative equations; therefore, they facilitate a proper understanding of the water pipe failure mechanism. Furthermore, physical models are easy to develop, especially when the model parametrization is simple yet accurate. However, the data required to develop a physical model is costly to get. The data are mainly obtained through site inspection, which is technically difficult to conduct because of the pipes' locations. Furthermore, the collected data may not fully represent the absolute condition of the pipes, as only a segment of the pipe can be practically investigated. For instance, the corrosion depth can vary spatially along the pipe length, which may not be captured by the data collected.

Another limitation of physical models is their unsuitability for global applications. This is due to the difference in environmental conditions of various geographical locations.

2.4.3.2.   **Statistical-based models**

The second category is statistical-based models. As the failure mechanism of the pipelines is of complex behavior, the development of statistical-based models does not require a proper understanding of such behavior. Unlike physical models, statistical-based models rely on historical data. This type of modeling is suitable for pipes with substantial historical failure data. The summary of the statistically based studies is presented in Table 2.12. Furthermore, in the case that more than one technique is used in a study, the one with the highest accuracy is reported in the Table. In order to choose an appropriate statistical method for fitting historical data to a model, the failure mode of the pipes in the network can be a deciding factor. For instance, the distribution of historical data associated with pipe failure due to corrosion pitting may be different from those associated with pipe deflection. Therefore, the statistical-based models are grouped into parametric and non-parametric models. The parametric models include Weibull distribution, Poisson distribution, Proportional hazard models, and Logistic regression. On the other hand, the non-parametric models include, Bayesian-based models, Hierarchical Beta Process (HBP) models, and other-statistical based models such as fuzzy analytical network process (FANP), preference ranking organization method for enrichment evaluation (PROMETHEE), among others. The details of these statistical-based models, as per the articles reviewed in this study, are presented in Appendix A.

(I) Advantages and limitations of statistical-based models

Since statistical-based models can be built on historical failure data of water pipes from installation until the out-of-service time, these models could be handy in modeling the entire life cycle of water pipes. Another benefit of this model category is that it is cost-effective. However, the prediction accuracy of statistical models depends on the historical data quantity

and quality. For instance, Bayesian models need to be updated continuously with the required data in order to sustain their prediction accuracy. Additionally, a limited number of variables, mostly due to a lack of data, are included in the development of statistical-based models, which can result in inappropriate predictions. Furthermore, statistical-based models require formulating some assumptions for model development, which may need some level of expertise in addition to their computational complexity.

2.4.3.3. **Artificial intelligence-based models**

In the past few decades, AI-based models have emerged as the way forward in modeling water pipe failure due to their robust predictive capacity. AI models are developed by simulating human intelligence on computer systems (Samoili et al., 2020). In this section, models relating to fuzzy and machine learning algorithms are reviewed and presented in Table 2.13. For instance, Fan et al. (2022) compared five machine learning algorithms (lightGBM, ANN, k-NN, SVM, and LR) to classify water pipes as broken or intact. The models used 13 factors, including 11 continuous and 2 categorical variables, to predict pipe failure probability on a 0-1 scale. Results showed that while most factors correlated with pipe failure, no single factor was dominant, supporting the complex nature of pipe failure mechanisms. The lightGBM algorithm performed best in terms of prediction accuracy and computational efficiency, followed by the ANN model. This study highlights the potential of machine learning in predicting water pipe failures and the importance of considering multiple factors in such predictions. Further reviewed articles relating to AI-based models are presented in appendix.

(I) Advantages and limitations of ML-based models

As an important benefit, many ML-based models (i.e., non-parametric) do not require pre-assumptions on the distribution and form of pipe failure data. Furthermore, ML-based models are able to handle heavy data sets effectively within a limited time.

Table 2. 12**:** Summary of statistically based studies

Authors	Methodology	Validated?	Evaluation metric	The most important factors considered	Type of data	Material type
(Al-Ali et al., 2020)	Logistic regression	No	0.74 (Acc)	Diameter, Material type	Historical data	Concrete, DI, PVC, FRP, and others
(Konstantinou & Stoianov, 2020)	Logistic regression and Linear Discriminant Analysis (LDA)	Yes	0.81 (AUC)	Pressure and Diameter	Historical data	AC, DI, and CI
(Weeraddana et al., 2020)	Gaussian Process Regression + Bayesian model	Yes	0.73 (AUC)	-	Historical data	AC and PVC
(Phan et al., 2019)	Weibull distribution model	No	-	-	Historical data	CI and DI
(Tchórzewska-Cieślak et al., 2019)	Bayesian model	No	-	-	Historical data	-
(Garcia et al., 2019)	Clustering-based spatiotemporal analysis	No	-	-	Historical data	CI
(Ismaeel & Zayed, 2018)	FANP and PROMETHEE + Probability theory	Yes	94.4 (VF)	Wall thickness, Leaks	Historical data + Expert's opinion	-
(Ward et al., 2017)	Weibull distribution model	Yes	96.0% (R^2)	-	Historical data	PE and others
(Chik et al., 2017)	Bayesian simple model	Yes	0.75 (AUC)	-	Historical data	CI
(Luo et al., 2017)	Hierarchical beta process	Yes	0.798 (AUC)	-	Historical data	-
(Elsawah et al., 2016)	Homogenous Poisson model	No	-	-	Historical data	CI, Steel, PVC, DI
(Shin et al., 2016)	Bayesian Inference + Markov chain Monte Carlo method	No	-	Diameter, Length	Historical data	Ductile cast iron
(Vladeanu & Koo, 2015)	Weibull distribution model	Yes	2.71% (PE)	-	Historical data	AC, CI, DI, Concrete, and PVC
(Lin et al., 2015)	Hierarchical beta process	Yes	0.827 (AUC)	-	Historical data	-
(Z. Li et al., 2013)	Hierarchical beta process	Yes	0.61 (AUC)	-	Historical data	-
(Singh & Adachi, 2012)	Homogenous Poisson model	No	-	-	Historical data	CI, DI, PVC, Concrete
(Friedl et al., 2012)	Logistic regression	No	-	Pressure, Diameter	Historical data	CI, DI, AC, Concrete, PE, PVC, Steel

Reference	Model		Acc/VF		Factors	Data	Material
(Karamouz et al., 2012)	Minimum Redundancy-Maximum Relevance + AHP	No	-		Diameter, Soil pH, and Length	Literature-based data	-
(Tchorzewska-Cieslak, 2012)	Weighting method	No	-		-	Historical data + Expert's opinion	CI, PVC, PE, Steel
(Kleiner & Rajani, 2012)	Bayesian model, Ordered list model and Logistic regression	Yes	-		Age, Length, Number of past failures	Historical data	CI, DI, AC
(Scheidegger et al., 2013)	Weibull-exponential + Bayesian Inference	Yes	-		Diameter and Material	Historical data	DI
(Scholten et al., 2014)	Weibull-exponential + Bayesian Inference	Yes	-		Diameter and Material	Historical data	DI
(Singh, 2011)	Bayesian model	No	-		-	Historical data	CI, DI, CC, AC, GI, PVC, and others
(Debón et al., 2010)	Proportional Hazard Model	Yes	0.76 (AUC)		-	Historical data	AC, CI, DI, and PE
(Carrión et al., 2010)	Proportional Hazard Model	Yes	-		-	Historical data	CI, PE, and AC
(Rogers & Grigg, 2007)	Non-homogeneous Poisson distribution model	No	-		-	Historical data	AC, CI, concrete, PVC, Steel, and other
(Economou et al., 2007)	Non-homogeneous Poisson distribution model + Bayesian model	Yes	17.3% (PE)		-	Historical data	AC
(Vanrenterghem-Raven, 2007)	Proportional Hazard Model	Yes	15% (PE)		Age, Length, Diameter, Previous break	Historical data	Steel and Non-steel
(Mailhot et al., 2000)	Proportional Hazard Model	No	-		Age	Historical data	-
(Cooper et al., 2000)	Logistic regression	No	-		Diameter, Soil corrosivity, Traffic load	Historical data	CI
(Lei & Sægrov, 1998)	Weibull distribution model	No	-		-	Historical data	CI, DI, Plastic, and others
(Andreou et al., 1987)	Proportional Hazard Model + Poisson model	No	-		-	Historical data	CI, Steel, Concrete

Acc : percentage of the pipes whose true conditions were predicted correctly

AUC: the area under the curve, VF: validation factor, PE: percentage error

Table 2. 13: Summary of AI-based studies

Authors	Methodology	Validated?	Evaluation metric	The most important factors considered	Type of data	Material type
(Fan et.al., 2022)	LightGBM, LR, SVM, ANN, k-NN	Yes	(0.81) AUC	Interval to last break, cold days, hot days, pipe length, pipe age	Historical data	CI, DI, and others
(T. Y.-J. Chen et al., 2022)	RF, Boosting trees, XGBoost	Yes	0.899 (AUC)	-	Historical data	CI, DI, PVC, and others
(Rifaai et al., 2022)	LR	Yes	0.680 (AUC)	Years from past failure, length, number of past failure	Historical data	AC, CI, DI, PVC, and others
(Raspati et al., 2022)	RF	Yes	-	Age, Length, Internal pressure, and Pipe material	Historical data	AC, CI, DI, GRP, PE, and PVC
(Weeraddana et al., 2021)	Random survival forest	Yes	0.719 (AUC)	-	Historical data	AC, CI, DI, PVC, PE
(Jara-arriagada & Stoianov, 2021)	LR	Yes	0.814 (AUC)	Pressure	Historical data	AC, CI, PE
(Giraldo-González & Rodríguez, 2020)	GBT, SVM, ANN and Bayes	Yes	0.998 (AUC)	Previous failure, Length, Precipitation	Historical data	AC and PVC
(Rahbaralam et al., 2020)	LR and XGBoost	Yes	0.859 (AUC)	Age, Material, Length	Historical data	DI, PE, Steel
(Kumar et al., 2018)	Gradient Boosting Decision Trees	Yes	0.62 (Precision)	Previous failure, Age, Diameter	Historical data	CI, DI, and others
(Konstantinou & Stoianov, 2020)	Gradient boosting, ANN, RF	Yes	1.0 (AUC)	Age, Length, Internal pressure	Historical data	AC, DI, and CI
(Al-Zahrani et al., 2016)	Fuzzy-based	No	-	-	Historical + Literature data	AC, PVC, Steel
(Salehi et al., 2021)	Fuzzy-based	No	-	-	Historical data + Literature data	CI, DI

Moreover, the introduction of automated machine learning tools such as TPOT, Orange, and RapidMiner, amongst others, has made ML applications easier and more accessible to individuals who are not experts in programming languages (Baharun et al., 2022; Demšar & Zupan, 2013; Randal et al., 2016). However, ML-based models are as good as the quality and quantity of the data used in developing them. Hence, low-quality and limited data will result in inaccurate predictions. Besides, ML models developed with limited data and an algorithm that optimizes its parameters may lead to overfitting (T. Y.-J. Chen et al., 2022). Hence, such a model cannot be generally applied to other historical failure data. Since ML-based models require a substantial amount of data, this implies that the models are most suitable for pipes that exhibit higher failure frequency. Additionally, some of the ML algorithms have zero to low interpretability, which might make it difficult to understand the relationship between the explanatory variables and the output of such models.

2.5. EXISTING FAILURE INDICATORS

Table 2.14 provides a summary of previous studies that have employed ML techniques to predict various failure indicators of water pipes. These studies span across different geographical locations and consider diverse pipe materials such as cast iron (CI), ductile iron (DI), asbestos cement (AC), polyvinyl chloride (PVC), polyethylene (PE), and others. The predicted outcomes range from condition index, wall thickness loss, remaining useful life, time to failure, to failure probability or rate. The performance metrics reported include accuracy, precision, recall, F1-score, area under the curve (AUC), mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), and R-squared (R^2) values. Various data splitting techniques, such as train-test split, k-fold cross-validation, and yearly forecasting, have been employed to evaluate the models' performance. In cases where more than one model is developed in a study, the performance metrics of the best model are

presented in Table 2.14. This table highlights the extensive efforts made by researchers to leverage ML for the proactive management of WDNs.

Based on the review, existing water pipe failure prediction models have focused on predicting the likelihood of failure events occurring, without distinguishing between failure types like leaks versus bursts (Rifaai et al., 2022; Robles-velasco et al., 2020). However, leaks and bursts have distinct failure mechanisms and driving factors. Leaks often result from corrosion-induced holes and joints over time, while bursts are sudden ruptures mostly from excessive internal pressure (Pękala & Pietrucha-Urbanik, 2018). Moreover, despite the significant advancements in modeling water pipe failures, there remains a scarcity of models that delve into the specific causes behind these failures. Current models predominantly focus on predicting the probability, timing, and rate of failure. A deeper understanding of the root causes via predictive modeling is essential for crafting precise intervention strategies and enhancing decisions regarding maintenance and replacement of water infrastructure.

Table 2. 14: Summary of previous studies predicting various outcomes as failure indicators of water pipes

Technique	Predicted outcome	Study location	Pipe material	Performance metrics	Data splitting	Reference
HBO-DL	Condition index	Hong Kong	CI, AC, GI, GIL, PE, PVC	MAE – 0.144 RMSE – 0.193 MAPE – 2.387%	Training – 80% Testing – 20%	(Mohammed Abdelkader et al., 2024)
XGBoost, Random Forest, Logistic regression	Failure probability	United Kingdom	AC, CI, PE, PVC, DI	Recall – 0.526 Precision – 0.389 F1-score – 0.226	5 folds (Cross validation)	(Beig Zali et al., 2024)
CatBoost, DT, RF, XGBoost, LightGBM	Wall thickness loss	Canada, USA	CI, DI, Steel	MAE – 3.013 MAPE – 24.044 R^2 – 0.904	Training – 80% Testing – 20%	(Models et al., 2024)
Logistic regression	Failure probability	Uganda	Steel, UPVC	Accuracy – 0.969 AUC – 0.996	Training – 60% Testing – 40%	(Auma et al., 2023)
RestNet, CNN	Failure probability	China	CI, Steel, PE, PCCP	Recall – 0.8571 Precision – 0.0207 AUC – 0.8703	Training – 90% Testing – 10%	(Liu et al., 2023)
Probabilistic LSTM, ARIMAX	Failure rate	USA	-	MSE – 3.09 UR – 3.19	Training – 1985-2012 data Testing – 2012-2019 data	(Fan et al., 2023)
ANN, LightGBM,	Failure probability	USA	CI, DI, and others	AUC – 0.81 Recall – 0.861	Training – 80% Testing – 20%	(Fan, Wang, Zhang, Xiong, et al., 2022)

LR, KNN, and SVC						
XGBoost, RF, BT	Failure probability	USA	CI, DI, PVC, and others	AUC = 0.8992	Training – 12 years data Testing –3 years data	(T. Y.-J. Chen et al., 2022)
MARS, GEP, and M5 Tree	Failure rate	Iran	AC, CI, PE	R = 0.981 RMSE = 0.544 Testing – 20%	Training – 80%	(Amiri-Ardakani & Najafzadeh, 2021)
WPHSM, RF, and RSF	Remaining useful life	Canada	AC, CI, and DI	C-Index = 0.925	Training – 80% Testing – 20%	(Snider & McBean, 2021)
ANN	Time to failure	Switzerland	CI, DI, and PE	R = 0.882	Training – 80% Testing – 20%	(Kerwin et al., 2020)
LR and SVR	Failure probability	Spain	CE, PL, and ME	AUC – 0.873 Recall – 0.848 Acc- 0.769	Training – 5 years data Testing – 2 years data	(Robles-velasco et al., 2020)
ANFIS and ANN	Remaining useful life	USA and Canada	AC, CI, DI, and Steel	MAE = 0.880 MAPE = 5.431 RAE= 0.007	Training – 75% Testing – 25%	(Tavakoli et al., 2020)
ANN, RF, and XGBoost	Time to failure	North America	AC, CI, DI, and PVC	R – 0.85 RMSE – 5.81	Training – 80% Testing – 20%	(Snider & McBean, 2018)

2.6.   RESEARCH GAPS

Despite the contribution of previous studies in relation to the failure of water pipes in WDNs, there are still some areas that are yet to be fully explored. Following the comprehensive review conducted in this study, four research gaps are identified and discussed as follows.

1. One of the notable research gaps identified is the limited attention given to the critical factors that influence water pipe failure and their interrelationships. While previous studies have made efforts to rank a few select factors, there remains a dearth of comprehensive investigations into the crucial factors that play a significant role in water pipe failure. Identifying and understanding a comprehensive list of failure factors is essential for developing effective strategies to mitigate failures and improve the overall management of WDNs. By bridging this gap, valuable insights can be gained into the intricate dynamics of water pipe failure and pave the way for more informed decision-making processes (Research objective 1). Furthermore, the impact of these failure factors on water pipe failure modes remains relatively unexplored within the existing literature. While some studies have examined the failure modes themselves, there is a lack of comprehensive research that delves into the intricate relationship between the failure factors and the resulting failure modes. Understanding how these factors contribute to specific failure modes is crucial for devising targeted preventive and maintenance strategies. By addressing this research gap, an improved understanding of the complex interactions between failure factors and failure modes can be gained, leading to more effective and efficient management of WDNs (Research objective 2).

2. In relation to previous studies that have investigated the probability of water pipe failure, ML-based models have been understudied compared to physical and statistical-based models. Therefore, it is important to explore the capabilities of ML to develop effective models for predicting the failure probability of water pipes. Furthermore, the

few previous ML-based models have some limitations: 1) there is a lack of systematic selection of hyperparameters for the ML models; 2) there is a lack of systematic selection of the best features (i.e., input variables) to include in the model development; 3) limited studies have interpreted the contribution of each feature to the predictive model; 4) there is a lack of a model deployment via a web-application, which can enhance the decision-making process in WDN management (Research objective 3a to answer the question – How to model the failure probability of water pipes?)

3. While some studies exist on the development of models to predict the probability of water pipe failure, the extant literature lacks studies investigating the specificity of the failure, such as developing models to predict the probability of leaks or bursts in a WDN (Research objective 3b to answer the question – How to model the probability of leak, burst, and no leak/burst?). Understanding the probability of leaks or bursts in a WDN is crucial for effective infrastructure management and maintenance. By accurately predicting these specific failure events, water utilities can proactively implement targeted measures to prevent and mitigate potential damages, reduce water loss, and ensure an uninterrupted water supply to consumers. However, the lack of models that specifically address the probability of leaks or bursts hinders understanding the underlying mechanisms and patterns associated with these specific failure events.

4. Understanding the causes of water pipe failure is crucial for the effective management and maintenance of WDNs. While previous studies have made valuable contributions by identifying various factors that contribute to pipe failure, there is a need to go beyond mere identification and delve into the development of predictive models. Predictive models offer a powerful tool for assessing and managing the risks associated with water pipe failure. By analyzing historical data and incorporating relevant factors, these models can provide insights into the specific causes of failure and enable proactive

measures to be taken to mitigate risks. However, despite the importance of such models, their development remains a research gap in the current literature (Research objective 4).

2.7. SUMMARY

This chapter presents a comprehensive review of previous studies relating to the failure of water pipes. Firstly, an extensive review of factors influencing water pipe failure is presented. A hybrid method of scientometric and systematic analysis is adopted for the literature review. The factors are categorized into pipe-related, operation-related, external-related, and soil-related factors. Subsequently, the common failure modes in water pipes are presented. Furthermore, the techniques used for predicting the failure probability of water pipes are reviewed, which are divided into physical, statistical, and AI-based models. The existing failure indicators predicted in previous studies are also reviewed. Consequently, four research gaps are identified based on the literature review, which forms the basis for the remaining objectives of the study.

Chapter 3

RESEARCH METHODOLOGY AND MODEL DEVELOPMENT[2]

3.1. INTRODUCTION

This chapter presents the methodologies adopted for this study. Firstly, it provides an overall framework that can be used to achieve the objectives of the research. Subsequently, the detailed methods and techniques employed to fulfill the objectives are explained alongside the novelty introduced by the study.

3.2. OVERALL FRAMEWORK

The overall research framework is depicted in Figure 3.1. The research study starts with a literature review to identify research gaps. Afterward, the research objectives are formulated based on the identified research gaps. As shown in Figure 3.1, this study entails four research objectives. Objective 1 deals with a scientometric and systematic review of failure factors and

---

[2] This chapter is largely based upon:

Taiwo, R., Zayed, T. & Ben Seghier, M. E. A. (2024). " Integrated intelligent models for predicting water pipe failure probability". Alexandria Engineering Journal, 86, 243-257, https://doi.org/10.1016/j.aej.2023.11.047

Taiwo, R., Yussif, A., Ben Seghier, M. E. A., & Zayed, T. (2024). "Explainable Ensemble Models for Predicting Wall Thickness Loss of Water Pipes". Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2024.102630

Taiwo, R., Zayed, T. & Adey, B.T. "Explainable deep learning models for predicting water pipe failure." Journal of Environmental Management (IF = 8.7, Q1). (Under review – 1st cycle)

Taiwo, R., Zayed, T. & Adey, B.T. "Interpretable ensemble models for predicting causes of water pipe failure." Reliability Engineering and System Safety (IF = 8.1, Q1). (Under review – 1st cycle)

Taiwo, R., Zayed, T. Elshaboury, N., & Abdelkader, E. M. "Promoting Sustainable Water Distribution Networks: Modeling of Water Pipe Failure Factors and Modes." Cleaner Engineering and Technology (IF = 5.3, Q1). (Under review – 2nd cycle)

failure modes of water pipes. The relevant literature are retrieved using Scopus and Web of Science databases. VOSviewer software is used for the scientometric review, while Nvivo software aids in the systematic review process. The detailed methodology adopted for the literature review has been explained in sections 2.2.1 and 2.4.1. While both scientometric and systematic reviews contribute to understanding water pipe failure, they serve distinct purposes in this research. The scientometric analysis provides a quantitative assessment of the research landscape through bibliometric indicators, publication patterns, and citation networks, helping identify research trends, influential authors, and emerging topics in the field. This approach offers a macro-level view of how knowledge of water pipe failure has evolved over time. In contrast, the systematic review follows a structured methodology to critically analyze and synthesize the actual content of the literature, focusing on identifying specific failure factors, modes, and their relationships. This micro-level analysis involves detailed examinations of methodologies, findings, and conclusions from individual studies to extract meaningful patterns and insights. The combination of these complementary approaches provides both breadth (scientometric) and depth (systematic) in understanding the current state of knowledge in water pipe failure research.

In the second objective, the identified factors influencing water pipe failure and common failure modes in WDNs are used to design a questionnaire survey. The questionnaire is sent to experts working in the domain of WDN, both in the industry and academic institutions. SPSS and SmartPLS software are used to analyze the questionnaire data. Partial least square structural equation modeling (PLS-SEM) is employed to model, rank, and investigate the relationship between the failure factors and failure modes. Hypotheses are proposed and validated using statistical tests.

The third objective entails two tasks. The first task is to develop an optimized model for predicting water pipe failure probability. The logistic regression (LR) algorithm is used as a

base model to develop the predictive model. Subsequently, the genetic algorithm (GA) is utilized to optimize the hyperparameters and select the best features for the LR model. In addition to the interpretability provided by the LR's coefficient, the SHapley Additive exPlanations (SHAP) framework is employed to explain the contribution of each feature to the model prediction. Furthermore, the best-optimized model is deployed in the form of a web application, which can assist water utility in the management of WDN. The second task deals with the development of models to predict the probability of a leak and burst of individual pipes in a network. Deep learning techniques such as deep neural network (DNN), convolution neural network (CNN) and TabNet are integrated with optimization techniques such as Bayesian Optimization (BO) to develop optimized models. Copeland algorithm is employed to select the best model. Consequently, the best model is deployed in the form of a web application.

The fourth objective deals with the development of an optimized model to predict the causes of water pipe failure. Ensemble state-of-the-art algorithms such as XGBoost are employed to develop predictive models for forecasting the causes of water pipe failure. The hyperparameters of the models is carefully selected using an optimization algorithm.

As the detailed methodology for objective two has been expounded in the literature review section, the adopted methodologies for the remaining three objectives are explained in detail in the subsequent sections.

3.3. PLS-SEM MODEL DEVELOPMENT

The adopted framework for this section of the study is depicted in Figure 3.2. The framework is divided into four phases. The first phase of the research dealt with the identification of the factors influencing water pipe failure, which were categorized into pipe-related, operation-related, external-related, and soil-related factors.
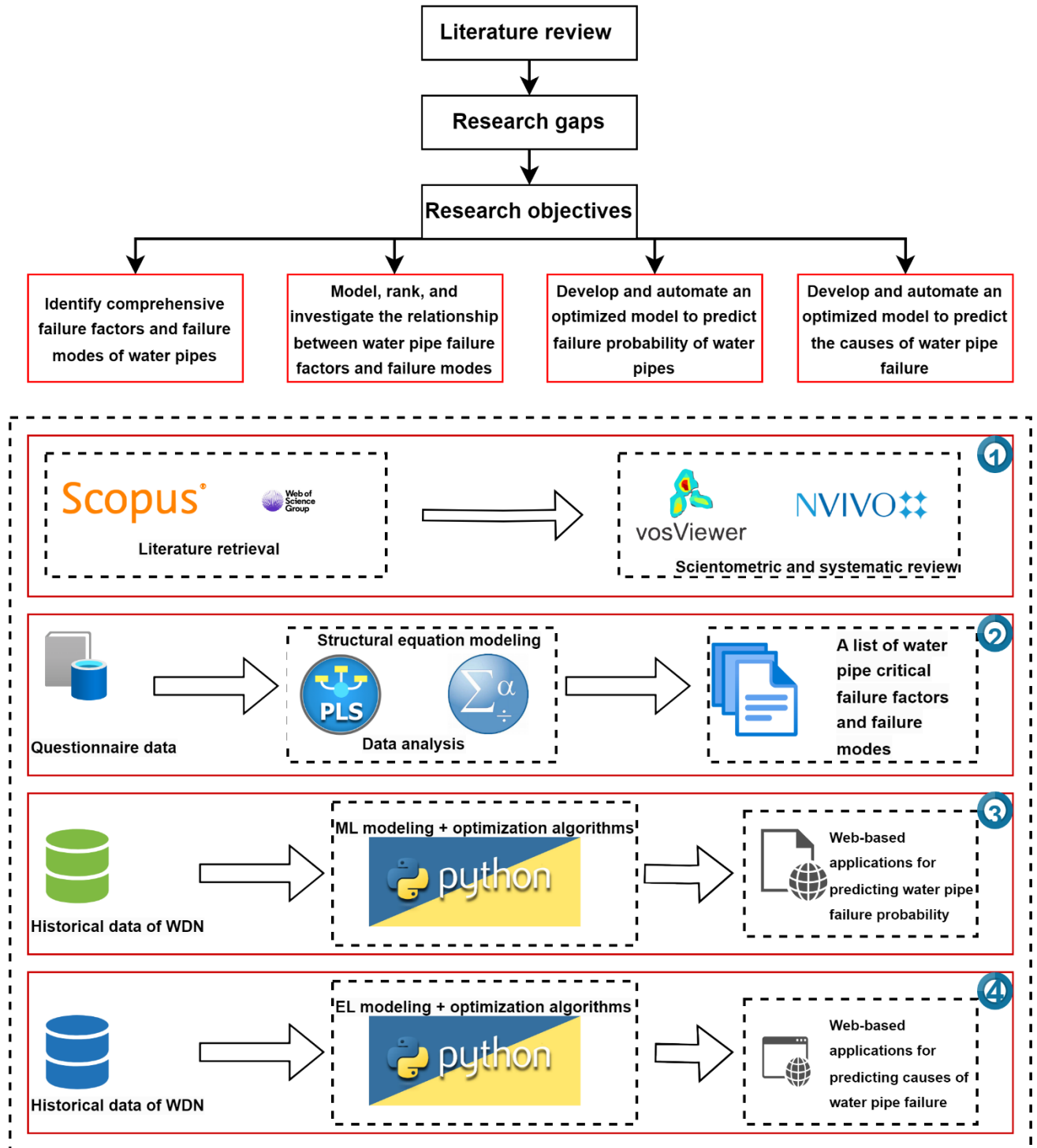
Figure 3. 1: Overall research methodology framework

Further, the failure modes were also identified herein. Following the literature survey (objective 1 of the study), the second phase entails the development of a conceptual model. The conceptual model is created to comprehend the impact of various factors on water pipe failure. This conceptualization is based on the key findings from the literature review. 30 failure factors

are identified and organized into four related constructs - pipe-related, operation-related, external-related and soil-related. The failure modes construct represents the different ways that pipes can fail. Connections are drawn between the factor constructs and failure modes to hypothesize relationships. This conceptualization builds on previous studies that have examined subsets of these factors using methods like AHP. However, this study's model offers a more comprehensive perspective by considering 30 factors across four constructs and their association with failure modes. The proposed conceptual model is illustrated in Figure 3.3. Thus, the hypothesis tested in this study states that "water pipe failure factors" possess a significant effect on "failure modes" because they could influence "failure modes." Statistical significance is determined through appropriate statistical tests, with the expectation that if the hypothesis holds true, these factors will demonstrate a measurable influence on specified failure modes. If the hypothesis is disproven, it will indicate that these "water pipe failure factors" do not have a relationship with the "failure modes." Subsequently, the questionnaire is designed and distributed to potential respondents. Data analysis utilizing partial least square structural equation modeling (PLS-SEM) represented the third stage. PLS-SEM is chosen as the analytical approach for this study due to its suitability for handling complex and interrelated constructs, which is inherent in the context of water pipe failure. Based on the analysis, the critical water pipe failure factors and failure modes were selected at stage four, and the hypothesis was validated. The details of the research methods are delineated in subsequent sections.

SEM algorithm was adopted as an analytical technique for analyzing survey responses. The SEM technique was preferred over other methods as it gives direct and indirect relationships between variables and constructs (Adabre et al., 2021). The validation process in SEM involves a two-step approach. Firstly, Confirmatory Factor Analysis (CFA) is utilized to confirm the measurement model, which employs measured indicators to provide evidence of significant

constructs. Secondly, a structural model is created to evaluate research hypotheses using path analysis (Ali et al., 2023). It should be noted that SEM analysis can be carried out using two distinct approaches. These are covariance-based SEM (CB-SEM) and partial least square SEM (PLS-SEM). PLS-SEM was employed due to its advantages over CB-SEM.
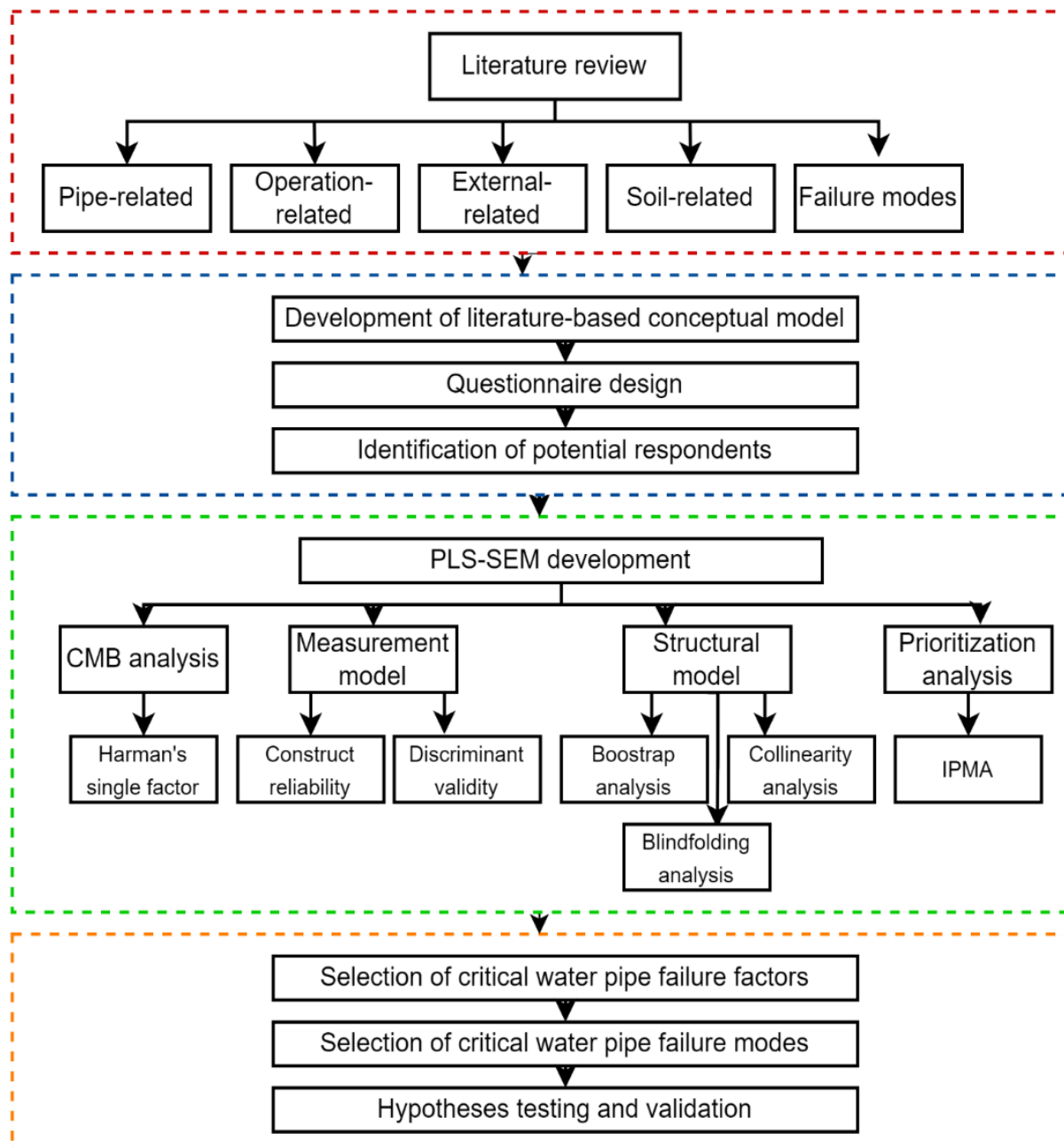
Figure 3. 2**:** Research framework for PLS-SEM model development
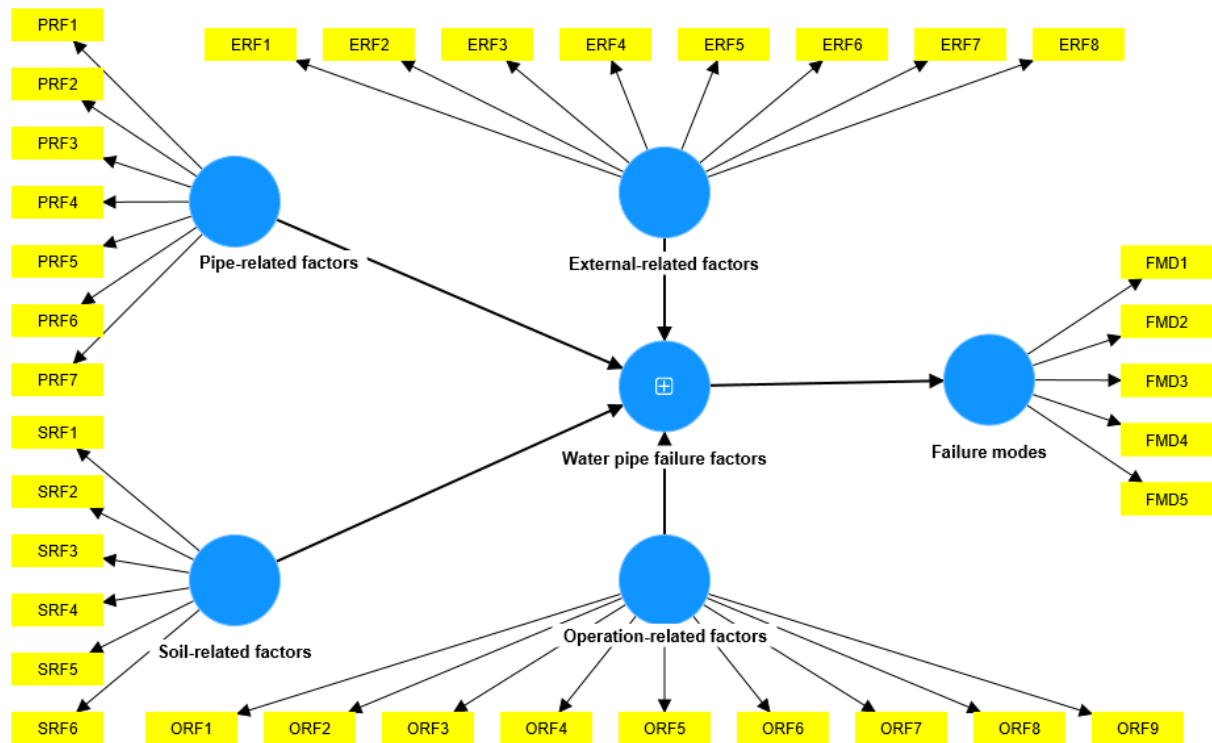
PRF1- Age, PRF2- Buried depth, PRF3- Diameter, PRF4- Length, PRF5- Material, PRF6- Protection efficiency, PRF7- Wall thickness, ERF1- Temperature, ERF2- Chemical substance, ERF3- Frost, ERF4- Land use, ERF5- Lime leaching, ERF6- Microbiologically induced corrosion, ERF7- Precipitation, ERF8- Traffic, SRF1- Bedding condition, SRF2- Soil moisture, SRF3- Soil pH, SRF4- Soil resistivity, SRF5- Soil type, SRF6- Soil aeration, ORF1- Installation and pump operation, ORF2- Internal pressure, ORF3- Maintenance practices, ORF4- Number of leaks, ORF5- Water acidity, ORF6- Water alkalinity, ORF7- Water hammer, ORF8- Water temperature, ORF9- Water velocity, FMD1- Bell splitting, FMD2- Blown-out hole, FMD3- Circumferential cracking, FMD4- Corrosion pitting, FMD5- Longitudinal cracking.

Figure 3. 3**:** A conceptual model for understanding failure factors of water pipe

These include the fact that 1) it has been found to account for experimental variance in a better manner; 2) it has been demonstrated to provide more accurate predictions; and 3) it offers a more robust statistical framework for evaluating multiple components (Hair et al., 2014, 2019). Prior to the SEM analysis, common method bias (CMB) was conducted to ensure the validity of the collected data. CMB can arise when respondents consistently exhibit similar response patterns, which may result in overestimating or underestimating the true relationships among the variables under study. Therefore, it is crucial to identify and address any potential common method bias to ensure the accuracy and dependability of the research findings. In this study, Harman's single-factor test was employed to examine the presence of CMB and its potential

impact on the results. SmartPLS version 4.0.9.1 and SPSS version 27 were used for all the computations conducted in this study. As highlighted earlier, the SEM model entails two main components: measurement and structural models, which are discussed in the subsequent sections.

3.3.1. **Measurement model**

A model that shows the association between the observable variables and the latent construct (such as pipe-related factors) is referred to as a measurement model. The observable variables could be defined as either formative or reflective. A confirmatory tetrad analysis (CTA) was conducted to ascertain this, and the analysis shows that the variables are reflective, as 80% of the p-values associated with the variables are statistically significant at 0.05 (Gudergan et al., 2008). In PLS-SEM, the measurement model is tested and validated utilizing construct reliability, convergent validity, and discriminant validity. The construct reliability and convergent validity can be evaluated using four different methods, including a) outer loading, b) reliability of each indicator using Cronbach's alpha, c) composite reliability, and d) average variance extracted (AVE) (Hair et al., 2017; Hair et al., 2014).

a) **Outer loading:** The outer loading method evaluates the strength of the relationship between the observed indicators and the latent construct they are intended to represent. This method assesses the extent to which the indicators measure the same construct and could provide a relative ranking for the indicators in the same construct. Table 3.1 reports different thresholds for various tests and their interpretations (Adabre et al., 2021; Ali et al., 2023). According to the table, an outer loading value of 0.5 is considered acceptable, indicating a reasonable level of association between the indicators and the construct. Conversely, an outer loading of 0.7 or higher is deemed highly satisfactory, indicating a strong and robust relationship between the indicators and the latent construct.

b) **Cronbach's alpha (α)**: This method evaluates the reliability of the indicators by measuring the internal consistency of the measurement items. Cronbach's alpha is a measure of how well the items within a construct are related to each other. According to Table 3.1, Cronbach's alpha of greater than 0.7 is deemed acceptable. The coefficient (α) can be calculated using Equation 3.1

$$\alpha = \frac{N - \bar{c}}{1 + (N-1) - \bar{c}}$$

3.1

where the number of indicators is denoted by N and $\bar{c}$ represents the inter-correlation among the indicators. A low average value of $\bar{c}$ typifies that the observable variables do not fit well in a construct, while a high average value shows that the variables measure the same construct satisfactorily (Hair et al., 2013).

c) **Composite reliability (ρ_c)**: Composite reliability is a reliable measure that evaluates the consistency and reliability of indicators in measuring a specific construct. It takes into account both the variance and covariance of the indicators, providing a more comprehensive assessment. A composite reliability value of 0.7 or above is commonly regarded as acceptable (Mohandes et al., 2022). This value indicates that at least 70% of the variance in the observed indicators can be attributed to the underlying construct they are intended to measure. Specifically, according to Table 3.1, composite reliability values of 0.8 and 0.9 are considered to show satisfactory and perfect reliability, respectively. The coefficient (ρ_c) can be calculated using Equation 3.2.

$$\rho_c = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum var(\varepsilon_i)}$$

3.2

where the term (λ_i) represents the component loading of each indicator to a construct, and the variation is represented by the term $var(\varepsilon_i)$, which is defined in Equation 3.3.

$$var(\varepsilon_i) = 1 - \lambda_i$$

3.3

d) **Average variance extracted (AVE):** AVE is another crucial metric of convergent validity that helps determine the proportion of variance in a construct that is explained by its indicators relative to the variance caused by measurement error. An AVE value of 0.5 or above is generally considered acceptable, indicating that its indicators are associated with at least 50% of the variance in the construct, and the measurement model is deemed to be reliable (Hair et al., 2017). The AVE can be computed using Equation 3.4.

$$AVE = \frac{\sum \lambda_i^2}{\sum \lambda_i^2 + \sum var(\varepsilon_i)} \qquad 3.4$$

where $\sum \lambda_i^2$ represents the sum of the squared factor loadings for the construct, and $\sum var(\varepsilon_i)$ represents the sum of the unique variances for the construct's indicators.

A measurement model's discriminant validity ensures that the measured constructs are distinct from each other and do not measure the same underlying construct. Discriminant validity is crucial in ensuring that the measurement model accurately represents the constructs being studied and does not cause any potential biases or confounding effects. In this study, the discriminant validity was evaluated using three methods: Heterotrait-monotrait (HTMT) ratio, Fornell-larcker criterion, and cross-loading, which are briefly highlighted below.

a. **Heterotrait-monotrait (HTMT) ratio:** To assess the discriminant validity, this approach compares the correlations between items from different constructs (heterotrait correlations) with the correlations between items from the same construct (monotrait correlations) (Henseler et al., 2015). An HTMT ratio value of less than 0.9 is considered an acceptable discriminant validity threshold (Kline, 2016). The formula for estimating the HTMT ratio is presented in Equation 3.5.

$$HTMT = \sqrt{\frac{HT^2}{MT1 \times MT2}} \qquad 3.5$$

where HT represents the average correlation between the items measuring different constructs, MT1 and MT2 refer to the average correlations between items measuring the same construct.

b. **Fornell-larcker criterion:** In this method, the validity is established by comparing the square root of AVE for each construct with the correlations between the construct and other constructs in the model (Fornell & Larcker, 1981). To ensure validity, it is expected that the square root of AVE for each construct surpasses the correlation value between the construct and any other construct in the model (Alshurideh et al., 2020).

c. **Cross loading**: This method scrutinizes the extent to which a specific indicator of a construct loads on its corresponding construct compared to other constructs (Hair et al., 2017). An indicator has satisfactory discriminant validity if it loads higher on its corresponding construct than on other constructs.

3.3.2. Structural model

The structural model examines the association between the latent constructs in the proposed theoretical framework. In essence, the four hypotheses developed in the conceptual framework will be tested and validated using the structural model. The strength of the relationship between the constructs is represented by the path coefficients. The path coefficients are validated using four methods. This includes a) a significance test using a t-test, b) a collinearity measure, c) explanatory power using R^2, and d) predictive relevance using Q^2. Bootstrap and blindfolding analyses were conducted for the t-test and Q^2 test, respectively.

a) **Significance test using t-test**: The t-test is utilized to determine if the path coefficients in the model are statistically significant. The significance level is usually set at 0.05 or 0.01, indicating the probability of obtaining the observed results by chance. A significant t-test indicates that the relationship between the two constructs is not due to chance and supports the hypotheses. A t-value greater than 1.96 indicates statistical significance at the 0.05 level, while a t-value greater than 2.58 indicates significance at the 0.01 level for a two-tailed test. (Hair et al., 2014).

b) **Collinearity measure:** Collinearity is a common problem in SEM, where high correlations exist between two or more independent variables (Adabre et al., 2021). This can lead to inaccurate parameter estimates and affect the model's overall validity. To assess collinearity, the Variance Inflation Factor (VIF) is used, and values greater than 3.5 indicate collinearity (Mohandes et al., 2022).

c) **Explanatory power using R^2**: R^2 serves as a measure to gauge the extent to which the exogenous variables in a model account for the variance in the endogenous variable. Ranging from 0 to 1, R^2 values offer insights into the model's explanatory prowess, with higher values indicative of a stronger explanatory power. Within the framework of SEM, R^2 plays a crucial role in evaluating the goodness of fit and assessing the strength of relationships among constructs. While the acceptable threshold for R^2 may vary depending on the specific context of the model, conventional interpretations consider R^2 values of 0.02, 0.13, and 0.26 as representative of weak, moderate, and substantial relationships, respectively (Cohen, 1992).

d) **Predictive relevance using Q^2:** Q^2 serves as a measure of the model's predictive capability and assesses its accuracy in predicting endogenous variables. It provides valuable insights into the model's predictive relevance. When the Q^2 value exceeds zero, it signifies that the model possesses predictive power. In line with the findings of Hair et al. (2014), Q^2 values of 0.02, 0.15, and 0.35 are considered indicative of small, medium, and large predictive strengths, respectively. To estimate Q^2, cross-validation is employed, where a subset of the data is used for model estimation, while the remaining data is reserved for model validation purposes. This methodology ensures robustness in evaluating the model's predictive performance.

3.4. DEVELOPMENT OF PROBABILITY OF FAILURE MODEL

This section explains in detail the methodologies adopted to develop probability of failure (POF) models. Figure 3.4 shows the research framework adopted in this section of the study in three main stages.

Table 3. 1: Threshold for various tests and their interpretations

Tests	Threshold	Interpretation
Outer loading	> 0.5	Acceptable
	> 0.7	Highly satisfactory
Cronbach's alpha (α)	> 0.7	Acceptable
	> 0.7	Acceptable
Composite reliability (ρ_c)	> 0.8	Satisfactory
	> 0.9	Perfect
Average variance extracted (AVE)	> 0.5	Acceptable
Heterotrait-monotrait (HTMT) ratio	< 0.9	Acceptable
Variance Inflation Factor (VIF)	< or = to 3.5	Acceptable
Path coefficient using t-value	> 1.96 at 0.05 p-value	Significant
	> 2.58 at 0.01 p-value	Significant
	> or = to 0.02	Weak
R^2 value	> or = to 0.13	Moderate
	> or = to 0.26	Substantial
	> or = to 0.02	Small
Q^2 value	> or = to 0.15	Medium
	> or = to 0.35	Large

The first stage deals with data collection and pre-processing. The data was collected from Hong Kong's Water Supply Department (HK WSD) and supplemented with climatic and traffic data from the Hong Kong Observatory and Transportation Department, respectively. Subsequently, the data was pre-processed by normalizing and standardizing it and inputting the missing records. The second stage involves the ML predictive models, which were developed using logistic regression and genetic algorithms. Different performance metrics derived from the confusion matrix, such as the Area Under the Curve (AUC) values, were used to evaluate the models. Subsequently, coefficients of the best LR model and SHAP values were used to measure the independent variables' contribution to the model's outcome. Detailed explanations of these steps are provided in subsequent sections. All the computations for this task were performed in Python 3.7 environment.

3.4.1. Predictive model using logistic regression

Logistic regression (LR) is a classic statistical model that is used to solve binary classification problems using logit as the link function (Rifaai et al., 2022).
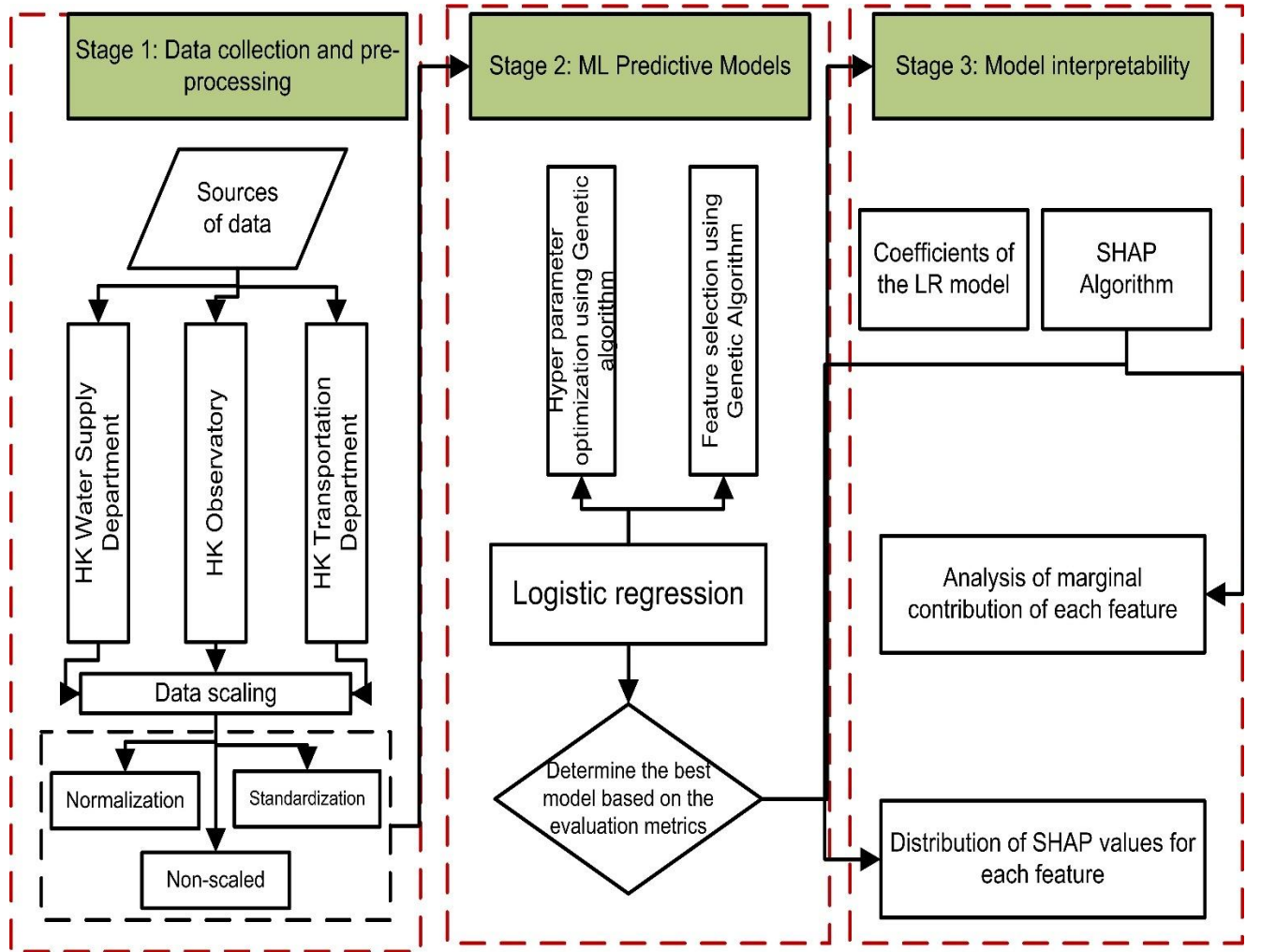
Figure 3. 4: Research framework for developing water pipes probability of failure models

Equation 3.6 mathematically represents the LR model (Robles-velasco et al., 2020).

$$p_i = \frac{1}{1+e^{-wx_i}} \hspace{6cm} (3.6)$$

where p_i is the probability of a sample belonging to a class, i denotes each observation in the dataset (i = 1, 2, 3,…N), x_i represents the vector of the independent variables, and their associated weights or coefficients are denoted by w. Equations 3.7-3.9 show that the response of the LR model is symmetrical.

$$P(y = 1 \mid x = x_i) = p_i; P(y = 0 \mid x = x_i) = 1 - p_i \hspace{3cm} (3.7)$$

$$1 - p_i = \frac{e^{-wx_i}}{1+ e^{-wx_i}} = \frac{1}{1+ e^{wx_i}} \tag{3.8}$$

$$p_i(x_i) = 1 - p_i(-x_i) \tag{3.9}$$

Equation 3.10, which shows the probability of a response, $y_i = (0,1)$, is used to assign a low probability to observations with $y_i = 0$ (i.e., pipelines with no failure) and a high probability to those observations having the characteristic of interest $y_i = 1$.

$$P(y_i) = p_i^{y_i}(1 - p_i)^{1-y_i} \tag{3.10}$$

The coefficients (also known as weights) of LR models are usually estimated by maximizing the log-likelihood function. The likelihood function can be obtained using Equation 8.

$$l = P\big(y_{1,\dots,}y_N\big) = \prod_{i=1}^{N} p_i^{y_i}(1 - p_i)^{1-y_i} \tag{3.11}$$

Equation 3.12 is derived by converting the likelihood function to the log-likelihood function and introducing $\log(\frac{p_i}{1-p_i}) = wx_i$ to Equation 3.11. After the weight estimation of each independent variable, the probability of failure of new observations can be computed using Equation 1. The class of the new observations can be determined using Equation 3.13. It should be noted that 0.5 is usually adopted as the risk threshold of many assets (Robles-velasco et al., 2020); however, it can be modified depending on the context of the problem.

$$Log(l) = \sum_{i=0}^{N} y_i wx_i - \log(1 + e^{wx_i}) \tag{3.12}$$

$$y_i = \begin{cases} 0 & if\ p_i \leq 0.5 \\ 1 & if\ p_i > 0.5 \end{cases} \tag{3.13}$$

3.4.2. Parameter optimization using genetic algorithm

GA is a biological evolutionary metaheuristic algorithm that works on the principle of natural selection and is capable of searching through the search space to provide an optimum solution to a problem. Due to its robustness, GA has been used to optimize various real-life models (Wu

et al., 2017; Liqian Yang et al., 2019). This research employs GA to optimize the hyperparameters of LR models and select the best features for predictions. Hence, the parameters that are optimized using GA are LR hyperparameters and features (i.e., input variables) of the model. GA is particularly useful when dealing with complex objective functions with multiple extrema. These functions can pose challenges for traditional optimization methods as they may converge to suboptimal solutions or get stuck in local optima. In contrast, the GA's population-based approach allows it to explore different regions of the search space simultaneously, increasing the chances of finding the optimal solution (Li Yang & Shami, 2020).

Figure 3.5 shows the schematic representation of how GA works. The algorithm works on the principle that the fittest individuals (using the adaptation function) will survive in an environment and pass the survival traits to the next generations. The steps involved in GA are summarized below (Liqian Yang et al., 2019).

Step 1 - Initialization of population: The first step is to randomly initialize a population of N individuals (i.e., chromosomes). Each chromosome is a potential solution to the problem of interest. The chromosome is denoted as a vector of the number of parameters to be estimated. For instance, if m parameters are to be estimated, the ith chromosome of the population can be defined by Equation 3.14. Consequently, Equation 3.15 represents the population of N chromosomes. The chromosomes are binary-coded, with 1 representing the presence of a parameter and 0 showing its absence.

$$X_i = \left[x_{1i}, x_{2i}, x_{3i}, \ldots, x_{pi} \ldots, x_{mi}\right]^T \tag{3.14}$$

$$Pop = [X_1, X_2, X_3, \ldots, X_i, \ldots X_N] \tag{3.15}$$

where x_{pi} is the pth parameter of the ith solution, and Pop refers to population.

Step 2 – Fitness function: The next step is to evaluate the fitness of each chromosome. The fitness value will determine if a chromosome will be selected for reproduction or not. Generally, chromosomes with high fitness values will be more competitive in the selection process for reproduction. In this study, accuracy and f1-score are used as fitness functions. More details about these functions are presented in Section 3.4.4.

Step 3 – Reproduction: To produce new offspring from the existing chromosomes, three processes are involved: selection, crossover, and mutation (Raharjo et al., 2019).

- **Selection**: This is the process of selecting the most competitive chromosomes for building a mating pool. This can be done using various strategies, including rank-based, truncation, roulette wheel, and tournament selection. Tournament selection was used because it helps to avoid early GA convergence. Early GA convergence often produces poor solutions.

- **Crossover**: At this stage, two parent chromosomes are randomly selected from the mating pool, and new offspring (i.e., chromosomes) are created by combining the genes of the two parent chromosomes with a crossover probability (p_c). Two-point crossover has been selected for this purpose.

- **Mutation**: The newly created chromosomes (i.e., offspring) are sent into a mutation operator. Mutation involves changing the gene of each chromosome such that each gene has a mutation probability (p_m). There are various mutation operators such as random setting, inversion, scramble, swap, and bit-flip mutation. However, bit-flip mutation was selected as it best suits the case of this study (i.e., binary encoded GA), and its effectiveness has been demonstrated in previous studies (Wu et al., 2017). Subsequently, the new offspring are added to a new population.

Step 4 – Elitism strategy: This strategy was used in the GA experiments to ensure that the fittest chromosomes (known as elites) were not lost from one generation to another. This implies that the fittest chromosomes in a previous population were copied into the next generation without undergoing crossover or mutation, thereby replacing the worst or least fitted chromosomes. To avoid early GA convergence, the elites consist of a small portion of the population. Steps 2 to 4 are repeated iteratively until optimum fitness is achieved (see Figure 3.5).

Based on various simulations, the parameters selected for all GA experiments in this study are enlisted in Table 3.2. The functions of each parameter have been described above.

Table 3. 2: Parameters of genetic algorithm

Parameters	Value
Population size	500
Fitness function	Accuracy and f1-score
Selection operator	Tournament selection (three chromosomes participate in each tournament)
Crossover operator	Two-point crossover
Mutation operator	Bit flip mutation
Crossover probability	0.8
Mutation probability	0.1
Number of iterations	3000
Elitist strategy	0.02

3.4.2.1. Optimization of logistic regression hyperparameters

Hyperparameters of a machine learning model are parameters that control the learning efficiency and weights of such a model. They are preselected before the model is trained. Hence, it is important to effectively search through the hyperparameters' space and select the best ones that fit the data. The optimization process usually includes four elements: a search space, an optimization algorithm to find the optimum hyperparameter combinations, an estimator with its objective function, and an evaluation function for comparing the efficiency of various hyperparameter combinations.
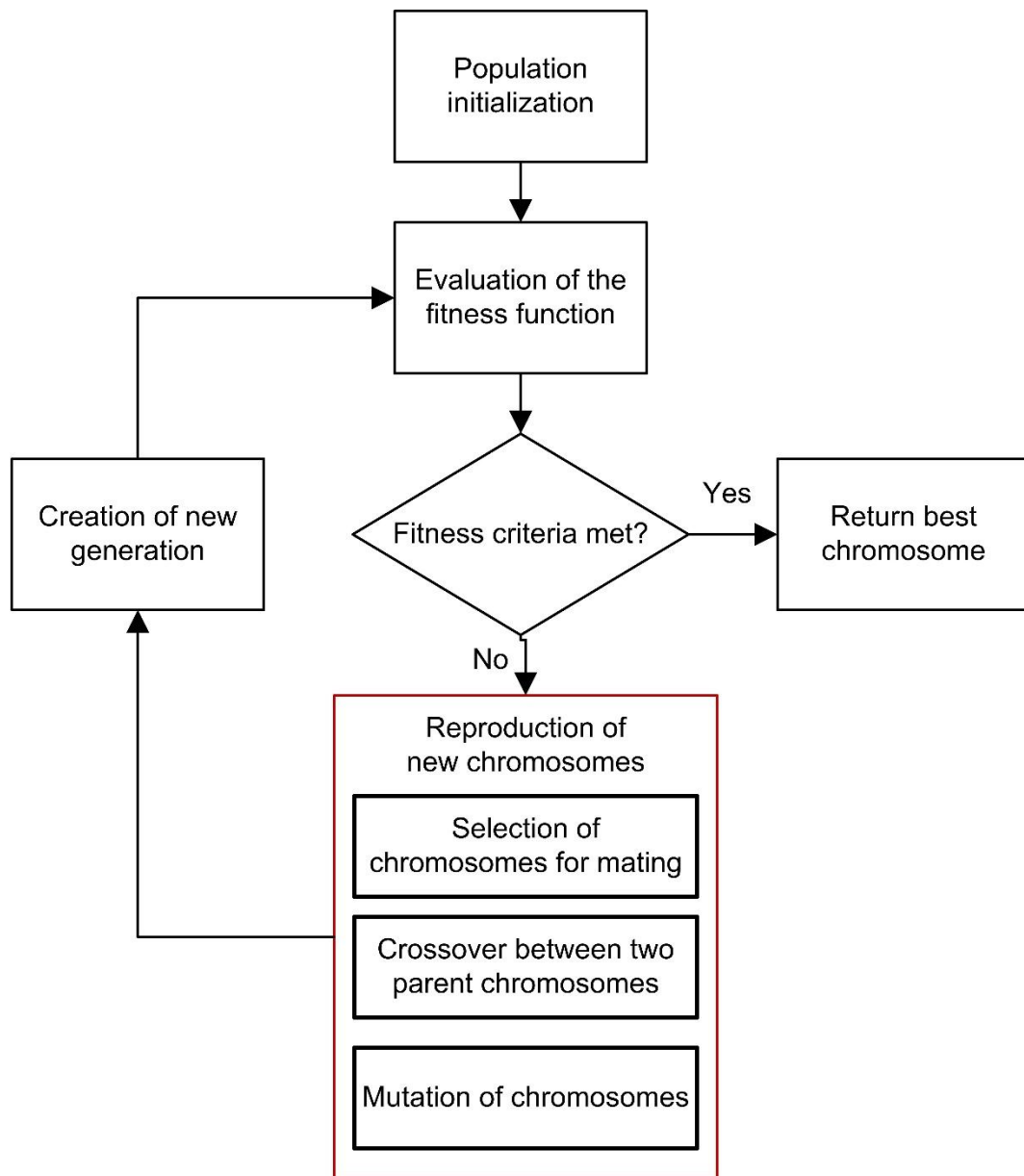
Figure 3. 5: Schematic representation of the genetic algorithm

In this research, GA was applied to optimize the hyperparameters of the LR models. The hyperparameters available in a typical LR model with their corresponding value range considered in this study are shown in Table 3.3. In this experiment, each chromosome in the genetic population has a value for the hyperparameters, which are randomly selected at

initialization. Figure 3.6 shows a typical chromosome for this experiment, which consists of 8 genes (i.e., hyperparameters).

Table 3. 3: Hyperparameters of the logistic regression model

Hyperparameter	Value
Penalty	[None, L1, L2, Elasticnet]
Dual	True or False
C	[1-500]
Solver	[lbfgs, newton-cg, liblinear, sag, saga]
Maximum iteration	[1-3000]
L1 ratio	[0-1]
Fit intercept	True or False
Verbose	[0-5]

Penalty	Dual	C	Solver	Max. Iteration	L1 ratio	Fit intercept	Verbose
L1	True	55.7	liblinear	2340	0.35	False	0

Figure 3. 6: A typical chromosome from experiment 1

3.4.2.2. **Optimization of logistic regression features**

Feature selection is an important aspect of ML modeling, as some features may be unnecessary and hence affect the performance of the model. Feature selection helps the model avoid overfitting, reduces computation time, and facilitates data visualization (Kotsiantis, 2014). The optimization process usually consists of two components: an optimization algorithm used to select the optimum subset of features and a performance criterion to identify the best features.

In this experiment, we extend the first experiment by using GA to optimally select the features (i.e., independent variables) needed for the LR models. Table 3.4 shows the structure of a typical chromosome in this experiment, consisting of 20 genes. The 21 genes are a summation of the LR hyperparameters and the independent variables. This implies that the optimization becomes more complex than experiment one, which increases the computation time. Since GA is a stochastic process that gives different results at different runs (i.e., simulations), the

program was run several times to determine the best feature combinations based on the performance criteria.

Table 3. 4: A typical chromosome in experiment 2

Parameter	Value
Penalty	L2
Dual	False
C	5
Solver	liblinear
Maximum iteration	2756
L1 ratio	0
Fit intercept	False
Verbose	0
Feature 1	1 (feature accepted)
Feature 2	0 (feature rejected)
Feature…	…
Feature 13	1 (feature accepted)

3.4.3.   Feature importance using SHapley Additive exPlanations

Compared to other ML algorithms, LR is often regarded as interpretable via its model coefficients. However, the interpretability is not direct, especially with the presence of multicollinearity, for non-statistical professionals in WDNs. Hence, SHapley Additive exPlanations (SHAP) is used in this study to explain the contribution of each independent variable to predictive models in a straightforward manner using clear visualizations (Lundberg & Lee, 2017).

SHAP is an explanatory algorithm based on game theory, presenting the contribution of each feature (i.e., independent variable) to a model's output in an intuitive graphical manner. After a model is developed, SHapley values of n features can be determined using Equation 3.16. In the context of game theory, the prediction is the game payout which is fairly distributed among the features (i.e., players). It should be noted that the SHAP value of a feature is relatively its contribution between the developed model $[f(X)]$ and a baseline model, often set as $E[f(X)]$,

where E represents a subset, and X refers to the explanatory variables. Further theoretical formulations of SHAP can be found in the paper authored by its originators (Lundberg & Lee, 2017).

$$\emptyset_i = \sum_{S \subset F} \frac{|S|!(|F|-|S|-1)!}{|F|!} [m(x_{S \cup i}) - m(x_S)] \qquad (3.16)$$

where

- $i = 1, 2, 3 \dots n,$

- the set that contains all the features is denoted by F,

- S represents a subset of F, and

- $m(.)$ represents a trained model.

The SHAP algorithm was implemented to complement the LR model's interpretability. The best model, in terms of the evaluation metrics, resulting from the first two experiments was interpreted using SHAP. Two tasks are involved in this experiment. The first task involves the determination of the marginal contribution of each feature to the model's output. The second task shows the direction of such contribution to the model's output to be either positive or negative.

All the experiments performed in this section of the study were conducted by dividing the dataset into two, with 80% being used for training and 20% for validation. The three main experiments are listed as follows:

- Optimization of logistic regression hyperparameters using genetic algorithm (experiment 1)

- Optimization of logistic regression hyperparameters and feature selection using genetic algorithm (experiment 2)

- Model interpretability using SHAP (experiment 3)

3.4.4.   **Evaluation metrics**

The output of the predictive models is a continuous number between 0 and 1, which denotes the failure probability of water pipes. As the dependent variable in the data set is binary-coded, a common approach to evaluating such models' output is establishing a class's threshold. A threshold value of 0.5 is chosen in this research. This implies that a pipe with a failure probability greater than 0.5 is predicted as a failed pipe. Based on the prediction and actual condition of the pipes (i.e., ground truth), a confusion matrix is established, as shown in Figure 3.7. In the Figure, "True Failure" represents the pipes that were correctly predicted as failed, while "False Failure" represents the pipes that were incorrectly predicted as failed, whereas they are intact. Similarly, "True Intact" denotes the pipes that were correctly predicted as intact, while "False Intact" denotes the pipes that were wrongly predicted as intact, whereas they have experienced failure. The "True Failure," "True Intact," "False Failure," and "False Intact" are generally known as "True Positives," "True Negatives," "False Positives," and "False Negatives," respectively. Based on the confusion matrix, five metrics are derived and used as evaluation indices in this study, as given in Equations 3.16 to 3.19. Figure 3.7 presents a domain-specific adaptation of the confusion matrix for water pipe failure classification. Unlike generic confusion matrices, this visualization explicitly shows how pipe conditions are categorized into failure/intact states, with specific implications for water utility decision-making. The customized matrix's structure can be related to practical consequences in WDN management - where false positives could lead to unnecessary maintenance costs, and false negatives could result in critical failures. This domain-specific interpretation is crucial for understanding the real-world implications of the model's predictions in water infrastructure management.

$$Accuracy = \frac{TF+TI}{TF+TI+FF+FI} \tag{3.16}$$

$$Precision = \frac{TF}{TF+FF} \tag{3.17}$$

$$Recall = \frac{TF}{TF+FI} \tag{3.18}$$

$$F1\ score = 2 \times \frac{1}{\frac{1}{Precision}+\frac{1}{Recall}} \tag{3.19}$$

where $TF, TI, FI, and\ FF$ denote "True Failure," "True Intact," "False Intact," and "False Failure." Accuracy measures the percentage of total correct predictions. This index might not be useful since equal importance is given to the two classes when the test data set is unbalanced, such as in the case of this study. Hence, recall, precision, and F1 score are useful in this case. Precision measures the ratio of correctly predicted failed pipes to all pipes predicted to have failed. On the other hand, recall measures the ratio of correctly predicted failed pipes to all pipes that have truly failed. From a practical point of view, high recall and precision values are desired, as they show the utility managers that truly failed pipes are correctly predicted. However, a low precision value indicates that a higher proportion of the pipes have been wrongly predicted as failed pipes, which may lead the utility managers to replace intact pipes, thereby unnecessarily increasing the maintenance cost. Moreover, a low recall value implies that a higher portion of failed pipes is misclassified as intact, misinforming utility managers about their maintenance schedule.

Furthermore, the F1 score finds the harmonic mean of precision and recall, producing information about the model's robustness (i.e., how many pipes are misclassified) and preciseness (i.e., how many pipes are classified correctly) (Tariq et al., 2022). Moreover, the Receiver Operating Curve (ROC) plots the True Positive Rate, TPR (i.e., recall), against the False Positive Rate, FPR (i.e., False Failure Rate) for different threshold values. Using the ROC curve, a model's performance in producing high TPR and low FPR is measured by the Area Under Curve (AUC). A perfect classifier will have an AUC of 1, while a classifier with an AUC

of 0.5 makes a random prediction. The five-evaluation metrics employed in this research have a value between 0 and 1. The closer the value to 1, the better the prediction.

		Predicted status	
		Predicted status is failure	Predicted status is intact
Actual status	Actual status is failure	True Failure (TF)	False Intact (FI)
	Actual status is intact	False Failure (FF)	True Intact (TI)

Figure 3. 7: Confusion matrix for classifying water pipe condition

3.5. DEVELOPMENT OF PROBABILITY OF LEAK AND BURST MODEL

The framework used to accomplish this objective is illustrated in Figure 3.8, which presents a comprehensive methodology for creating explainable deep learning (DL) algorithms for the probability of leak (POL) and probability of burst (POB) models. The framework is divided into five sequential steps, each contributing to creating a robust and interpretable predictive model. The first step is data preparation. This initial step involves the selection and processing of relevant data. The data is categorized into three types: pipe-related, environment-related, and operation-related. These datasets are then divided into a training (70% of the data) and a validation set (30%). Key pre-processing tasks such as data cleaning, outlier removal, imputation for missing values, normalization, and standardization are applied to ensure the data is suitable for training the DL models. Individual cleaning is performed on training and testing

datasets separately to prevent data leakage. The training and validation subset is subjected to a 10-fold cross-validation procedure to optimize model generalizability. In this process, the data is split into ten folds, wherein one fold serves as the validation set while the remaining nine folds comprise the training set. In the second step, DL architectures such as DNN, CNN, and TabNet are considered. Bayesian optimization is used to tune the hyperparameters of these models. This process involves training surrogate models on a training dataset and using them to predict the performance of the DL models on a validation dataset. The optimization process iteratively proposes new points (sets of hyperparameters) to find the set that maximizes performance on the validation set. In the third stage, the models' performance is assessed using a set of evaluation metrics: Accuracy, Recall, Precision, F1 score, Matthews Correlation Coefficient (MCC), and Cohen's Kappa. These metrics provide a broad overview of the models' predictive capabilities and performance. Subsequently, the Copeland method, a pairwise comparison ranking algorithm, is utilized to rank the models. Each model's performance is compared against the others', with wins, losses, and Copeland scores (the difference between the number of wins and losses) being calculated. This score effectively ranks the models according to their predictive abilities. The final step focuses on providing explainability for the chosen deep learning model using SHAP. The analysis of marginal contributions allows for a better understanding of the model's decision-making process. Additionally, the distribution of SHAP values can be used to identify the impact of the features on the model to be either positive or negative. The detail of the proposed methodology is explained in subsequent sections.

3.5.1. **Predictive model using deep learning algorithms**

3.5.1.1. **Deep Neural Network**

The Deep Neural Network (DNN) is a foundational model in the domain of DL algorithms.
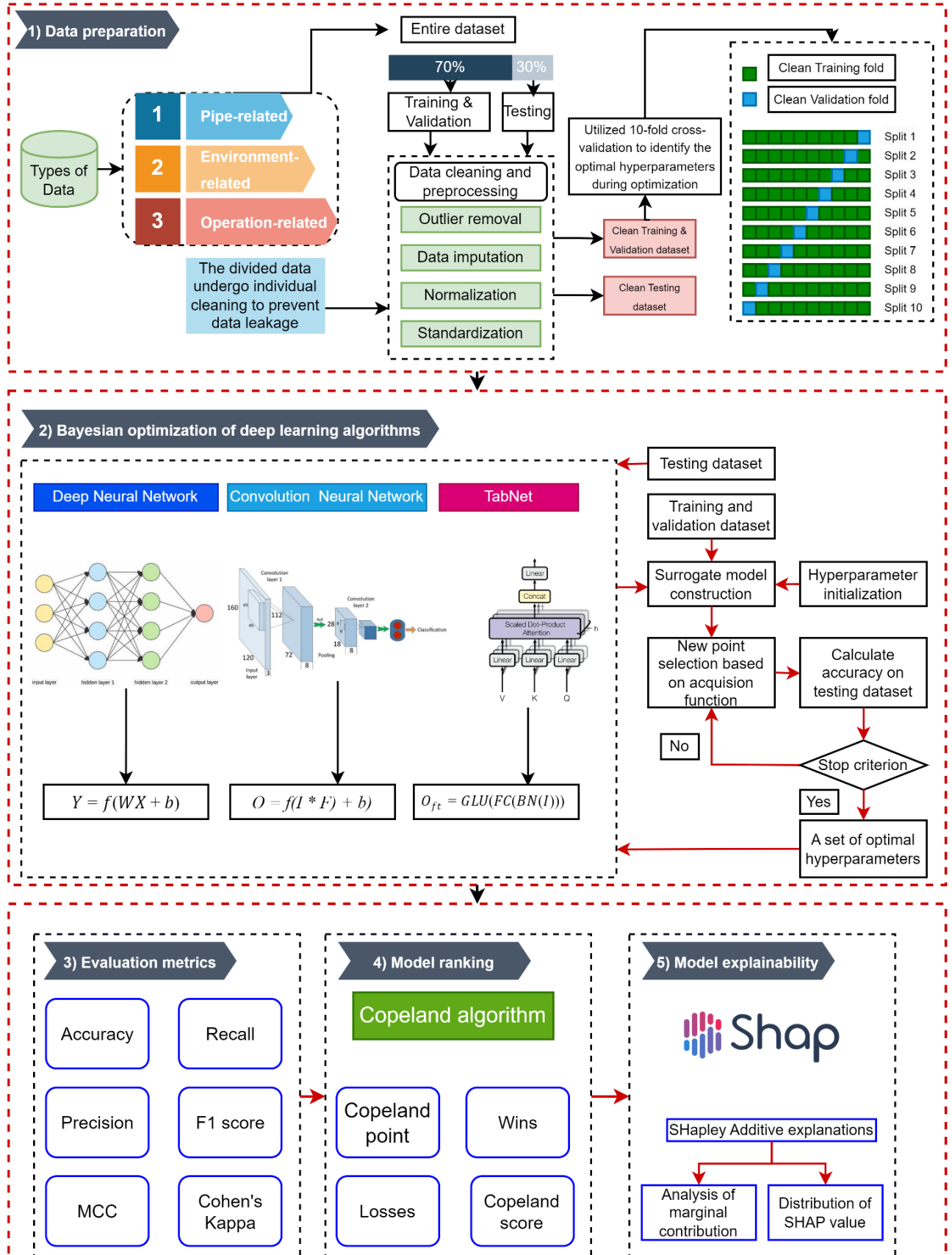
Figure 3. 8: Framework for predicting probability of leak and burst

DNNs are characterized by their depth, which comprises multiple hidden layers between the input and output layers, enabling them to model complex and high-level abstractions in data (Jun et al., 2017).

A standard DNN architecture is composed of an input layer X, multiple hidden layers H, and an output layer Y. Each layer consists of units or neurons, and each neuron in one layer is connected to every neuron in the subsequent layer, forming a dense network. The input layer receives the feature vectors derived from the data, which are then processed through the hidden layers using a series of weighted summations and non-linear activation functions.

The mathematical operations within a typical hidden layer l can be represented as follows:

$$\boldsymbol{H}^{(l)} = \sigma\left(\boldsymbol{W}^{(l)}\boldsymbol{H}^{(l-1)} + \boldsymbol{b}^{(l)}\right) \tag{3.20}$$

where $H^{(l-1)}$ is the output of the previous layer or the input data for the first hidden layer, $W^{(l)}$ denotes the weight matrix, $b^{(l)}$ is the bias vector, and σ represents the non-linear activation function, such as ReLU or sigmoid, applied element-wise. This process is iteratively conducted across all hidden layers.

The output layer of DNN provides the final prediction, framed as a classification task where the objective is to predict the probability of leak and burst for water pipes. For such classification tasks, the output layer typically employs an activation function, which outputs a probability distribution across the classes, which is represented mathematically as:

$$\boldsymbol{Y} = \sigma\left(\boldsymbol{W}^{(output)}\boldsymbol{H}^{(last)} + \boldsymbol{b}^{(output)}\right) \tag{3.21}$$

Here, Y is the vector of probabilities that the pipe section falls into each of the possible outcome classes.

3.5.1.2. **Convolution Neural Network**

The Convolutional Neural Network (CNN) architecture is specifically designed to process the multi-dimensional data associated with the characteristics of water pipes to predict their leakage/burst probabilities. The input layer is structured to receive pre-processed feature matrices representing the pipe-related, environment-related, and operation-related factors. The convolutional layers form the core of the CNN architecture. Multiple convolutional layers are employed to extract and learn features from the input data (Raziani & Azimbagirad, 2022; Tsai et al., 2022). The first convolutional layer applied a set of learnable filters (kernels) $\boldsymbol{W}^{(l)}$, where l represents the layer number. The convolution operation at each layer l for a given input matrix $\boldsymbol{X}^{(l)}$ is defined in Equation 3.22

$$\boldsymbol{F}^{(l)} = \boldsymbol{W}^{(l)} * \boldsymbol{X}^{(l)} + \boldsymbol{b}^{(l)} \tag{3.22}$$

where $\boldsymbol{F}^{(l)}$ is the feature map obtained after applying the kernel $\boldsymbol{W}^{(l)}$ to the input $\boldsymbol{X}^{(l)}$, and $\boldsymbol{b}^{(l)}$ represents the bias.

After each convolution operation, an activation function is applied to introduce non-linearity, enabling the network to learn complex patterns. Following the convolutional layers, pooling layers are used to reduce the spatial size of the representation, decreasing the number of parameters and computations in the network. Max pooling is utilized, which can be defined using Equation 3.23.

$$\boldsymbol{P}_{i,j}^{(l)} = max_{m,n \in M_{i,j}} \boldsymbol{A}_{m,n}^{(l)} \tag{3.23}$$

where $P^{(l)}$ is the pooled feature map and $M_{i,j}$ is the region in the activated feature map $A^{(l)}$ over which the pooling operation is performed.

The high-level reasoning in the network is performed by fully connected layers. The output from the final pooling layer is flattened and fed into a series of fully connected layers. The

operation at a fully connected layer l with the input vector $\boldsymbol{v}^{(l)}$ is given in Equations 3.24 and 3.25.

$$z^{(l)} = \boldsymbol{W}^{(l)} \boldsymbol{v}^{(l)} + \boldsymbol{b}^{(l)} \qquad (3.24)$$

$$o^{(l)} = \sigma\left(z^{(l)}\right) \qquad (3.25)$$

where $z^{(l)}$ is the linear combination of weights $\boldsymbol{W}^{(l)}$, biases $\boldsymbol{b}^{(l)}$, and input $\boldsymbol{v}^{(l)}$ and $o^{(l)}$ is the output after applying the activation function σ. It should be noted that the final fully connected layer acted as the output layer, with a single neuron using the sigmoid activation function to predict the probability p of pipe leakage or burst, given by Equation 3.26:

$$p = \frac{1}{1 + e^{z(output)}} \qquad (3.26)$$

where z^{output} is the input to the output neuron

3.5.1.3. **TabNet**

TabNet is a novel DL architecture for tabular data that uses sequential attention to learn which features to focus on during the learning process (Arık & Pfister, 2021). Figure 3.9 gives a schematic representation of TabNet architecture, which uses an encoder composed of multiple steps to determine relevant features from the data and generates a feature representation (H. V. Nguyen & Byeon, 2023). This representation is then aggregated to assist in decision-making. The model's input, consisting of batch-sized data with D-dimensional features, undergoes batch normalization before being processed by the feature transformer.

The feature transformer is structured with multiple gated linear unit (GLU) blocks. These GLU blocks, which include fully connected and batch normalization layers, are designed for robust learning, with some blocks being shared and others independent. To maintain stability and control variance, normalization is applied after each GLU block. The resulting transformed features are then fed into the attentive transformer.

The attentive transformer is composed of fully connected and batch normalization layers, followed by a prior scale and sparsemax layer. It uses the information from the prior step to calculate a mask layer for the current step, as shown in Equations 3.27 and 3.28:

$$P[i] = \prod_{j}^{i}(\gamma - M[j]) \tag{3.27}$$

$$M[i] = sparsemax\ (P[i-1] * h_i(a[i-1])) \tag{3.28}$$

where $P[i]$ represents the 'prior scale' at the i-th decision step, γ is the relaxation parameter, the mask from previous layers is represented by $M[j]$ and $h_i(\cdot)$ represents the trainable function of the fully connected and batch normalization layers. Sparsemax is used to ensure the sum of the mask coefficients is 1, contributing to sparse feature selection.

To control the sparsity of the features, sparsity regularization L_{sparse} is introduced in the form of entropy to the loss function, adding a small number $\in$ for numerical stability, which is depicted in Equation 3.29.

$$L_{sparse} = \sum_{i=1}^{N_{steps}} \sum_{b=1}^{B} \sum_{j=1}^{D} M_{bj}[i] \log(M_{bj}[i]) + \in \tag{3.29}$$

The feature transformer takes the masked features and outputs them for both the decision process and the next attentive transformer step, using Equation 3.30.

$$[d[i], a[i] = f_i(M[i] * f) \tag{3.30}$$

where $[d[i]$ is the decision step output, and $a[i]$ is the information for the subsequent step.

Regarding the interpretability, TabNet computes the importance of each step through an aggregation of the output vector, which is converted into a scalar, reflecting the step's significance to the final outcome. Local feature importance for a sample is derived by summing the results across all steps, and global feature importance is calculated using an aggregate mask.
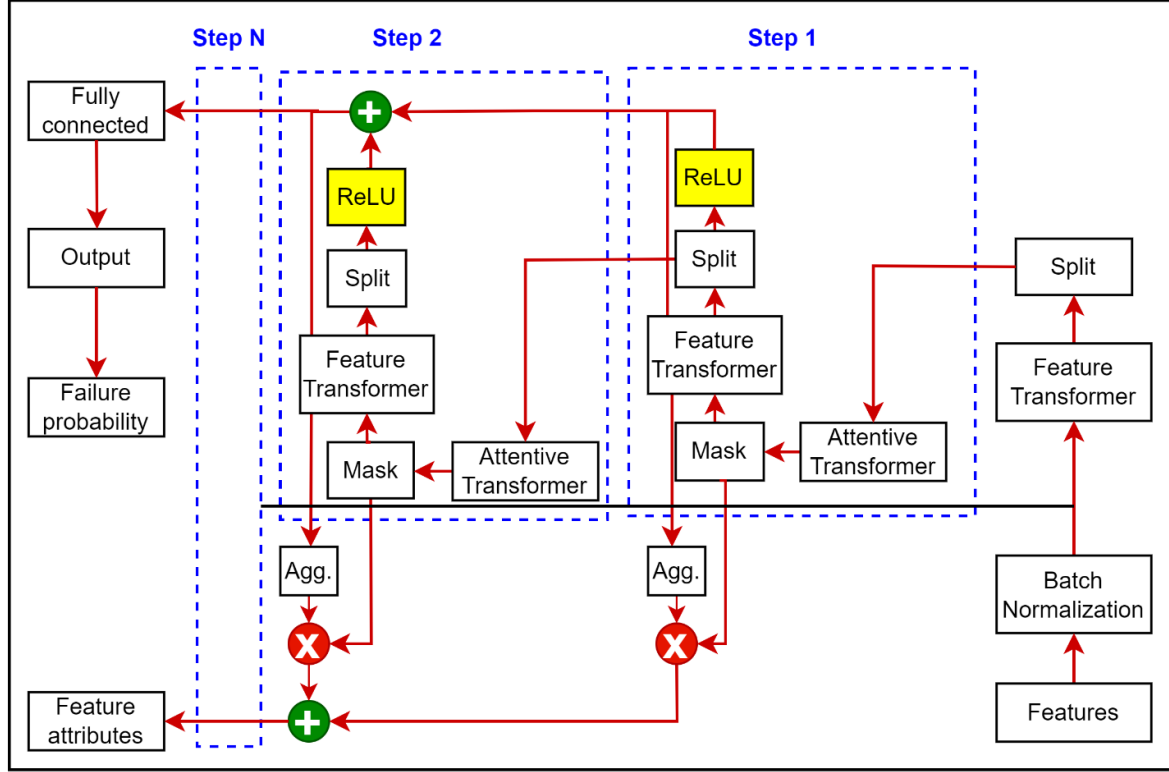
Figure 3. 9: Schematic representation of the TabNet architecture

3.5.2. **Hyperparameter optimization using Bayesian Optimization**

After building the DL models, their hyperparameters are optimized using Bayesian Optimization (BO), a strategy that enhances model performance by efficiently navigating the hyperparameter space. The hyperparameters and search space optimized for these models are presented in Table 3.5. BO seeks to identify the optimal set of parameters for a model. The approach aims to find the maximum value of an unknown objective function $\theta(p)$, at a given point p, within a defined search space Ω. The optimal sampling point can be represented by Equation 3.31 (Bello et al., 2023; Taiwo, Ben Seghier, et al., 2023a; Li Yang & Shami, 2020).

$$p^+ = \underset{p \in \Omega}{\mathrm{argmax}}\, \theta(p) \tag{3.31}$$

The BO process involves two key steps:

1. A surrogate probabilistic model is fitted to the objective function, often using a Gaussian Process (GP), which is then updated as new data points are sampled. This model is preferred for its flexibility, robustness, and analytic tractability.

2. An acquisition function is constructed from the posterior distribution of the surrogate model to balance the search space exploration with the exploitation of known good regions. The Expected Improvement (EI) is a common choice for the acquisition function. The optimization process iterates, continually updating the surrogate model with new findings, until a predefined stopping criterion, typically the maximization of the acquisition function, is met.

Table 3. 5: Details of the hyperparameter optimization

Model	Hyperparameter	Type	Range
DNN	Batch size	Integer	[8, 128]
	Epochs	Integer	[2, 50]
	Number of neurons	Integer	[8, 64]
	Optimizer	Categorical	[Adam, SGD, Adagrad, RMSprop]
	Learning rate	Continuous	[0.01, 1]
CNN	Batch size	Integer	[8, 128]
	Epochs	Integer	[2, 50]
	Number of neurons	Integer	[8, 64]
	Optimizer	Categorical	[Adam, SGD, Adagrad, RMSprop]
	Number of filters	Integer	[8, 128]
	Kernel size	Integer	[2, 20]
	Learning rate	Continuous	[0.01, 1]
TabNet	n_d	Integer	[8, 128]
	n_a	Integer	[8, 128]
	n_steps	Integer	[8, 128]
	gamma	Continous	[1, 10]
	lambda_sparse	Continuous	[0.01, 1]
	Batch size	Integer	[8, 128]
	Step size	Continuous	[0.01, 1]

3.5.3. Evaluation of the DL models

The predictive algorithms generate a continuous score ranging from 0 to 1, reflecting the likelihood of a pipe experiencing a leak or burst. Given that the target variable within the dataset is dichotomous, it is standard practice to select a cutoff point for categorizing the outcomes (Robles-velasco et al., 2020; Taiwo et al., 2024a). In this study, pipes with predicted failure probabilities above 0.5 are classified by the model as failures, while those below 0.5 are classified as non-failures. A confusion matrix is constructed to compare the model's predictions against the actual conditions of the pipes, as depicted in Figure 3.10 (Mazumder, Salman, & Li, 2021; Robles-velasco et al., 2020). This matrix is then used to calculate six different performance metrics, which are detailed in Table 3.6.

		Predicted status	
		Predicted status is leak/burst	Predicted status is intact
Actual status	Actual status is leak/burst	True Leak/Burst (TL/B)	False Intact (FI)
	Actual status is intact	False Leak/Burst (FL/B)	True Intact (TI)

Figure 3. 10: Confusion matrix for classifying water pipe status

3.5.4. Ranking of the DL models

The Copeland method, a paired comparison approach, is employed to rank the performance of the DL models. This method involves comparing each model against every other in a series of head-to-head matchups. The process of evaluation is highlighted as follows (Furxhi et al., 2019):

Table 3. 6: Evaluation metrics for the DL models

Evaluation metric	Mathematical expression
Accuracy	$\dfrac{TL/B + TI}{TL/B + TI + FL/B + FI}$
Precision	$\dfrac{TL/B}{(TL/B + FL/B)}$
Recall	$\dfrac{TL/B}{(TL/B + FI)}$
F1 score	$2 * \dfrac{(Precision * Recall)}{(Precision + Recall)}$
MCC	$\dfrac{(TL/B \times TI - FL/B \times FI)}{\sqrt{TL/B + FL/B} \times (TL/B + FI) \times (TI + FL/B) \times (TI + FI)}$
Cohen's Kappa	$\dfrac{Accuracy - Expected\ accuracy}{1 - Expected\ accuracy}$

$$Expected\ accuracy = \frac{((TL/B + FI) \times (TL/B + FL/B) + ((TI + FL/B) + (TI + FI)}{(TL/B + TI + FL/B + FI)^2}$$

1. **Pairwise Comparisons**: Each DL model is compared with every other model for each performance metric. In these comparisons, models are awarded points based on their performance relative to one another.

2. **Points Allocation**: A model earns a point for each performance metric where it outperforms another model. For instance, if Model A has a higher precision than Model B in their comparison, Model A receives a point.

3. **Win/Loss Record**: The outcome of each pairwise comparison is a win, loss, or tie for the models involved. A win is recorded for a model if it accrues more points than the other model in their comparison. Similarly, a loss is noted when a model earns fewer points than its competitor, while a tie is considered if both models accumulate an equal number of points.

4. **Copeland Score Calculation**: The Copeland score for each model is calculated by subtracting the total number of losses from the total number of wins:

$$Copeland\ Score = Total\ Wins - Total\ Losses \qquad\qquad (3.32)$$

5. **Ranking of Models**: Models are then ranked based on their Copeland scores, with the model boasting the highest score at the top rank, indicating it has the best performance across the evaluated metrics.

After determining the most effective model using the Copeland ranking method, the SHAP framework is utilized to interpret the chosen model. This framework provides insights into how each feature contributes to the predictive outcome, offering a clear explanation of the model's decision-making process. A detailed explanation of the SHAP framework is provided in Section

3.6. DEVELOPMENT OF CAUSES OF FAILURE PREDICTIVE MODEL

As per the fourth objective of this study, this section explains the methodologies adopted to develop an optimized model for predicting the causes of the failure (COF) of water pipes. The proposed framework employed to achieve this objective is shown in Figure 3.11. This comprises five sequential phases, including data preparation, ML modelling, evaluation metrics, model ranking, and model explainability. The data preparation stage shows that the dataset comprises three categories: pipe-related, environment-related, and operation-related data. The data preparation phase involves several steps, including outlier removal, data imputation for handling missing values, and standardization of the data. The cleaned dataset is then split into training (80%) and testing (20%) subsets. The training data is further divided using 10-fold cross-validation to identify the optimal hyperparameters during model optimization.

The next phase focuses on the optimization of various ML algorithms, including XGBoost, AdaBoost, Random Forest, LightGBM, and CatBoost, using the Tree-structured Parzen Estimator (TPE) technique. The TPE algorithm is employed to find the optimal set of

hyperparameters for each model. This optimization process follows an iterative approach, where a surrogate model is initially constructed using the training and validation datasets. Subsequently, the TPE algorithm selects new hyperparameter values based on an acquisition function, which guides the search towards promising regions of the hyperparameter space. The performance of the optimized models is evaluated using various metrics, including accuracy, macro precision, macro recall, macro F1 score, weighted precision, weighted recall, weighted F1 score, and average area under the curve (AUC). To systematically select the best-performing model, the Copeland algorithm is employed. This algorithm ranks the models based on their pairwise comparisons, considering factors such as Copeland points, wins, and losses. The ranking helps identify the top-performing model for predicting the causes of water pipe failures. The last phase deals with model interpretability. This is achieved through the use of SHapley Additive exPlanations (SHAP), a game-theoretic approach that provides insights into the marginal contribution of each feature to the model's output. The SHAP analysis involves examining the distribution of SHAP values for each of the features to understand their impact on the predictive model. The specifics of the proposed methodology are described in the subsequent sections.

3.6.1. **Predictive model using ensemble learning**

To develop accurate and robust predictive models for identifying the causes of water pipe failures, this study employs five state-of-the-art ensemble learning (EL) techniques: AdaBoost, Random Forest, XGBoost, LightGBM, and CatBoost. These ensemble models have demonstrated superior performance in various prediction tasks by combining multiple individual models, thereby reducing bias, variance, and overfitting issues (Hancock & Khoshgoftaar, 2020; Winkler et al., 2018). For the multi-class classification problem of predicting pipe failure causes, the one-vs-rest (OvR) approach was employed to build the classifiers.
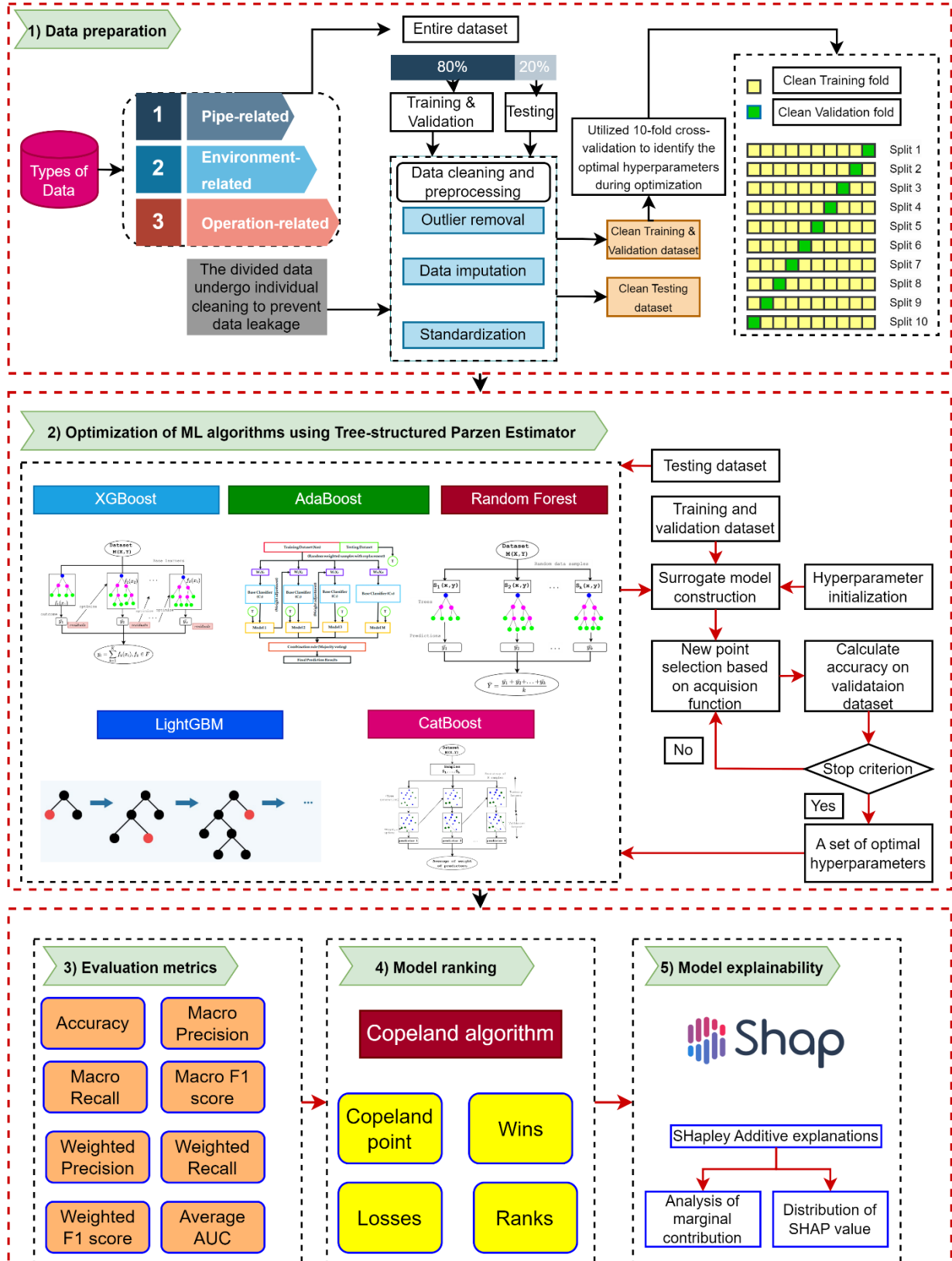
Figure 3. 11: Framework for predicting causes of water pipe failure

In the OvR strategy, a separate binary classifier is trained for each class, discriminating that class from the combined rest of the classes. During prediction, the class with the highest confidence score from the respective binary classifier is assigned as the predicted class.

3.6.1.1. **AdaBoost**

AdaBoost, short for Adaptive Boosting, is an ensemble learning (EL) algorithm proposed by Freund and Schapire in 1997 (Freund & Schapire, 1997). It is a meta-algorithm that combines multiple weak learners (e.g., decision trees) into a strong classifier. AdaBoost operates in an iterative manner, training successive weak learners on instances that were misclassified by the previous learners, thereby focusing on the difficult-to-classify examples. The schematic representation of AdaBoost is shown in Figure 3.12.

The AdaBoost algorithm works as follows:

1. Initialize the weights for each observation $w_i = \frac{1}{N}$ for $i = 1, 2, \ldots, N$ where N is the total number of observations.

2. For each iteration $m = 1 \ to \ M$ (where M is the number of weak learners):

   a. Train a weak learner $G_m(x)$ using the current weights w_i on the training data.

   b. Calculate the weighted error rate of the weak learner:

   $$err_m = \sum_{i=1}^{N} w_i \prod(y_i \neq G_m(x_i)) \tag{3.33}$$

   c. Compute the weight for the weak learner:

   $$\alpha_m = \log\left(\frac{1 - err_m}{err_m}\right) \tag{3.34}$$

   d. Update the weights for each observation:

   $$w_i \leftarrow w_i \cdot \exp(\alpha_m \cdot \prod(y_i \neq G_m(x_i))) \tag{3.35}$$

   e. Normalize the weights for each observation:

   $$w_i \leftarrow \frac{w_i}{\sum_{j=1}^{N} w_j} \tag{3.36}$$

3. The final model is a strong classifier that combines the weak learners through a weighted majority vote:

$$G(x) = sign\left(\sum_{m=1}^{M} \alpha_m G_m(x)\right) \qquad (3.37)$$
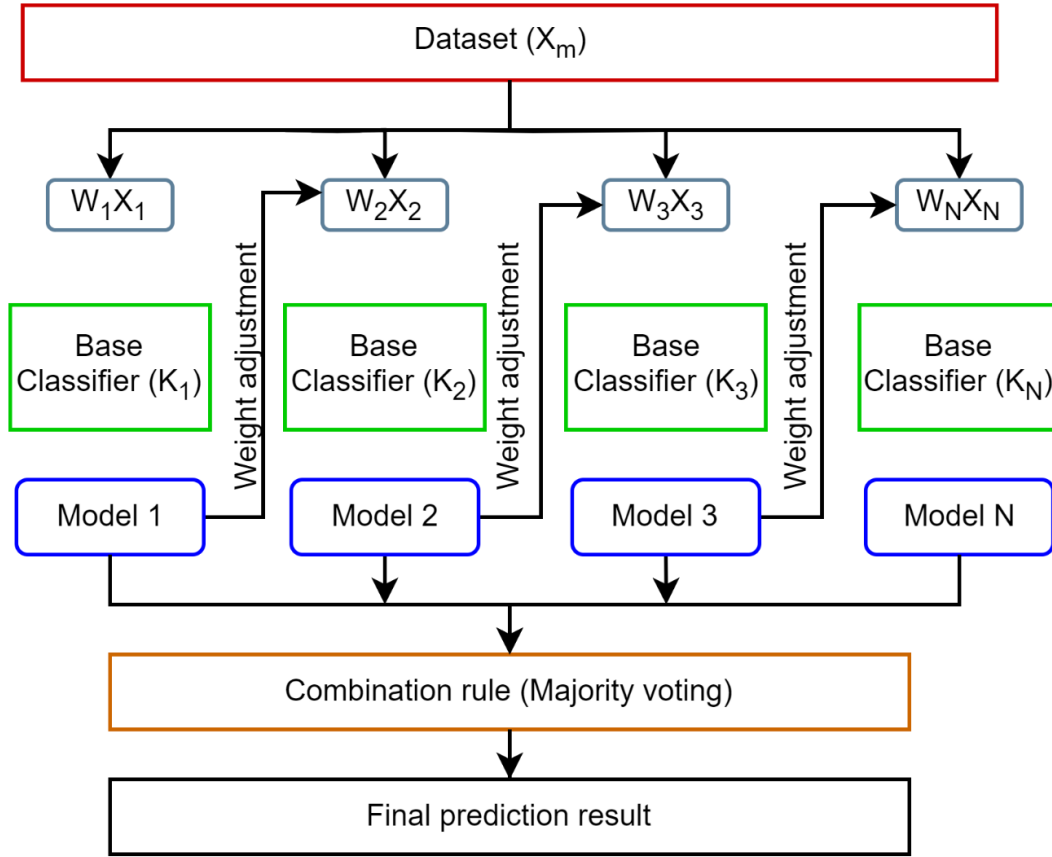
Figure 3. 12: Schematic representation of AdaBoost algorithm

3.6.1.2. **Random Forest**

Random Forest (RF) is another EL algorithm that constructs multiple decision trees and combines their predictions through majority voting for classification tasks (Breiman, 2001). It is a bagging-based ensemble method, where each decision tree is trained on a different bootstrap sample of the original data, and a random subset of features is considered for splitting at each node (see Figure 3.13). This randomization process helps to reduce the variance and overfitting issues associated with individual decision trees.

The Random Forest algorithm works as follows:

1. Draw n bootstrap samples from the original data, each containing N instances sampled with replacement.

2. For each bootstrap sample i ($i = 1, 2, \ldots, n$):

   a. Grow an unpruned decision tree T_i on the bootstrap sample, with the following modification: at each node, instead of considering all features, randomly select m features from the total M features ($m \ll M$). Choose the best feature from the m features to split the node.

   b. Repeat the process until the minimum node size is reached or the tree is fully grown.

3. After constructing all n trees, predictions for new instances are made by aggregating the predictions of individual trees:

$$\hat{C}(x) = majority\ vote\ \{C_i(x)\}_{i=1}^{n} \tag{3.38}$$

In the above equation, $\hat{C}(x)$ represents the predicted class for a new instance x. It should be stated that during training, each decision tree learns to discriminate between one class and the combined rest of the classes. During prediction, the class with the highest number of votes from the ensemble of trees is assigned as the predicted class.

3.6.1.3. **XGBoost**

Extreme Gradient Boosting (XGBoost) is a highly efficient and scalable implementation of the gradient boosting framework. It is an EL algorithm that sequentially builds weak decision tree models, with each successive model attempting to correct the errors made by the previous ones (T. Chen & Guestrin, 2016). XGBoost has gained widespread popularity due to its superior performance, parallelization capabilities, and ability to handle various data types, including sparse and complex data. Figure 3.14 gives a schematic representation of the algorithm.
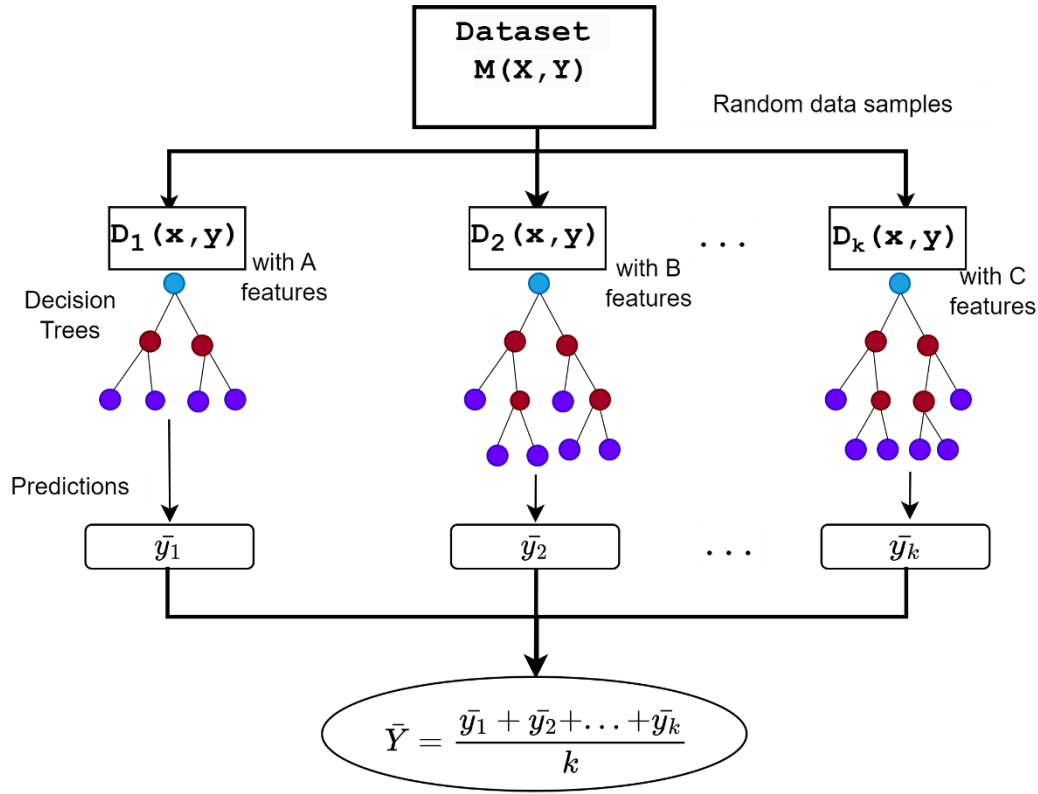
Figure 3. 13: A visual diagram illustrating the Random Forest algorithm

The XGBoost algorithm aims to minimize the following regularized objective function:

$$\mathcal{L}^{t} = \sum_{i=1}^{n} \ell[(\mathbf{y_i}, \hat{\mathbf{y}_i})^{(t-1)} + f_t(\mathbf{x_i})] + \Omega(f_t) \tag{3.39}$$

where l is a differentiable loss function (e.g., logistic loss for classification), y_i and $\hat{y}_i^{t-1}$ are the true and predicted values for the ith instance, respectively, f_t is the new tree model to be added, and $\Omega(f_t)$ is the regularization term that controls the complexity of the tree model.

The regularization term $\Omega(f_t)$ is defined as:

$$\Omega(f_t) = \gamma T + 0.5\lambda \sum_{j=1}^{T} w_j^2 \tag{3.40}$$

where T is the number of leaves in the tree, w_j is the score associated with the jth leaf, and γ and λ are hyperparameters controlling the pruning and ridge regularization, respectively.

The boosting process in XGBoost involves the following steps:

1. Initialize the predictions $\hat{y}_i^{(0)}$ with a constant value (the log-odds ratio is employed).

2. For iterations $t = 1,2,\ldots,T$:

   a. Compute the residuals $r_{ti} = \dfrac{\partial l(y_i,\hat{y}_i^{t-1})}{\partial \hat{y}_i^{t-1}}$ for all instances $i = 1,2,\ldots,n$.

   b. Fit a decision tree model $f_t(x)$ to the residuals r_{ti} by minimizing the regularized objective function.

   c. Update the predictions: $\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta\, f_t(x_i)$, where η is the learning rate.

During prediction, the class with the highest confidence score from the respective binary classifier is assigned as the predicted class.
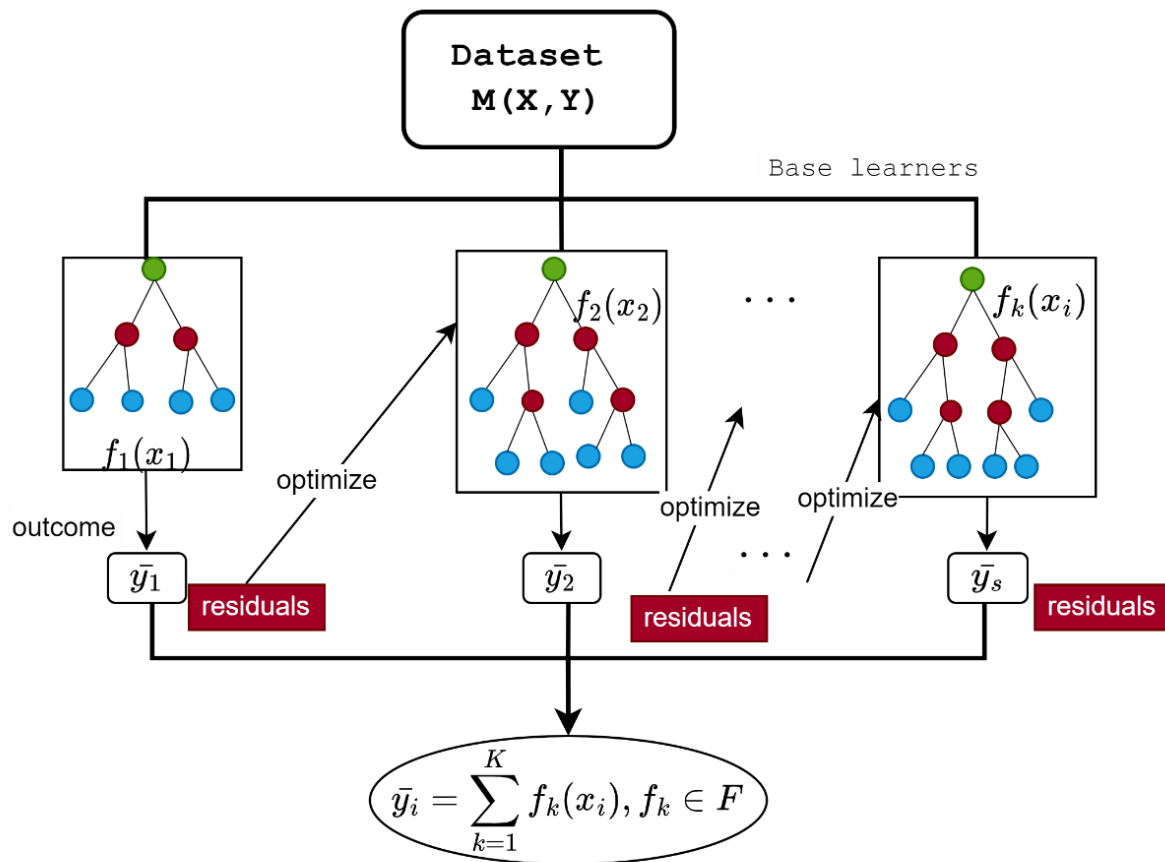
Figure 3. 14: A schematic representation of XGBoost

3.6.1.4. **LightGBM**

Light Gradient Boosting Machine (LightGBM) is a gradient-boosting framework that uses tree-based learning algorithms (Ke et al., 2017). While it shares similarities with XGBoost in the underlying gradient boosting principles, LightGBM introduces several novel techniques to improve computational efficiency, particularly for large-scale data.

Gradient-based One-Side Sampling (GOSS) is utilized to decrease the number of data instances needed for gradient computations by sorting instances based on their gradient magnitudes and retaining a small portion with large gradients, while randomly sampling instances with smaller gradients to preserve the data distribution (C. Chen & Seo, 2023). Exclusive Feature Bundling (EFB) is another technique where mutually exclusive features, which are rarely non-zero at the same time, are combined into a single feature, thus reducing dimensionality without significant information loss, making it particularly effective for sparse datasets. Unlike traditional algorithms that grow trees level by level, LightGBM adopts a leaf-wise tree growth strategy (see Figure 3.15), selecting the leaf for splitting that minimizes loss, resulting in potentially deeper and more complex trees but often with improved accuracy and lower computational costs. Additionally, LightGBM employs histogram-based split finding, where histograms represent gradient statistics of features, allowing for quicker split finding and reduced memory usage compared to precise split finding methods.
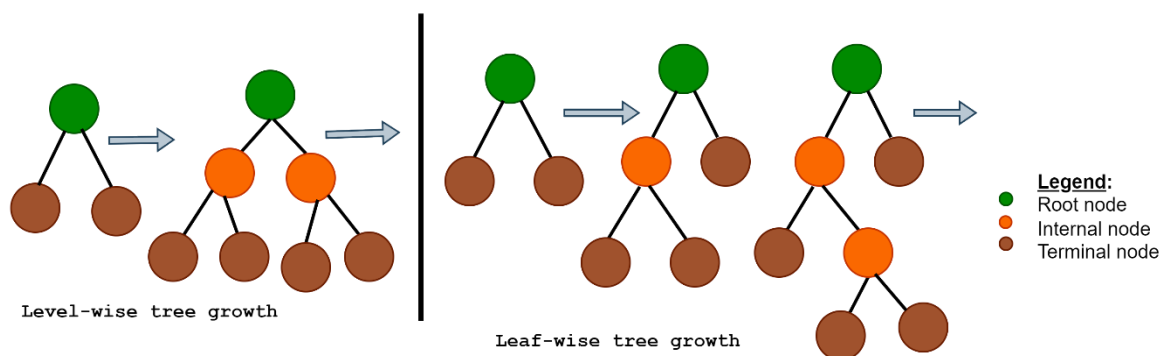
Figure 3. 15: Tree Growth by Levels (Left) and Tree Growth by Leaves (Right)

3.6.1.5. **CatBoost**

Categorical boosting (CatBoost) is an open-source ML algorithm developed in 2017 (Prokhorenkova et al., 2018). It is a gradient-boosting decision tree algorithm that constructs an ensemble of decision trees sequentially, with each tree aiming to correct the errors of its predecessors (see Figure 3.16). A standout feature of CatBoost is its advanced handling of categorical features through Ordered Target Encoding (OTE). This technique replaces categorical values with corresponding target statistics, reducing the risk of overfitting and effectively managing high-cardinality categorical features.

The training process of CatBoost for multi-class classification involves:

1. **Loss Function**: CatBoost supports multiple loss functions for multi-class classification, such as cross-entropy loss or multi-class logistic loss. For a dataset with N instances and K classes, the multi-class logistic loss is defined as:

$$L(y, f) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} [y_{ik} \log(p_{ik})] \tag{3.41}$$

Where y_{ik} is an indicator variable (0 or 1) indicating whether instance i belongs to class k, and p_{ik} is the predicted probability of instance i belonging to class k, obtained from the model's output $f(x_i)$.

2. **Gradient Boosting**: CatBoost iteratively builds an ensemble of decision trees by minimizing the loss function. At each iteration t, a new decision tree $h_t(x)$ is added:

$$F_t(x) = F_{t-1}(x) + \eta \times h_t(x) \tag{3.42}$$

where $F_t(x)$ is the updated ensemble model, η is the learning rate, and $h_t(x)$ is the new decision tree trained to approximate the negative gradient of the loss function.

3. **Ordered Target Encoding**: This technique replaces categorical values with target statistics in a specific order to prevent data leakage and overfitting.

4. **Regularization**: CatBoost includes L2 regularization, random subsampling of instances (bagging), and random subsampling of features (feature bagging) to prevent overfitting.

5. **Overfitting Detection**: The algorithm uses a holdout set or out-of-fold instances to monitor and prevent overfitting during training.

6. **Multi-Class Prediction**: After training, CatBoost makes multi-class predictions by applying the decision tree ensemble to new instances and computing the predicted probabilities for each class, with the highest probability determining the final prediction.

3.6.2. Hyperparameter optimization using Tree-Structured Parzen Estimator

Hyperparameter tuning is a crucial step in optimizing the performance of ML models. In this study, the Tree-Structured Parzen Estimator (TPE) approach, implemented in the Optuna library, is employed to optimize the EL algorithms (Akiba et al., 2019; Models et al., 2024). TPE is a sequential model-based optimization (SMBO) technique that has proven effective in finding optimal hyperparameter configurations for various machine learning algorithms. TPE is based on the idea of modeling the relationship between hyperparameter values and the associated objective function (i.e., weighted F1 score) using two probabilistic models: one for the promising hyperparameter configurations $(l(x))$ and another for the unpromising configurations $(g(x))$. These models are updated after each iteration of the optimization process based on the observed objective function values.
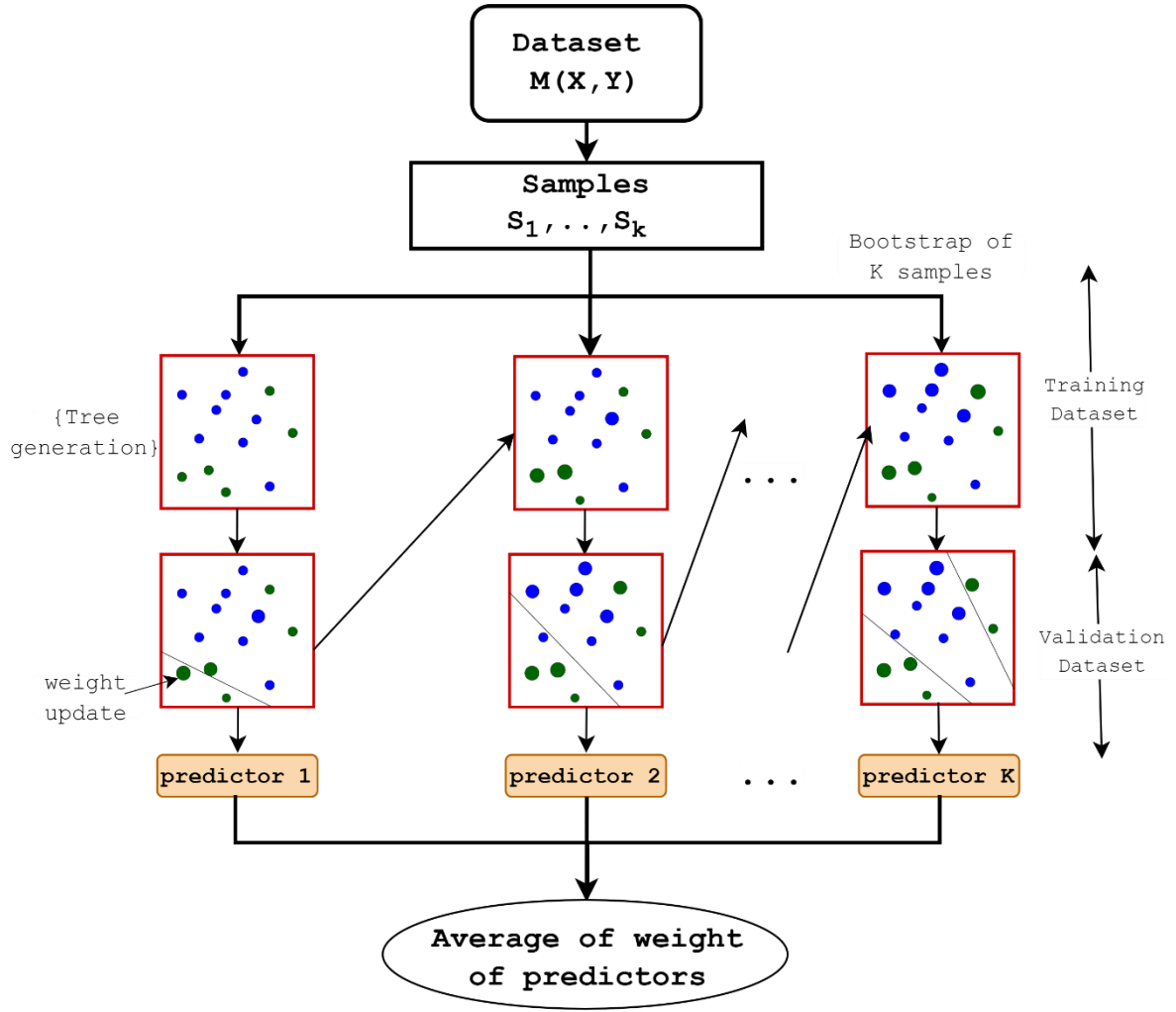
Figure 3. 16: A schematic representation of CatBoost algorithm

The TPE algorithm proceeds as follows:

1. Initialize the probabilistic models $l(x)$ and $g(x)$ with uniform priors.

2. Draw a new set of hyperparameter values x' from $l(x)/(l(x) + g(x))$, which favors sampling from the more promising regions of the hyperparameter space.

3. Evaluate the objective function $f(x')$ with the sampled hyperparameters x'.

4. Update the probabilistic models $l(x)$ and $g(x)$ based on the observed objective function value $f(x')$:

   - If $f(x')$ is among the best observations so far, increase the density of $l(x)$ at x'.

- Otherwise, increase the density of $g(x)$ at x'.

5. Repeat steps 2-4 for a specified number of iterations or until convergence.

The advantages of TPE include efficient exploration of the hyperparameter space, flexibility in handling various hyperparameter types, and the ability to parallelize the optimization process. The hyperparameters optimized and the search space are denoted in Table 3.7.

Table 3. 7: Hyperparameters of EL algorithms

Ensemble algorithms	Hyperparameters	Type	Range
AdaBoost	Number of estimators	Integer	[1, 500]
	Maximum depth	Integer	[1, 50]
	Learning rate	Float	[0.1, 3]
	Boosting algorithm	Categorical	[SAMME, SAMME.R]
RF	Maximum depth	Integer	[1, 50]
	Minimum samples leaf	Integer	[2, 20]
	Maximum features	Categorical	[auto, sqrt, log2]
	Minimum samples split	Integer	[1, 50]
	Number of estimators	Integer	[1, 500]
XGBoost	Learning rate	Float	[0.1, 3]
	Number of estimators	Integer	[1, 500]
	Maximum depth	Integer	[1, 50]
	Subsample	Float	[0.1, 3]
	Column sampling by tree	Float	[0.1, 1]
LightGBM	Learning rate	Float	[0.1, 3]
	Maximum depth	Integer	[1, 50]
	Number of estimators	Integer	[1, 500]
	Subsample	Float	[0.01, 1]
	Column sampling by tree	Float	[0.1, 1]
CatBoost	Learning rate	Float	[0.1, 3]
	Subsample	Float	[0.1, 3]
	Maximum depth	Integer	[1, 50]
	Number of estimators	Integer	[1, 500]
	Column sampling by level	Float	[0.1, 1]

3.6.3. Evaluation of the EL models

To evaluate the performance of the developed models for predicting the causes of water pipe failures, various performance metrics were employed, derived from the confusion matrix for each multi-class classifier. The confusion matrix is a critical tool in assessing the performance of classification models, as it provides a detailed breakdown of the model's predictions across different classes (Taiwo et al., 2024; Yussif et al., 2023).

The confusion matrix for a multi-class classifier is a $K \times K$ matrix, where K is the number of classes. In this study, K is 4. Each element CM_{ij} in the matrix represents the number of instances of class i that were predicted as class j. The diagonal elements CM_{ii} represent the correctly classified instances for each class, while the off-diagonal elements represent misclassifications.

The following metrics are calculated from the confusion matrix and used to evaluate the models' performance:

- **Accuracy:**

$$Accuracy = \frac{\sum_{i=1}^{K} CM_{ii}}{\sum_{i=1}^{K} \sum_{j=1}^{K} CM_{ij}} \tag{3.43}$$

- **Macro Precision:**

$$Macro\ Precision = \frac{1}{K} \sum_{k=1}^{K} Precision_k \tag{3.44}$$

$$\text{where } Precision_k = \frac{CM_{kk}}{\sum_{i=1}^{K} CM_{ik}} \tag{3.45}$$

- **Macro Recall:**

$$Macro\ Precision = \frac{1}{K} \sum_{k=1}^{K} Recall_k \tag{3.46}$$

$$\text{where } Recall_k = \frac{CM_{kk}}{\sum_{i=1}^{K} CM_{kj}} \tag{3.47}$$

- **Macro F1 Score:**

$$Macro\ F1\ Score\ =\ 2\ \times\ \frac{Macro\ Precision\ \times Macro\ Recall}{Macro\ Precision + Macro\ Recall} \tag{3.48}$$

- **Weighted Precision:**

$$Weighted\ Precision =\ \frac{\sum_{k=1}^{K}(Precision_k \times \sum_{j=1}^{K} CM_{kj})}{\sum_{i=1}^{K}\sum_{j=1}^{K} CM_{ij}} \tag{3.49}$$

- **Weighted Recall:**

$$Weighted\ Recall =\ \frac{\sum_{k=1}^{K}(Recall_k \times \sum_{j=1}^{K} CM_{kj})}{\sum_{i=1}^{K}\sum_{j=1}^{K} CM_{ij}} \tag{3.50}$$

- **Weighted F1 Score:**

$$Weighted\ F1\ Score\ =\ 2\ \times\ \frac{Weighted\ Precision\ \times Weighted\ Recall}{Weighted\ Precision + Weighted\ Recall} \tag{3.51}$$

- **Average Area Under the Curve (Avg AUC):** The average of the area under the receiver operating characteristic (ROC) curve for each class, calculated as the mean of the individual AUC scores for each class.

It should be noted that macro metrics, such as macro precision, recall, and F1 score, calculate the average of the respective metric for each class without considering class imbalance, treating each class equally. Weighted metrics, on the other hand, account for class imbalance by weighting the metric for each class by the number of actual instances in that class, providing a more representative measure of the model's performance across all classes. Apart from these 8 metrics, the training time for each model was also included in the performance evaluation.

3.6.4.   Model selection and interpretability

Selecting the most efficient EL model for deployment is crucial for application success. To achieve this, the Copeland algorithm was utilized to conduct a pairwise comparison of the five

EL models under consideration. This comparison was based on eight evaluation metrics outlined in Section 3.6.3, along with considerations of training time, to determine the optimal model. The specifics of the Copeland algorithm, including its methodology and implementation details, are discussed in Section 3.5.4. Once the best model was identified, its decision-making process was interpreted using the SHAP framework. This framework provides an in-depth look at the contribution of each feature to the predictive outcomes, offering valuable insights into the internal workings of the model. Detailed information about the SHAP framework, including its theoretical underpinnings, is available in Section 3.4.3. This interpretability is essential for understanding the driving factors behind the model's predictions, thereby enhancing trust and transparency in its deployment.

3.7. SUMMARY

This chapter presents in detail the adopted methodologies to achieve the objectives of this research. Firstly, a scientometric and systematic review of the literature is conducted to identify water pipe failure factors and failure modes. The second objective is achieved using PLS-SEM algorithm to model, rank, and investigate the relationship between the failure factors and failure modes of water pipes. Different hypotheses are proposed and validated using statistical tests. The first task of the third objective is achieved by fussing logistic regression with genetic algorithms to develop an optimized model to predict the failure probability of water pipes. The contribution of each feature to the prediction is explained using the SHAP framework. The second task of the third objective is realized by fussing DL models with BO to build optimized models for predicting the probability of leaks and bursts. Subsequently, the fourth objective is achieved by utilizing EL algorithms to develop models to predict the causes of water pipe failure. These models are optimized using TPE.

Chapter 4

DATA COLLECTION[3]

4.1. INTRODUCTION

This chapter describes the data collection and processing methodology employed in this study. Four distinct data sources were utilized to achieve the research objectives. A questionnaire survey was conducted among global water utility professionals to validate and rank the identified failure factors and investigate their relationships with failure modes, with the complete survey instrument provided in Appendix B. Historical water network data from the Hong Kong Water Distribution Network (HK WDN) containing operational records and pipe characteristics was collected for developing and validating failure prediction models (POF, POL, POB). Climate data from the Hong Kong Observatory encompassing temperature, rainfall, and humidity parameters was obtained to account for environmental factors in failure prediction. Traffic data from the Hong Kong Transportation Department was gathered to

---

[3] This chapter is largely based upon:

Taiwo, R., Zayed, T. & Ben Seghier, M. E. A. (2024). " Integrated intelligent models for predicting water pipe failure probability". Alexandria Engineering Journal, 86, 243-257, https://doi.org/10.1016/j.aej.2023.11.047

Taiwo, R., Yussif, A., Ben Seghier, M. E. A., & Zayed, T. (2024). "Explainable Ensemble Models for Predicting Wall Thickness Loss of Water Pipes". Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2024.102630

Taiwo, R., Zayed, T. & Adey, B.T. "Explainable deep learning models for predicting water pipe failure." Journal of Environmental Management (IF = 8.7, Q1). (Under review – 1st cycle)

Taiwo, R., Zayed, T. & Adey, B.T. "Interpretable ensemble models for predicting causes of water pipe failure." Reliability Engineering and System Safety (IF = 8.1, Q1). (Under review – 1st cycle)

Taiwo, R., Zayed, T. Elshaboury, N., & Abdelkader, E. M. "Promoting Sustainable Water Distribution Networks: Modeling of Water Pipe Failure Factors and Modes." Cleaner Engineering and Technology (IF = 5.3, Q1). (Under review – 2nd cycle)

incorporate external loading effects. The integration of these diverse data sources enables comprehensive analysis of water pipe failure mechanisms while ensuring robust model development and validation. The detailed data collection for each of the objectives is explained in the following sections.

4.2. DATA COLLECTION FOR PLS-SEM MODEL

4.2.1. Questionnaire design

A closed-ended questionnaire was developed and administered to fulfill the second objective of this study. The survey was developed using the Qualtrics platform, and it was distributed online to experts working in the domain of WDN. One advantage of a closed-ended questionnaire is that it allows for structured responses and is time-efficient (Adabre et al., 2021). The questionnaire contains three sections. The first section collects the demographic details of the respondents. Moving on to the subsequent section, respondents are presented with a series of carefully curated failure factors that have been identified as potentially influential in the occurrence of water pipe failures. Using a 5-point Likert scale, participants are requested to rate the perceived influence of these factors on water pipes. The scale ranges from 1, indicating very low influence, to 5, denoting very high influence. By employing this structured rating system, the participants' subjective evaluations of the various failure factors can be gauged effectively, providing valuable insights into their perceived significance. In the final section of the questionnaire, participants are tasked with assessing the severity of different water pipe failure modes. This section aims to ascertain the participant's perception of the severity associated with these failure modes. Respondents are prompted to assign a rating on a scale of three, with 1 indicating low severity, 2 signifying moderate severity, and 3 denoting high severity. The questionnaire design is shown in Appendix B.

4.2.2.　Sampling and survey respondents

As there is no comprehensive list of professionals working on WDN, a random sampling approach was used to identify the potential respondents for the survey. A random sampling approach has the advantage of ensuring that every individual in the population has an equal chance of being selected (Olanrewaju et al., 2022). The sampling frame for this study consisted of professionals working in the WDN industry, including engineers, managers, and academics. To reach a wider pool of potential respondents, the survey was distributed through professional networks, industry associations, and academic-based media platforms. A total of 320 potential respondents were identified through the sampling approach. After excluding those who did not meet the inclusion criteria, such as those who were not directly involved in the management of WDN, a total of 210 respondents were invited to participate in the survey. Hitherto, out of the invited respondents, 160 completed the survey within 10 weeks, resulting in a response rate of 76%. The high response rate may indicate that the survey was well-received and the participants found that understanding the relationship between water pipe failure factors is necessary and important.

The adequacy of the sample size in this study, consisting of 160 respondents, aligns with the recommended requirements for conducting an SEM analysis (Wolf et al., 2013). Ott & Longnecker (2016) and Wolf et al. (2013) emphasized that a minimum of 30 responses is necessary for SEM analysis, a criterion fulfilled by the sample size in this study. As a general guideline (Hair et al., 2017), the recommended sample size is typically estimated to be ten times greater than the maximum value derived from two criteria: 1) the highest number of connections between an observed variable and a latent construct, and 2) the highest number of connections between a latent construct and other latent constructs. In this study, the maximum number of connections between an observed variable and a latent construct is 9 (i.e., operation-related factors). Additionally, there are six latent constructs, each potentially establishing a

maximum of five relationships with other constructs. Based on these criteria, the estimated required sample size would be 90 respondents. However, it is worth noting that the actual sample size used in this study was 160 respondents, surpassing the recommended threshold. This larger sample size ensures a more comprehensive exploration of the relationships between observed variables and latent constructs, enhancing the reliability and generalizability of the study findings. Therefore, the sample size of 160 respondents is deemed adequate for conducting the SEM analysis, demonstrating a rigorous approach to data collection and analysis in this study.

4.2.3.   **Profile of the respondents**

Figure 4.1-4.5 elucidates the demographic characteristics of the respondents. It indicates that about half of the respondents were academics (51.3%), while the remaining half are working in the industry related to WDN as engineers (17.76%), government officials (13.82%), site supervisors (9.21%), consultant (5.26%), and manager (2.63%). Most respondents (42.48%) had 11-15 years of experience, 19.61% had 16-20 years of experience, and only 11.76% had less than 5 years of experience. Regarding their degree qualification, 37.25, 31.37, and 22.22% had Doctorate, Master, and Bachelor's degrees, respectively. Further, the majority of the respondents (63.82%) had a major in Civil Engineering, while 11.84 and 9.21% had a background in Project Management and Environmental Engineering, respectively. As shown in Figure 4.5, the respondents were from different countries across the globe, with 29.45, 21.92, and 19.86% from Hong Kong, the USA, and Canada, respectively. Based on this demographic information, it can be concluded that the respondents had the necessary knowledge, experience, and qualifications to contribute to the objectives of this study. The diverse background of the respondents from different countries adds to the robustness and generalizability of the study's findings.

Current position

Position	Percentage (%)
Government official	13.82
Site supervisor	9.21
Engineer	17.76
Manager	2.63
Consultant	5.26
Academic	51.32

Percentage (%)
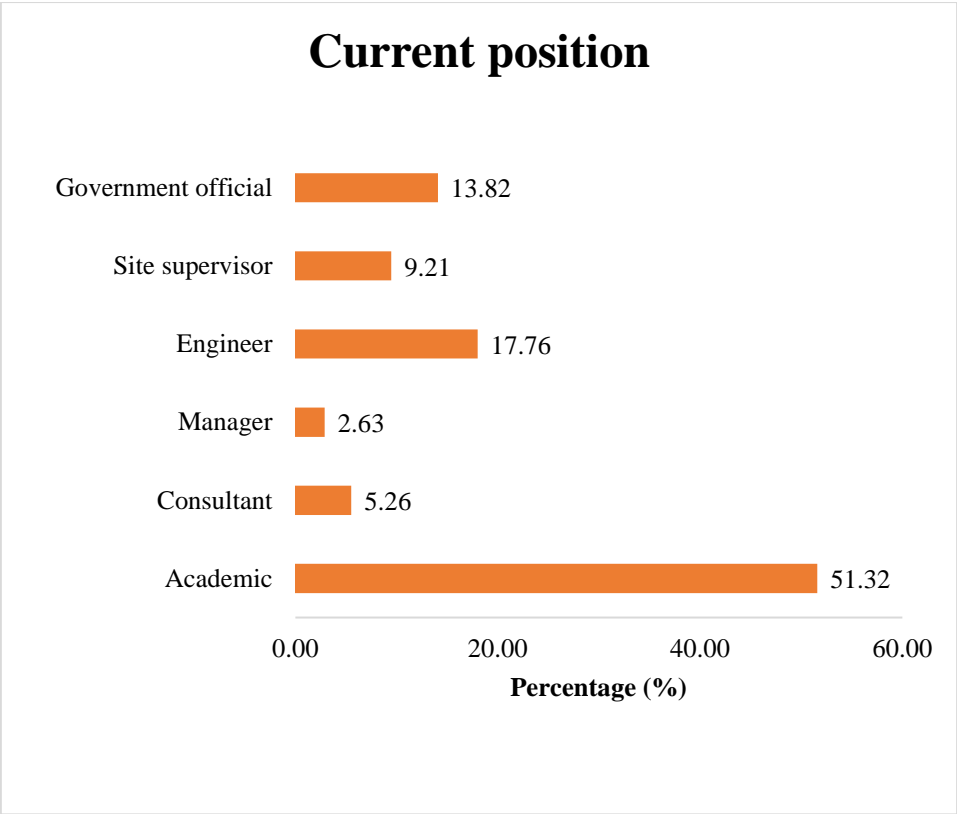
Figure 4. 1: Current position of the respondents

Experience

Experience	Percentage (%)
<5 years	11.76
5-10 years	16.34
11-15 years	42.48
16-20 years	19.61
>20 years	9.80

Percentage (%)
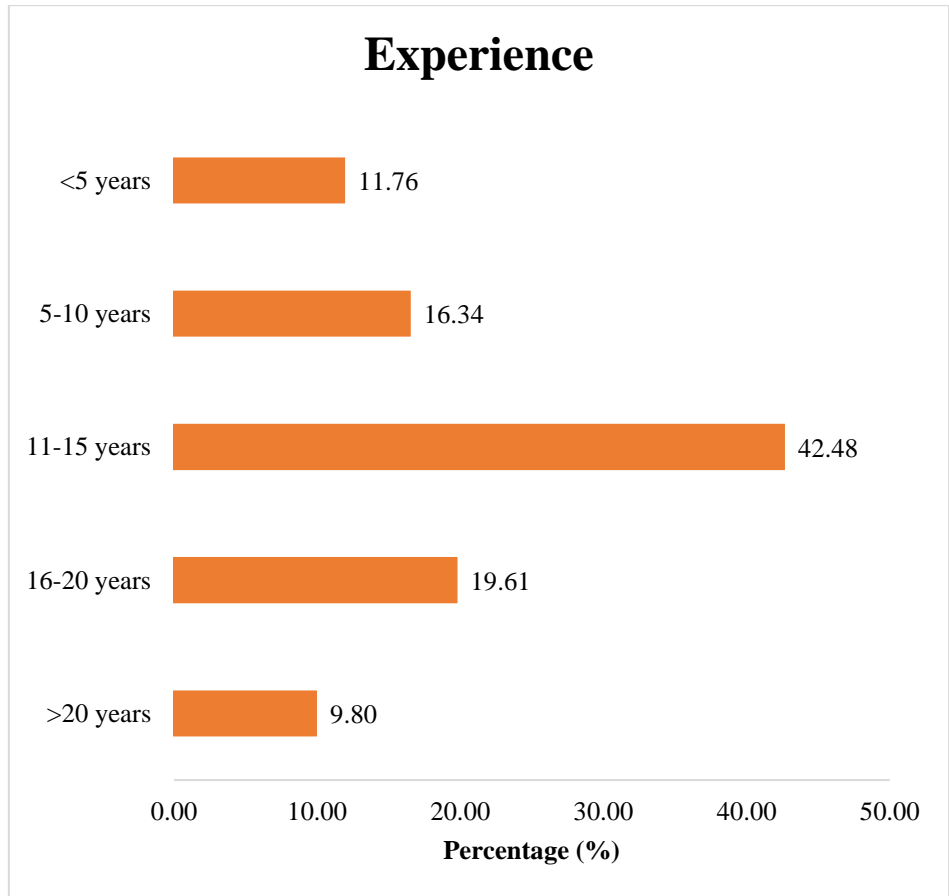
Figure 4. 2: Experience of the respondents

118
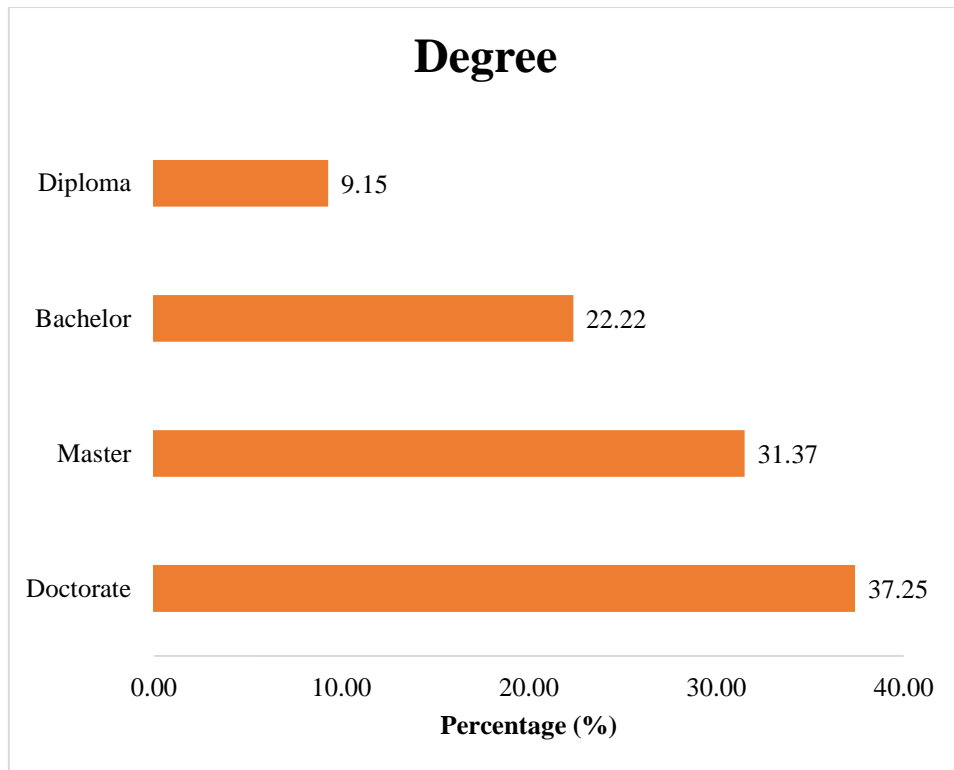
Figure 4. 3: Degree of the respondents
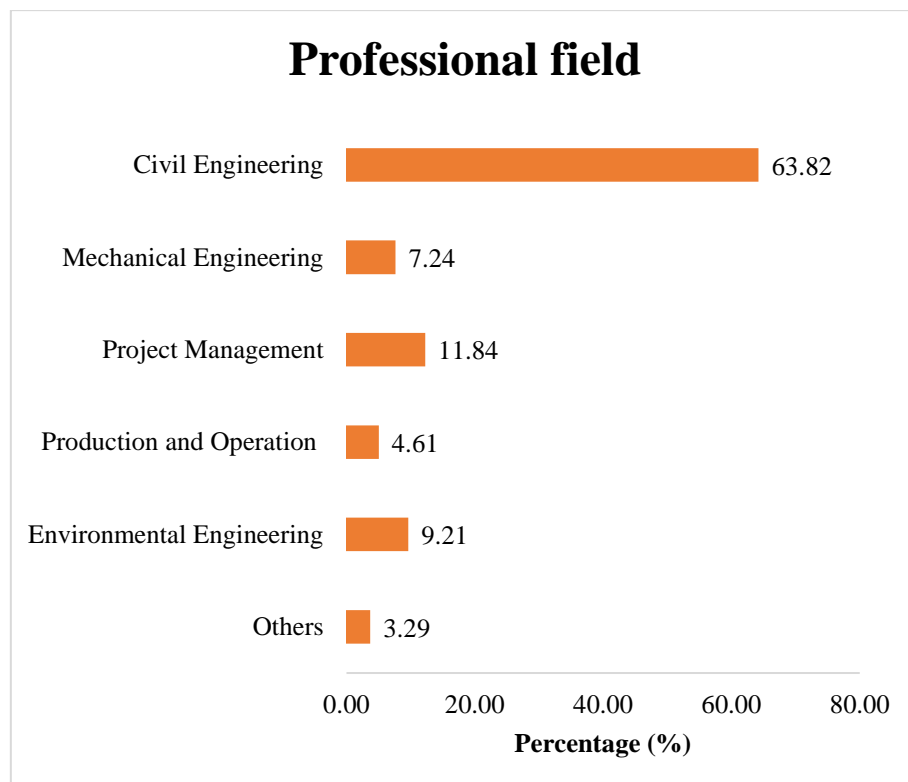
Figure 4. 4: Professional field of the respondents
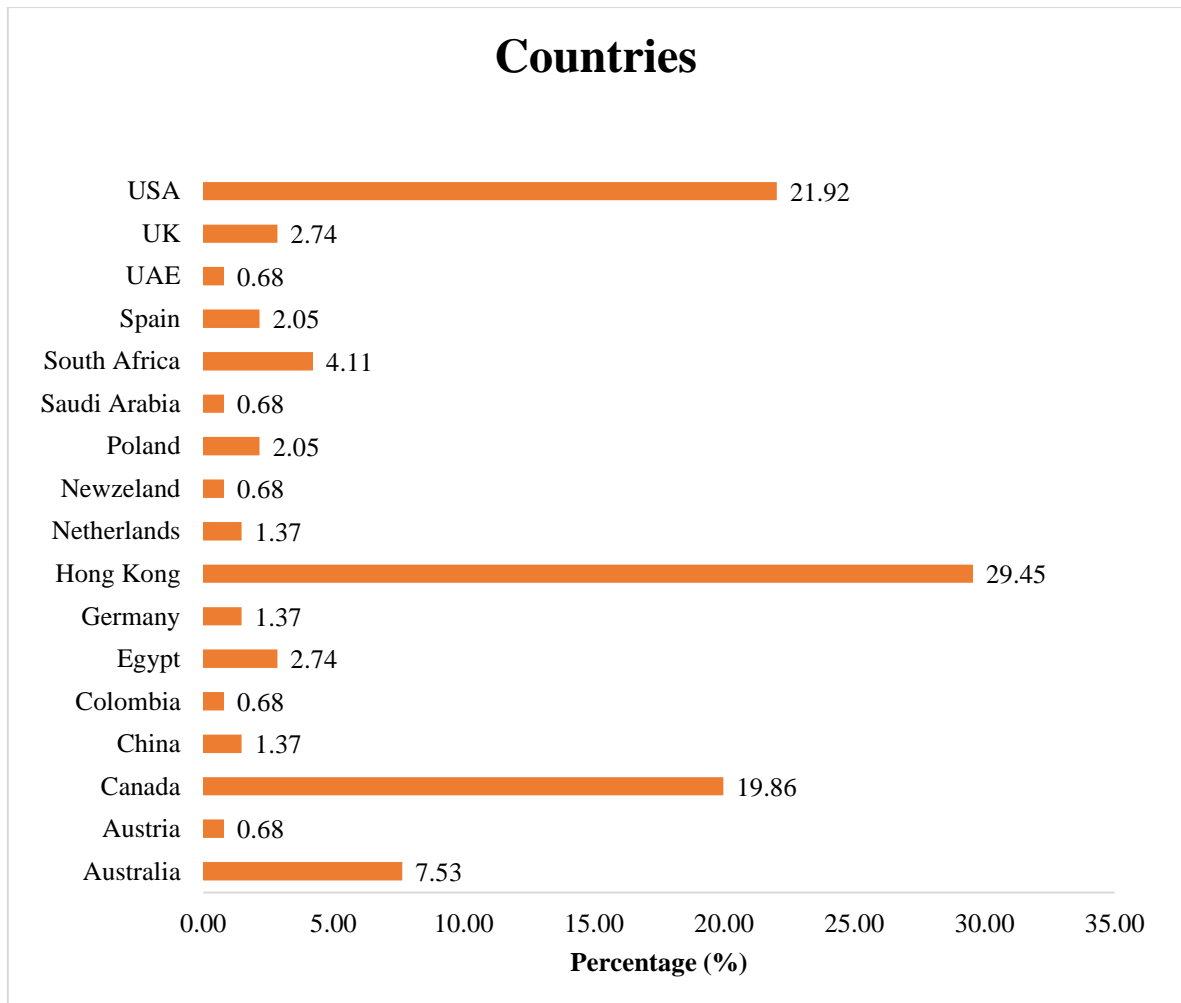
Figure 4. 5: Country of the respondents

4.2.4. **Common method bias**

In order to affirm that the data used in this study is not influenced by common method bias (CMB), several steps were taken. Firstly, the data collection process was designed to be anonymous to minimize the social desirability bias. Secondly, the data was collected at different time intervals to reduce the potential for same-source bias. Thirdly, Harman's single-factor test was conducted using SPSS software to examine the presence of CMB. The result of the CMB test is presented in Table 4.1 and showed that the first factor explained only 28.08% of the total variance, which is less than the recommended threshold (i.e., 50%) (Hair et al., 2013). This suggests that CMB is not a major concern in this research.

Table 4. 1: Result of common method bias

Extraction sums of squared loadings		
Total	% of variance	Cumulative %
9.829	28.08	28.08

4.3. DATA COLLECTION FOR THE POF MODEL

4.3.1. Case study data

The data used in this section of the study are collected from the Water Supply Department in Hong Kong, which is responsible for the management of Hong Kong's WDN. The WDN provides both fresh and saltwater to about 7.41 million individuals (Water Supplies Department, 2021). The HK WSD provided two GIS files. The first one consists of the network, while the second one shows the pipes that have experienced failure from 2010 to 2020. After matching the two files together, the data comprise 1,089,232 pipes, with only 51,568 of them as failed pipes. This means only 4.73% of the pipes have experienced failure in the past.

The WDN is composed of over 8300 km of pipes, with approximately 80% of it carrying freshwater, and the others are responsible for saltwater. In order to have a comprehensive understanding of factors affecting the failure of water pipes, the data provided by the HK WSD was supplemented by data available from different open sources. One of the gaps in previous studies is the lack of data aggregation from different sources to produce robust models. This gap is filled by aggregating the WSD data with the climatic and traffic data provided by HK Observatory and Transportation Department. Although the supplementary data are available online, it requires enormous efforts to combine them together since they are available in different formats. After the data aggregation, the data consisted of 13 factors, divided into pipe-related, environment-related, and operational-related factors. The pipe-related factors include pipe length, diameter, material, and age. The environment-related factors are soil corrosivity, road type, land use, temperature, precipitation, and annual average daily traffic (AADT), while

the operation-related factors include pressure and availability of cathodic protection. The description of the data used in developing the predictive models is presented in Table 4.2, together with statistics of the numerical data. As can be inferred from Table 4.2, the categorical factors take a value of 0 when the factor is not present in an observation and 1 otherwise.

The network comprises different materials, including ductile iron (DI), cast iron (CI), plastic (Polyvinyl chloride and polyethylene pipes), steel, galvanized iron (lined and unlined), and asbestos cement (AC) pipes. Figure 3 shows the distribution of the pipes in relation to the water types in HK WDN. As shown in Figure 4, GI (26.43%), DI (25.6%), and plastic (23.06%) pipes dominate the freshwater pipes, while DI (39.36%), plastic (33.98%), and CI (10.53%) are mostly used to distribute saltwater to the consumers.

4.3.2. **Data pre-processing**

Data pre-processing is an essential step in ML model development. It is known that an ML model is as good as the data used in its development. Quality data has the potential to produce accurate predictive models. Hence, outlier detection, data imputation, normalization, and standardization were conducted to tackle the problem of data redundancy, noise, and heterogeneity. The pre-processing starts by detecting the outliers in the data. The detection was done using data visualizations (i.e., box plots and scatter plots) and descriptive statistics, and the outliers were eliminated. For instance, 99% of the pipes have an age between 0-70 years, and only 1% of the pipes have an age between 115-120 years. These few data instances are outliers, which were eliminated.

Subsequently, the missing data for the numerical factors was filled using the mean value, while the mode was used to complete the categorical data.

Table 4. 2: Description of the data used for model development

Factor	Source	Description	Mean	Std	Min	Max
Length	HK WSD	This shows how long the pipe is, and it is measured in meters.	8.04	20.25	1.00	200.00
Diameter	HK WSD	This describes the outside diameter of the pipes, and it is measured in millimeters.	132.47	171.11	20.00	3000.00
Material	HK WSD	This describes the material type of each pipe.	-	-	0	1
Age	HK WSD	This refers to the number of years since the installation	22.45	16.20	0	70
Soil corrosivity	HK WSD	This typifies the aggressiveness of the soil, and it can be noncorrosive, moderately, or highly corrosive.	-	-	0	1
Road type	HK WSD	This shows the type of road that is above the buried pipes, and it can be footway, carriageway, or other locations	-	-	0	1
Land use	HK WSD	This describes the usage of land in the vicinity of the buried pipes, and it can either be urban or rural.	-	-	0	1
Temperature	HK OBV	This refers to the mean temperature (^{o}C) at various pipe locations	24.26	6.95	2.50	30.5
Precipitation	HK OBV	This is the accumulated amount of rainfall at various pipe locations, and it is measured in millimeters.	10.36	3.33	2.34	13.1
AADT	HK TRD	This describes the traffic above each buried pipe.	20337.97	24346.09	1.00	179400.00
Pressure	HK WSD	This refers to the water pressure, and it is measured in bars	6.22	2.421	0.05	21.069
Water type	HK WSD	This shows the type of water carried by the pipe, and it can either be fresh or saltwater	-	-	0	1
CP	HK WSD	This describes if cathodic protection is applied to a pipe or not.	-	-	0	1

HK OBV refers to Hong Kong Observatory, and HK TRD refers to Hong Kong Transportation Department
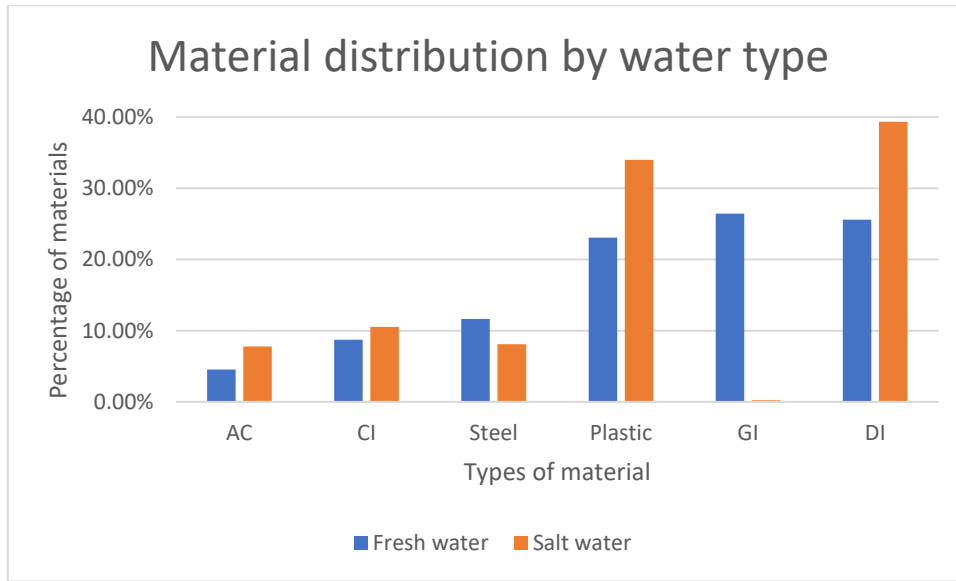
Figure 4. 6: Material distribution by water types in Hong Kong WDN

There are two methods of coding categorical data in ML modeling: one-hot encoding (i.e., dummy coding) and label coding. The former refers to creating binary variables [0,1] for each of the categorical features, meaning that 0 will be applied when the feature is present in a sample, and 1 will be adopted otherwise. The latter refers to the process of assigning an arbitrary number to each unique attribute of a categorical feature. Based on previous research and a series of experiments, one-hot encoding is used for coding the categorical features in this study (Rifaai et al., 2022; Robles-Velasco et al., 2021). The output of the models developed in this study is the status of an individual pipe, which is either failed or not (i.e., intact). The output variable is binary-coded in the sense that 1 represents a failed pipe, and 0 depicts an intact pipe. It should be stressed that the probability of belonging to either of the two classes (i.e., failed and non-failed classes) is taken as the probability of failure. Hence, a pipe is considered failed when its predicted probability of failure is greater than 0.5. A common issue of ML modeling of WDNs is the nature of the dataset – it is usually imbalanced. This is because most of the pipes in the networks have not yet experienced failure. However, an ML model will be biased towards the majority class if the model is trained on imbalanced data, thereby reducing the

accuracy and precision of such a model. As previously mentioned, the data is highly unbalanced, as about 5% of the data belongs to the failure class; hence, an under-sampling technique was adopted to rebalance the training data. This technique has been used in a previous related study (Robles-velasco et al., 2020).

After the data-cleaning process described above, 871,802 pipes were found appropriate for the model development. Since classifiers are to be developed in this study and the numerical factors occur on different scales (i.e., value ranges), data scaling was conducted to unify the data. Without applying data scaling, the learning algorithm will be biased towards numerical factors with higher magnitude, hence, producing an inaccurate model. Therefore, data scaling is an important step in ML modeling, as it allows the learning algorithm to understand the data well at a fast rate.

Two data scaling methods were adopted – normalization and standardization – based on their efficiency in previous studies (Almheiri et al., 2021; Robles-velasco et al., 2020), and their results were compared. The data were normalized and standardized using Equations 4.1 and 4.2, respectively. Normalization ensures that the data lies within a certain range based on the maximum and minimum values of a feature. In our case, the data ranges between 0 and 1. On the other hand, standardization ensures that the standardized feature has a standard deviation of 1 and a mean of 0. It should be noted that only the numerical features were standardized. Standardizing the one-hot encoding features means that a distribution is assigned to them, which is inaccurate.

$$x_{norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{4.1}$$

$$x_{stand} = \frac{x_i - x_{mean}}{x_{std}} \tag{4.2}$$

where x_i, x_{norm} and x_{stand} refer to the unscaled, normalized, and standardized value of a data instance and x_{min}, x_{max}, x_{mean} and x_{std} represent the minimum, maximum, mean, and standard deviation of a feature.

4.4.   DATA COLLECTION FOR THE POL AND POB MODELS

4.4.1.   Case study data

An updated data from HK WSD was collected, which shows the exact type of failure that each pipe has experienced in the past. The data provision from the WSD comprised two GIS files, one detailing the configuration of the water network and the other cataloging incidences of leaks and bursts across the network. Upon integration of these data, the resulting compilation included records for 1,089,232 pipes. Of these pipes, 37,767 were identified to have suffered from leaks, while bursts were noted in 1,552 cases. This indicates that leakages and bursts affected a mere 3.47% and 0.142% of the network, respectively. The WDN spans over 8,300 kilometers and is the primary water supply infrastructure for a population exceeding 7.41 million (Water Supplies Department HKSAR, 2021).

Additional data were sourced from various open databases to form a more holistic view of the variables influencing pipe failures. This supplementary data encompasses climatic variables obtained from the HK Observatory and traffic information sourced from the HK Transport Department. Subsequent to the dataset aggregation, a total of 13 variables were identified and categorized into three groups: those pertaining to the pipes themselves, environmental conditions, and operational characteristics. The group of pipe-specific variables encompassed attributes such as length, diameter, material, and service age. Environmental variables included the corrosiveness of the soil, the type of roadway above, the surrounding land's utilization, meteorological factors such as temperature, precipitation, and humidity, and the annual average daily traffic (AADT) values. The operational variables consisted of the water pressure within

the pipes, water type, and whether CP was implemented or not. The dataset's descriptive statistics are presented in Table 4.3. A binary system was employed for categorical variables, assigning a '1' to denote the presence of a characteristic in a given data point and a '0' to signify its absence (i.e., one-hot encoding).

Table 4. 3: Descriptive statistics of the data

Factor	Unit	Mean	Std	Min	Max
Length	Metre (m)	7.43	17.43	1.00	200.00
Diameter	Millimeters (mm)	131.21	167.44	20.00	3000.00
Material	AC, CI, DI, PE	-	-	0	1
Age	Years	24.17	20.29	0	70.00
Soil corrosivity	Non-corrosive, Mildly corrosive, Highly corrosive	-	-	0	1
Road type	Footway, Carriageway, Other locations	-	-	0	1
Land use	Urban, Rural.	-	-	0	1
Temperature	ºC	24.26	6.96	2.30	30.20
Precipitation	Millimeters (mm)	10.56	9.07	1.02	792.00
Traffic	AADT	11427.88	15970.03	1.00	179400.00
Pressure	Bars	6.24	2.42	0.53	24.60
Water type	Freshwater, Saltwater	-	-	0	1
CP	-	-	-	0	1

4.4.2. Data pre-processing

Pre-processing of data is a critical stage in the development of DL models, as the accuracy of predictions is greatly influenced by the quality of the input data. To address issues of redundancy, noise, and heterogeneity in the data, several pre-processing techniques were employed. Initial steps included identifying and removing outliers, utilizing methods such as box and scatter plots, and a thorough examination of descriptive statistics. For instance, pipe

data indicating an age of 115 years, which represented a scant proportion of less than 1%, were deemed outliers and thus removed. For the treatment of missing entries, numerical data points were substituted with the mean of their respective feature, while the most frequent values, or modes, were used for categorical data. The rationale behind these methods is to maintain the integrity of the dataset's distribution. In the scaling phase, both normalization and standardization procedures were executed, drawing on their demonstrated effectiveness in prior research (Almheiri et al., 2021; Robles-velasco et al., 2020). The mathematical expressions for these scaling methods are presented in Equations 4.1 and 4.2.

4.5. DATA COLLECTION FOR THE COF MODEL

4.5.1. Case study data

The Water Supplies Department (WSD) provided updated data on the Hong Kong WDN, which includes over 1 million assets. Of these, 64,076 assets have documented records of failures, including detailed causes. The dataset from Hong Kong WSD incorporates intrinsic pipe data and operational data. To further enrich this dataset, climatic data from the Hong Kong Observatory and traffic data from the Hong Kong Transport Department were added. After integrating these diverse sources of data, the final dataset emerged with 13 input variables and 1 output variable detailing the causes of water pipe failures.

After data cleaning, the remaining number of data instances was 62,738. Table 4.4 presents the descriptive statistics of the dataset, including the mean, standard deviation, minimum, and maximum values for each variable. To better understand the dataset, the distribution of pipe materials and causes of water pipe failures were analyzed. Figure 4.7 illustrates the distribution of pipe materials for the failed pipes, while Figure 4.8 depicts the distribution of the causes of water pipe failures. Figure 2 shows lined galvanized iron (GIL) as the most prevalent pipe material, followed by galvanized iron (GI) and unplasticized Polyvinyl Chloride (UPVC).

Figure 3 reveals that corrosion was the leading cause of failures, highlighting its significance, followed by faulty workmanship and external loading.

Table 4. 4: Descriptive statistics of the data

Variable	Unit	Mean	Standard deviation	Minimum	Maximum
Length	Metre (m)	23.25	25.33	1.00	200.00
Diameter	Millimeters (mm)	99.15	98.14	100	2200
Age	Years	30.23	21.45	0.00	99.00
Material	AC, CI, DI, GI, GIL, S, SS, PE, UPVC	-	-	0	1
Soil corrosivity	Highly corrosive, Mildly corrosive, Non corrosive	-	-	0	1
Road type	Footway, Carriageway, Other locations	-	-	0	1
Water type	Saltwater, Freshwater	-	-	0	1
Pressure	Bar	6.40	2.60	0.67	20.11
Landuse	Urban, Rural	-	-	0	1
Traffic	AADT	21935	23573	0.00	147930
Temperature	ºC	22.95	4.99	3.00	30.80
Relative humidity	%	78.09	7.45	51.00	100.00
Precipitation	Millimeters (mm)	171.59	162.98	0.00	893.00
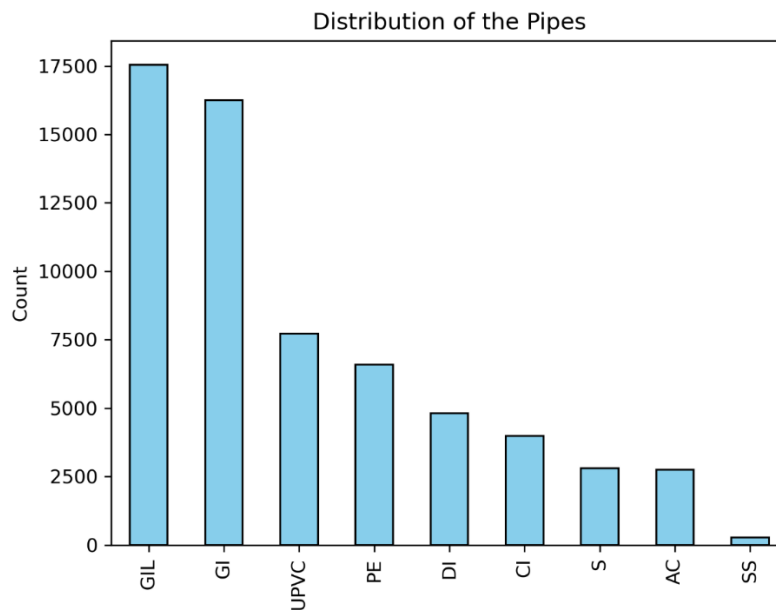
Figure 4. 7: Material distribution for the failed pipes
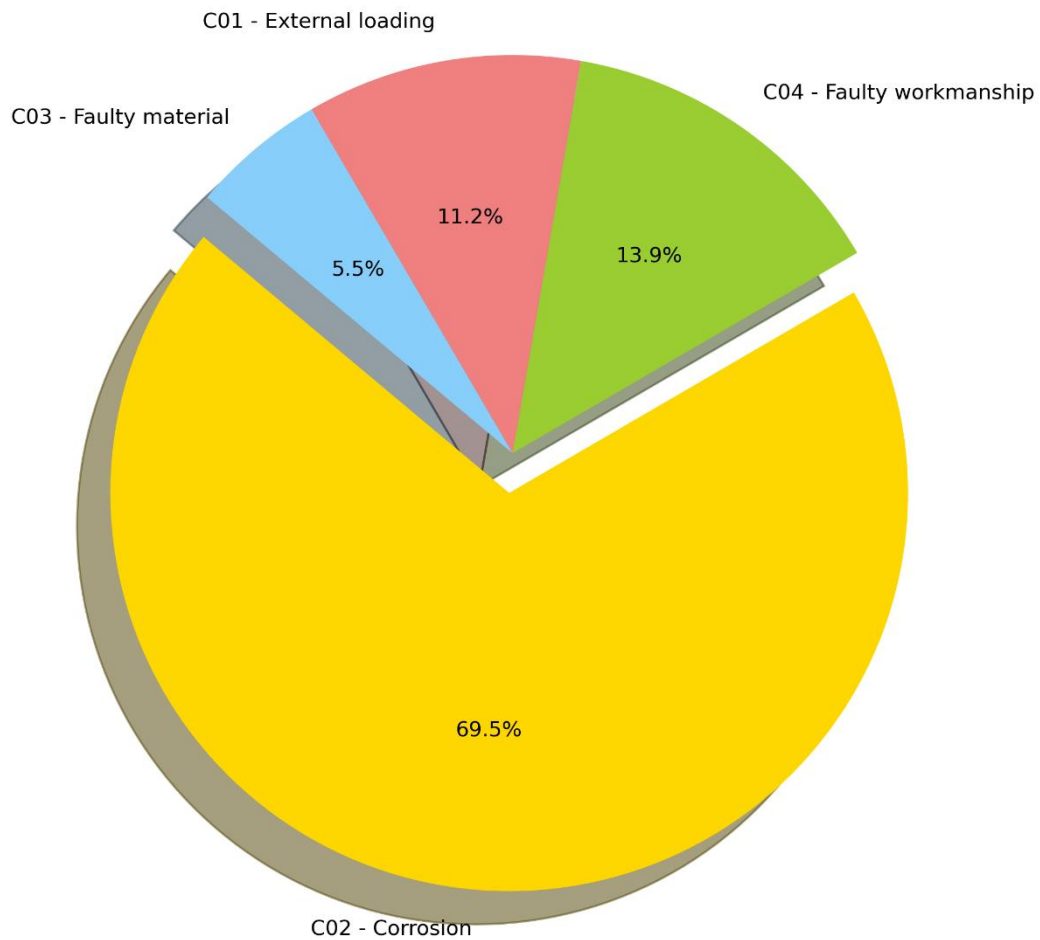
Figure 4. 8: Distribution of water pipe failure causes

4.5.2. Data pre-processing

The input variables consisted of four pipe-related factors: length, diameter, age, and material. The environment-related factors included soil corrosivity, road type, land use, annual average daily traffic (AADT), temperature, relative humidity, and precipitation. The operational-related factors were water type and internal pressure. Data pre-processing was carried out to ensure data quality and suitability for modeling. This involved removing outliers and duplicate records, as well as imputing missing values using the mean for numerical variables and the mode for categorical variables. Based on the outcomes of the previous objectives of this study,

categorical data was pre-processed by creating dummy variables, while numerical data was standardized using Equation 4. 2.

4.6. **SUMMARY**

This chapter outlines the data collection methods utilized to develop models for objectives 2, 3, and 4 of this study. To achieve the second objective, a questionnaire survey was conducted to gather data necessary for developing the PLS-SEM model. The chapter also describes the collection of data used to develop an optimized predictive model that estimates the likelihood of failure in water pipes, addressing the first task of the third objective. Additionally, updated data specifying the type of failure was collected to aid in the development of the Probability of Leak (POL) and Probability of Burst (POB) models, meeting the requirements of objective 3b. Lastly, recent data detailing the causes of water pipe failures in the Hong Kong Water Distribution Network (HK WDN) was gathered and utilized to develop the Causes of Failure (COF) model.

Chapter 5

RESULTS AND DISCUSSION[4]

5.1. INTRODUCTION

This chapter provides the results of the research methodologies developed to fulfill the objectives of the current study. First, the results from the systematic review of water pipe failure factors and failure modes are presented. Subsequently, the results of the PLS-SEM model are highlighted and discussed. The implementation of the POF, POL, and POB models is discussed and elaborated. The results of the predictive models for causes of water pipe failure are also presented to fulfill the fourth objective of this study.

---

[4] This chapter is largely based upon:

Taiwo, R., Zayed, T. & Ben Seghier, M. E. A. (2024). " Integrated intelligent models for predicting water pipe failure probability". Alexandria Engineering Journal, 86, 243-257, https://doi.org/10.1016/j.aej.2023.11.047

Taiwo, R., Yussif, A., Ben Seghier, M. E. A., & Zayed, T. (2024). "Explainable Ensemble Models for Predicting Wall Thickness Loss of Water Pipes". Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2024.102630

Taiwo, R., Zayed, T. & Adey, B.T. "Explainable deep learning models for predicting water pipe failure." Journal of Environmental Management (IF = 8.7, Q1). (Under review – 1st cycle)

Taiwo, R., Zayed, T. & Adey, B.T. "Interpretable ensemble models for predicting causes of water pipe failure." Reliability Engineering and System Safety (IF = 8.1, Q1). (Under review – 1st cycle)

Taiwo, R., Zayed, T. Elshaboury, N., & Abdelkader, E. M. "Promoting Sustainable Water Distribution Networks: Modeling of Water Pipe Failure Factors and Modes." Cleaner Engineering and Technology (IF = 5.3, Q1). (Under review – 2nd cycle)

5.2. RESULTS OF SYSTEMATIC REVIEW OF WATER PIPE FAILURE FACTORS AND FAILURE MODES

A comprehensive literature review of water pipe failure factors and failure modes is conducted and presented in the second chapter of this thesis. 30 failure factors and five failure modes are identified. These factors and failure modes are presented in this section using fault tree logic (FTL).

5.2.1. Fault-tree logic for mapping failure factors of water pipe

This section presents a mapping model for the failure factors of water pipes. The model was developed using the FTL as it can show the link between the failure of a water pipe and its influencing factors. This model can be beneficial to infrastructure managers in focusing on measures or strategies that can be used to mitigate the occurrence of the basic causes of the system failure. As shown in Figure 5.1, the FTL diagram summarises the output of the systematic review presented in chapter two. The interpretation of the symbols used in the FTL diagram is shown in Table 5.1. It could be observed that the FTL diagram consists of several intermediate and basic events and one undeveloped event. Several benefits can be derived from the use of FTL in mapping the failure factors, as highlighted below:

- **Structured Approach**: FTL provides a structured procedure for mapping and analyzing the failure factors in a complex water pipe system. With the FTL model, all the relevant causes of water pipe failure are considered, and a systematic examination of the interdependencies between various failure factors can be established (Kim et al., 2021).

- **Risk Identification**: The use of FTL helps identify the likelihood of failure events and failure modes, allowing decision-makers to prioritize areas for improvement and allocate resources effectively. The results of the analysis can also be used to develop
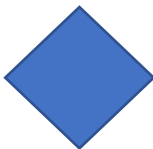
risk management strategies, such as identifying critical components and taking measures to improve their reliability (Lindhe et al., 2009).

- **Improved Decision-Making**: FTL can provide deeper insight into the behavior of a WDN and the factors that contribute to its failures. This improved understanding can help inform decisions related to system design, maintenance, and operation.

The FTL diagram starts with the top event named "failure factors of water pipes," which is directly connected to four intermediate events — pipe-related, soil-related, external-related, and operation-related factors — by the "OR" gate, indicating that any of the intermediate events can cause the occurrence of the top event. For instance, "pipe diameter" — an element of a pipe-related intermediate event and "internal water pressure" — an element of an operation-related event caused the failure of water pipes, as discussed by Hekmati et al. (2020) and Robert et al. (2017), respectively. The first intermediate event – pipe-related factors – consists of another intermediate event named "Protection Efficiency" and 6 basic events that are connected to the intermediate event with the "OR" gate. The same case is applicable to the intermediate event of the operation-related factors. Moreover, the intermediate event of soil-related causes is connected by the "OR" gate to five basic events and another intermediate event named "Resistivity". It should be known that the "Resistivity" event is connected by the "AND" gate to three basic events, typifying that these three events must occur simultaneously to produce the "Resistivity" event (Arriba-Rodriguez et al., 2018). Subsequently, the intermediate event of the external-related factors is connected by the "OR' gate to three intermediate events: climate-related, location-based, and biological and chemical-related factors, indicating that each of the events can cause water pipe failure. For example, "temperature" — an element of a climate-related event and "microbiologically induced corrosion" — an element of a biological and chemical-related event caused the failure of water pipes as indicated in the studies by Trickey et al. (2016) and San et al. (2012), respectively. Similarly, the intermediate events of

"climate-related factors" and "biological and chemical-related factors" are individually connected to three basic events by the "OR gate," while the event of the location-based factors is connected by the "OR gate" to a basic event — traffic — and an undeveloped event, land use. An undeveloped event is used for "land use" because very limited literature has suggested it to be a contributing factor to the failure of water pipes.

Table 5. 1: FTL event symbols, logic gates, and interpretations

Event symbols and logic gates		Interpretation
Basic event	⬤	It represents a component of failure that cannot be developed further.
Intermediate event	▬	It represents a fault event that is between basic and top events. It is developed through logic gates.
Undeveloped event	◆	It represents an event that cannot be developed further due to minimal information available about it.
OR gate	OR	It is used to show that the output event will occur if any of the inputs occur.
AND gate	AND	It is used to show that the output event occurs only when all the inputs exist at the same time.

5.2.2. Fault-tree logic for mapping failure modes of water pipe

Figure 5.2 presents the FTL diagram for mapping the failure modes of water pipes. According to the diagram, the top event, "water pipe failure modes," is connected to five basic events – circumferential cracking, longitudinal cracking, bell splitting, corrosion pitting, and blown-out hole – by the "OR" gate, indicating that any of the basic events stand as an independent failure mode for water pipes.

Figure 5. 1: Fault tree logic (FTL) diagram for failure factors of water pipe
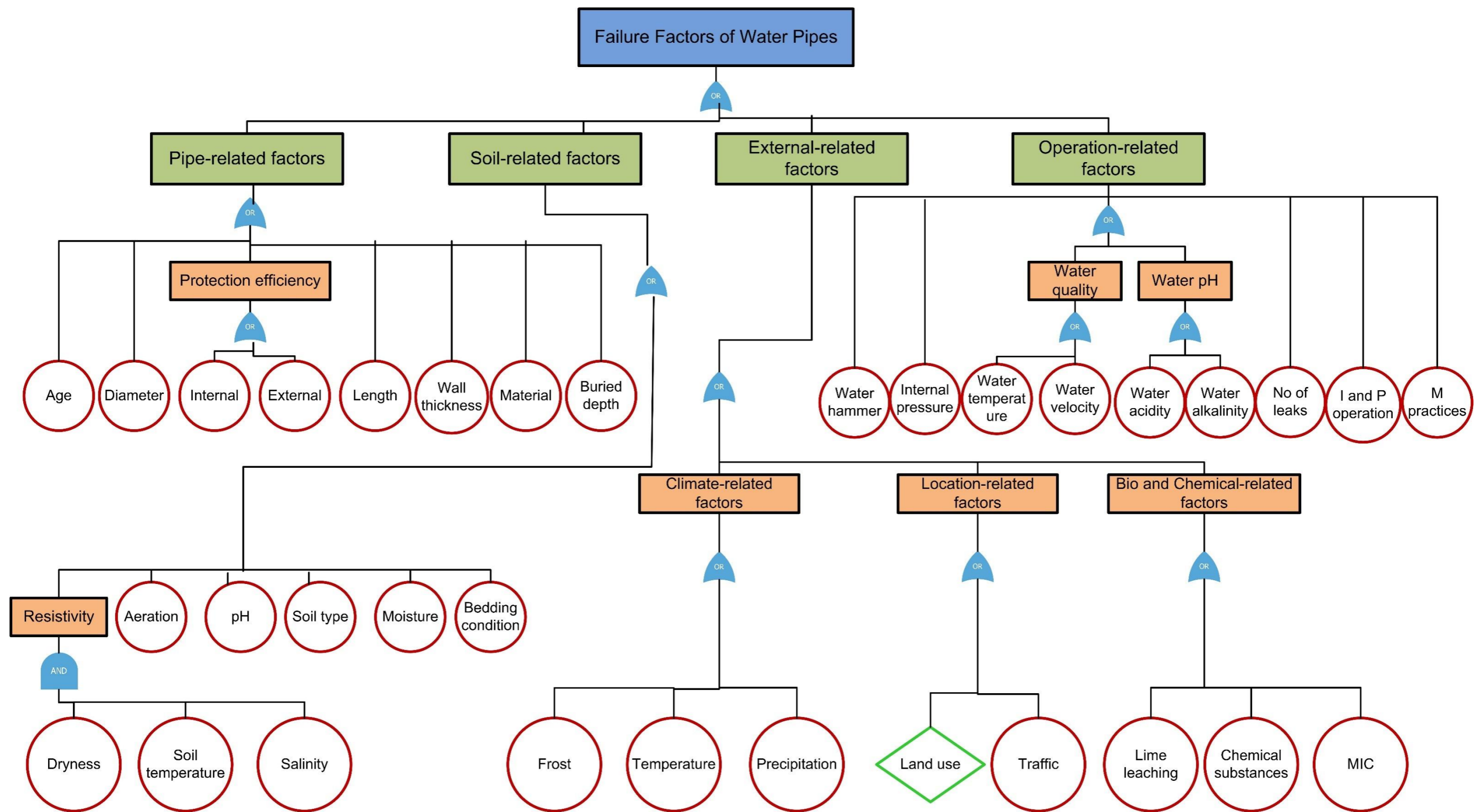
I and P operation = Installation and pump operation; M practices = Maintenance practices; MIC = Microbiologically induced corrosion
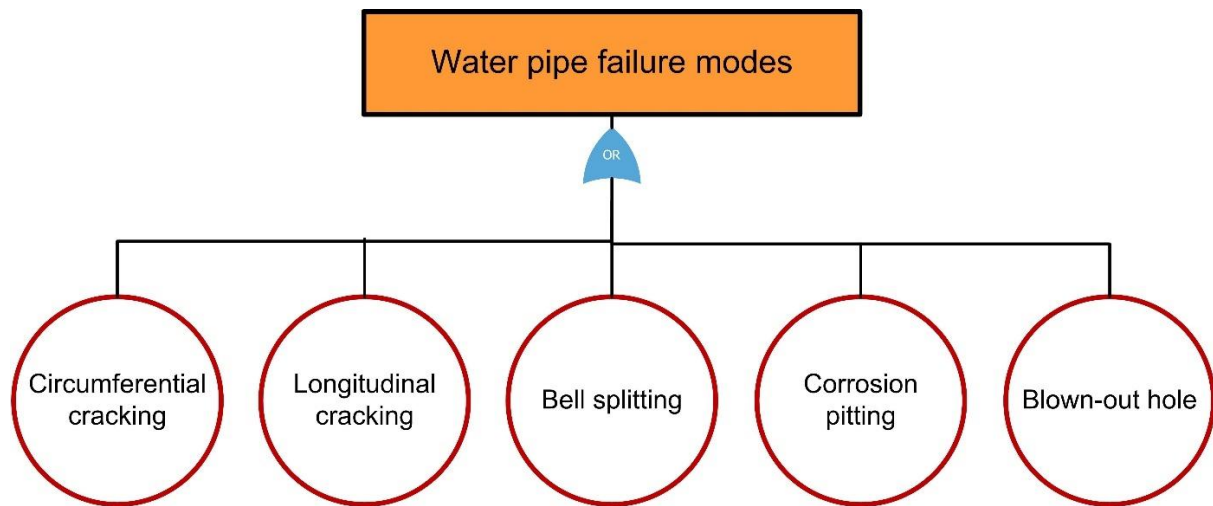
Figure 5. 2: Fault tree logic diagram for water pipe failure modes

5.3. RESULTS OF THE PLS-SEM MODEL

5.3.1. Result of the measurement model

As previously indicated, the measurement model was developed using the PLS-SEM algorithm in SmartPLS software. The results of the model were assessed and validated using construct and convergent validity and discriminant validity, which are presented in the subsequent sections.

5.3.1.1 Construct reliability and convergent validity

(I) Outer loading: Table 5.2 expounds on the result of the construct reliability and convergent validity. The initial measurement model contained some indicators with outer loadings below the 0.5 threshold. While the literature review suggests that these indicators may contribute to pipe failure, they were not as critical as the ones retained in the final model. Removing these less critical indicators enhances the overall model reliability by focusing on the most significant factors influencing pipe failures. Figure 5.3 demonstrates the final measurement model consisting of reliable indicators. The result indicates that 81% (18) of the reliable indicators had a factor loading of 0.7 and above, which is considered highly satisfactory (see Table 5.2).

The remaining three indicators exhibited a factor loading greater than 0.5, which is considered acceptable. By retaining only the most reliable indicators, the model ensures that it accurately captures the most influential factors, leading to more robust and trustworthy results.

(II) Cronbach's alpha (α): As previously indicated, the acceptable threshold for Cronbach's alpha is 0.7. Table 5.2 indicates that all the constructs have a Cronbach's alpha greater than 0.8, which is more than the required threshold. It should be noted some indicators were deleted even though they have an outer loading greater than 0.5 to ensure that all the constructs meet the validity requirement.

(III) Composite reliability (ρ_c): According to Table 3.1, a construct with a composite reliability value of 0.7, 0.8, and 0.9 is regarded as acceptable, satisfactory, and perfect. This assesses the internal consistency of measurements, providing a more accurate measure than traditional Cronbach's alpha by accounting for different indicator loadings. In this connection, all the constructs exhibited a composite reliability value greater than 0.8, indicating that the reliability of the measurement model is deemed satisfactory.

(IV) Average variance extracted (AVE): Table 5.2 states that all the constructs exhibited an AVE value greater than the required threshold of 0.5. This implies that the construct accounts for more variance than measurement error, thereby indicating that the indicators are dependable and valid measures of the construct.

The construct reliability and convergent validity findings establish that the analytical model is both convergent and consistent, providing evidence to support the notion that the indicators used in the model are reliable and valid measures of the constructs they represent. Additionally, the relationships between the constructs in the analytical model have been carefully analyzed, providing further evidence to support the theoretical relationships proposed in the study.

Table 5. 2: Result of the construct reliability and convergent validity

Construct	Code	Outer loading		Cronbach's alpha	Composite reliability	AVE
		Initial	Modified			
Pipe-related factors	PRF 1	0.755	0.780	0.812	0.828	0.572
	PRF2	0.546	deleted			
	PRF3	0.776	0.832			
	PRF4	0.766	0.809			
	PRF5	0.708	0.707			
	PRF6	0.247	deleted			
	PRF7	0.687	0.643			
External-related factors	ERF1	0.846	0.905	0.800	0.841	0.634
	ERF2	0.466	deleted			
	ERF3	0.795	0.816			
	ERF4	0.361	deleted			
	ERF5	0.430	deleted			
	ERF6	0.645	0.584			
	ERF7	0.774	0.851			
	ERF8	0.159	deleted			
Soil-related factors	SRF1	0.583	0.571	0.803	0.862	0.551
	SRF2	0.791	0.815			
	SRF3	0.744	0.746			
	SRF4	0.789	0.811			
	SRF5	0.751	0.733			
	SRF6	0.597	deleted			
Operational-related factors	ORF1	0.400	deleted	0.809	0.817	0.572
	ORF2	0.639	0.676			
	ORF3	0.694	0.721			
	ORF4	0.696	0.748			
	ORF5	0.795	0.777			
	ORF6	0.864	0.861			
	ORF7	-0.024	deleted			
	ORF8	0.548	deleted			
	ORF9	0.578	deleted			
Failure modes	FMD1	0.344	deleted	0.829	0.844	0.745
	FMD2	0.799	0.825			
	FMD3	0.788	0.858			
	FMD4	0.865	0.910			
	FMD5	0.436	deleted			

5.3.1.2. Discriminant validity

The discriminant validity of the model was assessed using three distinct and complementary approaches. The Heterotrait-Monotrait (HTMT) ratio evaluates the correlation between

constructs relative to the correlation within constructs, offering a more stringent assessment than traditional methods.
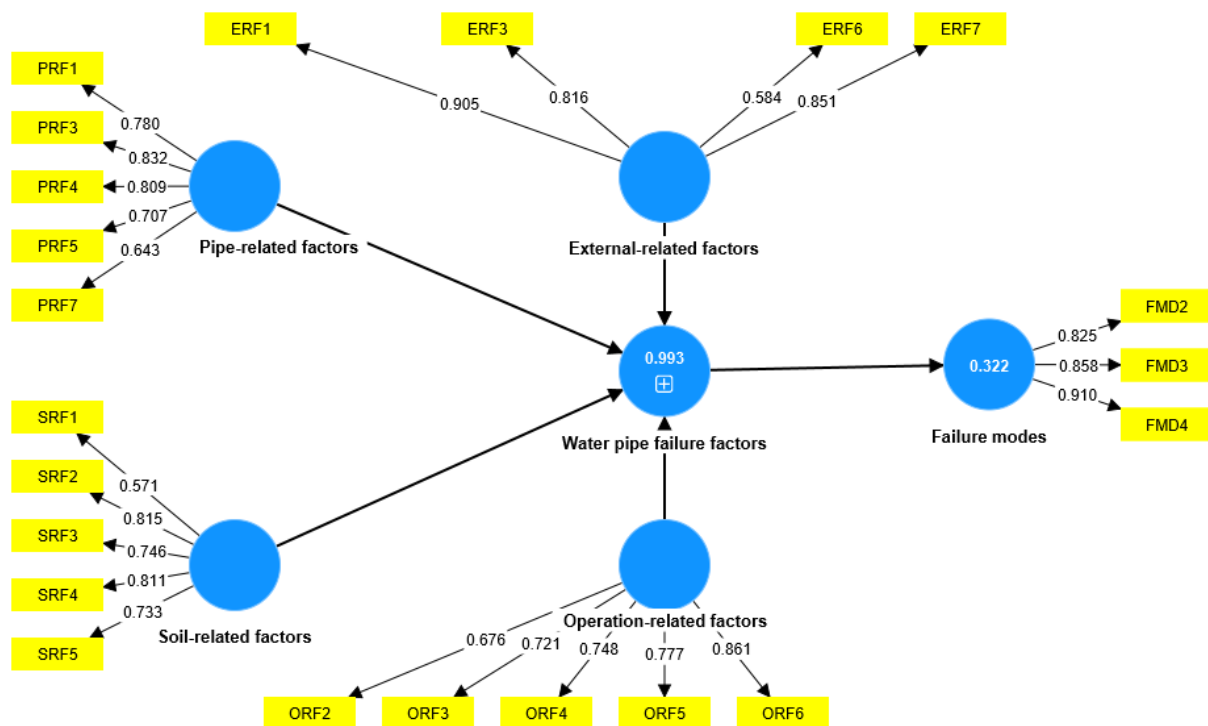
Figure 5. 3**:** Final measurement model

The Fornell-Larcker criterion compares the square root of AVE values with latent variable correlations, specifically examining whether a construct shares more variance with its indicators than with other constructs. Cross-loading analysis provides a detailed item-level assessment by examining if indicators load more strongly on their assigned constructs than on others, thereby confirming the uniqueness of each construct's measurement.

(I) Heterotrait-monotrait (HTMT) ratio: As depicted in Table 5.3, the HTMT value for all the constructs typifies that the validity is confirmed as their values are less than 0.9 (Kline, 2016).

(II) Fornell-larcker criterion: In order to further confirm the discriminant validity, another assessment was conducted using the Fornell-larcker criterion. According to the result shown in Table 5.4, it is demonstrated that the diagonal values, which represent the correlations between the same constructs (i.e., values in bold in Table 5.4), are the highest among all the values in

the table. This implies that the constructs have a stronger correlation with themselves than with any other construct, implying satisfactory discriminant validity. Therefore, it can be concluded that the constructs are distinct and not measuring the same underlying concept (Adabre et al., 2021; Chin, 1998).

Table 5. 3**:** Heterotrait-monotrait (HTMT) ratio

Constructs	External-related factors	Failure modes	Operation-related factors	Pipe-related factors	Soil-related factors
External-related factors					
Failure modes	0.610				
Operation-related factors	0.886	0.694			
Pipe-related factors	0.842	0.725	0.799		
Soil-related factors	0.869	0.523	0.839	0.681	

Table 5. 4: Fornell-larcker criterion

Constructs	External-related factors	Failure modes	Operation-related factors	Pipe-related factors	Soil-related factors
External-related factors	**0.766**				
Failure modes	0.574	**0.814**			
Operation-related factors	0.728	0.469	**0.767**		
Pipe-related factors	0.741	0.595	0.726	**0.736**	
Soil-related factors	0.732	0.400	0.693	0.680	**0.722**

(III) Cross-loading: Finally, the discriminant validity of the measurement model is investigated using the cross-loadings of the observable variables. Table 5.5 explicates the result of the analysis. As per the result, the discriminant validity is confirmed as all the observable variables have the highest loadings in their respective construct (Hair et al., 2017).

Table 5. 5: Cross loadings of the indicators

Codes	External-related factors	Failure modes	Operation-related factors	Pipe-related factors	Soil-related factors
ERF1	**0.934**	0.525	0.776	0.780	0.648
ERF3	**0.823**	0.314	0.624	0.523	0.617
ERF6	**0.565**	0.232	0.454	0.365	0.548
ERF7	**0.840**	0.543	0.559	0.682	0.632
FMD2	0.496	**0.837**	0.499	0.622	0.376
FMD3	0.327	**0.843**	0.345	0.438	0.255
FMD4	0.456	**0.910**	0.487	0.532	0.418
ORF2	0.470	0.330	**0.681**	0.613	0.445
ORF3	0.464	0.323	**0.734**	0.472	0.468
ORF4	0.494	0.521	**0.760**	0.619	0.504
ORF5	0.544	0.298	**0.755**	0.399	0.536
ORF6	0.708	0.420	**0.850**	0.539	0.711
PRF1	0.488	0.434	0.578	**0.788**	0.455
PRF3	0.746	0.613	0.582	**0.823**	0.658
PRF4	0.715	0.597	0.489	**0.819**	0.548
PRF5	0.393	0.363	0.515	**0.723**	0.400
PRF7	0.428	0.214	0.518	**0.628**	0.569
SRF1	0.320	-0.156	0.289	0.101	**0.567**
SRF2	0.711	0.465	0.672	0.627	**0.852**
SRF3	0.469	0.208	0.438	0.366	**0.778**
SRF4	0.712	0.422	0.596	0.721	**0.829**
SRF5	0.383	0.077	0.433	0.294	**0.721**

5.3.2.   **Result of the structural model**

After confirming the reliability and validity of the measurement model, path analysis was conducted to test the study hypothesis through the structural model. The path coefficients indicate the strength of relationships between constructs, with values from 0.5-1.0 suggesting strong effects (Murari, 2015). The analysis showed water pipe failure factors have a strong influence ($\beta = 0.567$) on failure modes. This result provides quantitative evidence supporting the hypothesized relationship between these two constructs.

(I) Significance test using t-test: To determine the significance of the path coefficients, a bootstrapping analysis was conducted using the t-test. The bootstrapping analysis, involving 5000 subsamples, provides a non-parametric approach to assess the precision of PLS estimates

by examining the variability of the estimates across multiple resamples of the data. Table 5.6 presents the direct relationships between the constructs for hypothesis testing. The analysis revealed that the tested hypothesis in this study was supported, as the t-value between the related constructs exceeded 2.58 at a significance level of 0.01 (Adabre et al., 2021; Hair et al., 2014). Therefore, the results provide strong evidence in support of the proposed relationships between the constructs.
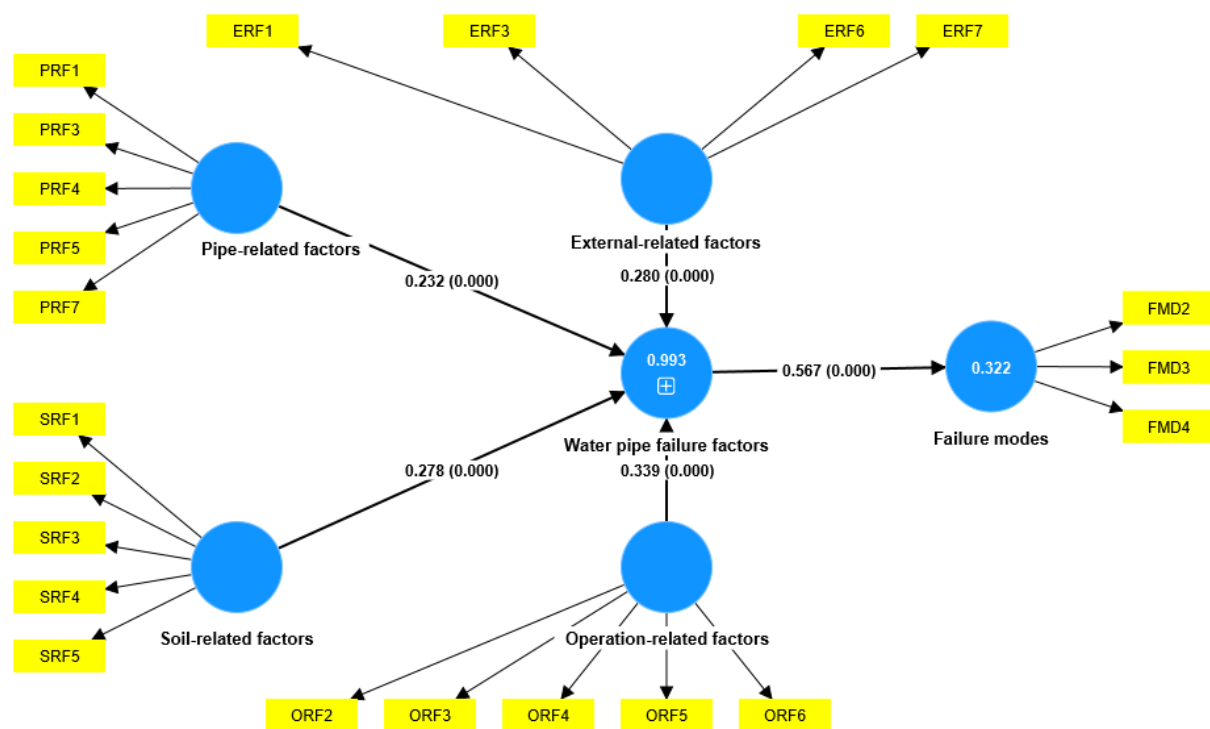
Figure 5. 4: Structural model

Table 5. 6: Direct relationship between the constructs for hypotheses testing and VIF

Hypothesis	Relationships	β values	Standard deviation	T statistics	P values	VIF
HT1	Water pipe failure factors -> Failure modes	0.567	0.064	8.912	0.000	1.308

(II) Collinearity measure: An assessment of the inner VIF was done to check the existence of multicollinearity. VIF specifically quantifies the severity of multicollinearity among predictor

variables, ensuring the independence of each factor's contribution. As shown in Table 5.6, it can be observed that the value of the VIF is less than the acceptable threshold (i.e., 3.5). This suggests that there was no issue of collinearity among the constructs, and the model's estimates of the path coefficient were not biased due to high correlations among the independent variables. Thus, the findings of the study can be considered reliable and valid.

(III) Explanatory power using R^2: Besides assessing the significance and collinearity of the constructs, the explanatory power of the model was evaluated using the coefficient of determination (R^2). As depicted in Figure 5.4, the R^2 values were 0.322 and 0.993, indicating that the model explained a significant portion of the variance in the dependent variable (Hair et al., 2017). Specifically, the result indicates that 32.2% of the failure modes variance is explained by the water pipe failure factors, which is deemed satisfactory and significant (Ali et al., 2023).

(IV) Predictive relevance using Q^2: Another important aspect of assessing the quality of a model is its predictive relevance. To evaluate this, the Q^2 measure was used, which estimates the predictive ability of the model based on a cross-validation procedure (Hair et al., 2017). The Q^2 values were calculated using the cross-validation method in the blindfolding protocol in SmartPLS software. The results presented in Table 5.7 indicate that all the constructs had positive Q^2 values, ranging from 0.135 to 0.361, suggesting that the model has satisfactory predictive relevance (Adabre et al., 2021). Specifically, the Q^2 value of 0.254 for the failure modes construct implies that the model can predict approximately 25.4% of the variance in the failure modes variable, which is considered acceptable (Hair et al., 2014). Therefore, the results provide evidence that the proposed model has both explanatory and predictive relevance, indicating its potential usefulness for understanding and predicting water pipe failure modes.

Table 5. 7**:** Result of predictive relevance of the model

Constructs	SSO	SSE	Q² (=1-SSE/SSO)
External-related factors	660.000	428.762	0.361
Failure modes	490.000	373.569	0.254
Operation-related factors	820.000	604.215	0.267
Pipe-related factors	820.000	810.000	0.135
Soil-related factors	820.000	653.766	0.202

SSO – total sum of squares and SSE – sum of square errors

(V) Assessing the effect size using f^2: To determine the strength of the impact of the independent constructs on the dependent construct, effect size (f^2) was used. This measure assesses the influence of one construct on another in terms of R^2. The changes in R^2 were evaluated to calculate the f^2, which helped identify any substantial impact between the constructs. According to (Cohen, 2013), the effect size between two constructs is regarded as weak, moderate, and substantial if f^2 ranges between 0.02 to 0.15, 0.15 to 0.35, and a value above 0.35, respectively. Therefore, as per the f^2 result shown in Table 5.8, it can be seen that a substantial impact exists between failure factors and failure modes.

Table 5. 8**:** Result of the f^2 test

Hypothesis	Relationships	f^2
HT1	Water pipe failure factors -> Failure modes	0.474

5.3.3. **Importance-performance analysis (IPMA)**

As an addition to affirming the significance of the relationship between the constructs, it is crucial to investigate the importance and performance of these constructs on the target construct (i.e., failure modes). Hence, the importance-performance map analysis (IPMA) was carried out in SmartPLS. The result of the analysis is shown in Figure 5.5 and Table 5.9. According to the result, the construct of pipe-related factors has the highest performance and total effects (i.e., importance) on the failure modes, followed by operation-related factors and external-related factors. This analysis provides valuable insights into the relative importance of each construct

and can be used to guide decision-making and resource allocation efforts in the management

of water pipe failure.

Table 5. 9: Importance-performance map analysis result

Constructs	Performance	Importance
External-related factors	74.172	0.297
Operation-related factors	78.886	0.305
Pipe-related factors	83.661	0.879
Soil-related factors	73.081	0.256
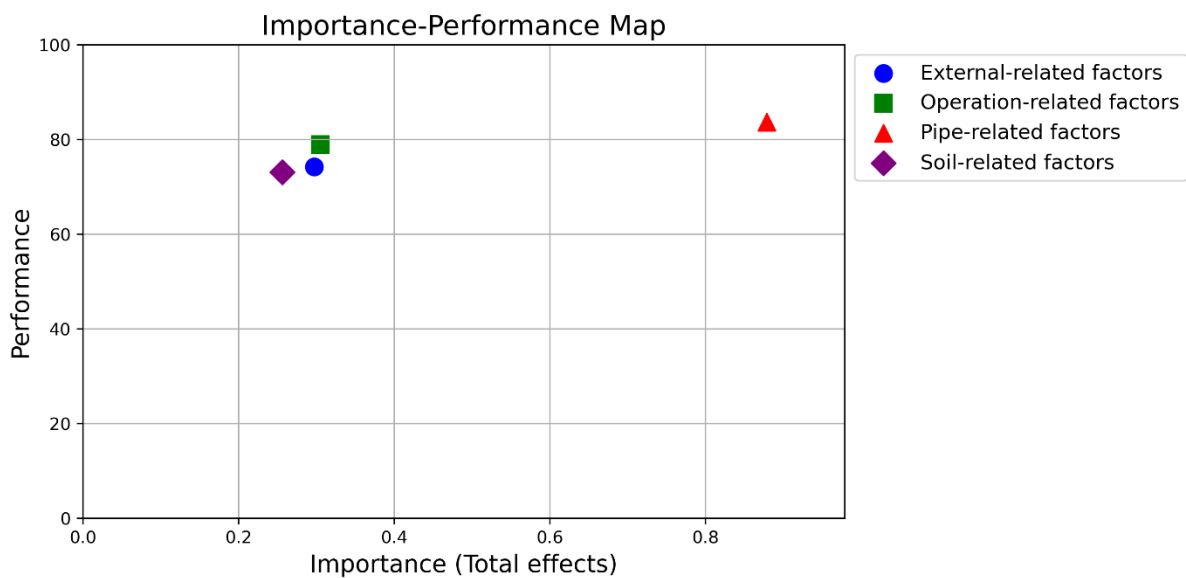
Figure 5. 5: Importance-performance map analysis of the endogenous constructs on the target construct

5.3.4. Discussion of critical water pipe failure factors and failure modes

This section presents an integrated discussion of the results obtained from the measurement

and structural models. The discussion is based on the significant indicators in each construct

and confirmation of the hypotheses made in this study.

5.2.4.1 Pipe-related factors construct

Out of seven factors reviewed as pipe-related influencing factors for water pipe failure, five

were found to be critical for water pipe failure. These factors, together with their corresponding

loadings, are PRF1 – "age" (0.780); PRF3 – "diameter" (0.832); PRF4 – "length" (0.809); PRF5 – "material" (0.707); and PRF7 – "wall thickness" (0.643).

Based on the loadings, pipe diameter was found to be the most critical pipe-related factor causing the failure of water pipes. This finding is in agreement with previous studies that have documented pipe diameter as an influential water pipe failure factor (Hekmati et al., 2020; Kutyłowska & Hotloś, 2013; Zywiec et al., 2019). Generally, an inverse relationship exists between pipe failure and diameter (Taiwo et al., 2023). That is, the higher the diameter, the lesser the pipe failure. As Hekmati et al. (2020) reported, a 600mm-diameter pipe experienced failure twice a year, whereas a pipe with a diameter of 150 mm failed 43 times during the same period. The thin wall thickness and higher pressure fluctuations exhibited by smaller-diameter pipes have been attributed to their higher failure rates (Bruaset & Sægrov, 2018; Ellison & Spencer, 2016).

The result of the analysis demonstrated that pipe length is the second most critical factor contributing to water pipe failure. It has been illustrated that the likelihood of failure of water pipes increases as the length increases (Almheiri et al., 2020; Vipulanandan et al., 2012). (Zamenian et al., 2017) found that the failure rate of some pipes increased by 0.041 when the length increased by 1 km. This may be due to the fact that long pipes are subjected to more bending moments, which can cause them to buckle or break under excessive loads. The bending stress in a pipe is proportional to the pipe length, and as the length increases, the stress on the pipe also increases. This stress can lead to pipe deformation, cracking, and, ultimately, failure (Taiwo et al., 2023). Further, longer pipes have a higher number of joints and connections, which are potential weak points where leaks or bursts can occur. The more connections a pipe has, the higher the risk of failure due to corrosion, wear and tear, and other issues.

Pipe age is another critical factor affecting water pipe failure. Generally, pipe deterioration increases with age. According to the result reported by (Ellison & Spencer, 2016), pipes whose age ranged between 35-49 years had 10% less failure compared to pipes within the age bracket of 50-65 years. This can be explained by the fact that pipes undergo degradation as they become older. This degradation can be associated with corrosion, wear, tear, and fatigue (Fares & Zayed, 2010; Hekmati et al., 2020).

The material of the pipe is another critical factor that could contribute to pipe failure. Common pipe materials used in WDN are asbestos (AC), cast iron (CI), ductile iron (DI), concrete, galvanized iron (GI), polyethylene (PE), polyvinyl chloride (PVC), and steel (Clair & Sinha, 2014; Taiwo et al., 2023). Different pipe materials have different properties, such as strength, resistance to corrosion, and durability, which affect their susceptibility to failure (Fares & Zayed, 2010). For example, pipes made from cast iron are more susceptible to failure than pipes made from newer materials like DI or PVC. CI pipes are prone to corrosion, leading to pipe thinning, leaks, and eventual failure. On the other hand, DI pipes are more corrosion-resistant and have better mechanical properties, making them less susceptible to failure.

Additionally, the wall thickness of the pipe has been recognized as a significant factor affecting pipe failure (Bruaset & Sægrov, 2018). Pipes with thin walls are more susceptible to cracks, leaks, and bursts due to excessive internal pressure and external loads. Thin-walled pipes also have lower resistance to corrosion, erosion, and wear, which can weaken the pipe structure and reduce its lifespan. Besides, thinner pipes are more susceptible to bending and buckling under external loads, such as soil movement, traffic loads, and temperature changes. These deformations can lead to structural damage, including cracks and ruptures, which can cause water loss and service disruptions.

## 5.2.4.2.	**Operation-related factors construct**

The analysis indicates that the operation-related factors construct has 5 significant factors. These factors in conjunction with their corresponding loadings are ORF2 – "internal pressure" (0.676); ORF3 – "maintenance practices" (0.721); ORF4 – "number of leaks" (0.748); ORF5 – "water acidity" (0.777); and ORF6 – "water alkalinity" (0.861).

The result shows that internal pressure contributes significantly to water pipe failure. When water flows through a pipe, it exerts pressure on the inner walls of the pipe, and if the pipe is unable to withstand this pressure, it can lead to the failure of the pipe. The extant literature has established a linear relationship between internal pressure and pipe failure (Kabir et al., 2015; Poojitha & Jothiprakash, 2022; Rezaei et al., 2015). According to an experiment conducted by (Rathnayaka et al., 2017) on two exhumed pipes, the pipes exhibited small cracks and leakage upon applying 3.25 and 3.60 Mpa pressure on them. (Ellison & Spencer, 2016) found that pipes subjected to low and moderate pressures experienced about 2.5 and 1.8 times less failure compared to pipes subjected to high pressures.

Additionally, both maintenance practices and the number of leaks have been found to influence pipe failure. Neglecting regular maintenance activities can accumulate debris, sediment, and other materials that can block the pipe or cause internal corrosion. This negligence will not only lead to leaks and bursts but also cause water quality issues (Barton et al., 2019). Small water pipe breaks can go undetected for a very long time if the WDN is not properly maintained, resulting in water loss and damage to the surrounding areas. Regular inspections and leak detection programs can help identify leaks early on and prevent pipe failure. Similarly, a linear relationship between the number of leaks and ultimate pipe failure has been established in the existing literature (Rathnayaka et al., 2017; Taiwo et al., 2023). The more leaks a pipe has, the more likely it is to fail due to the cumulative effect of the damage. Leaks facilitate

water loss, decreased water pressure, and water supply contamination. Furthermore, a leaking pipe can gradually cause erosion of the surrounding soil, which can negatively impact the bedding condition of the system (K. T. H. Liyanage & Dhar, 2017). This erosion process can lead to increased stress on the wall thickness of the pipe.

Water acidity and alkalinity are other operation-related critical factors affecting the failure of water pipes. When the pH level of water is low (i.e., acidic), the likelihood of corrosion is high. Acidic water may react with the metallic components of pipes, such as iron, zinc, and copper, leading to the formation of metal ions and the release of hydrogen gas (Jun et al., 2020). This process is known as chemical corrosion. It was demonstrated by (Zraick et al., 2019) that the failure rate of pipe increases as the corrosion rate increases. The corrosion of pipes due to water acidity potentially facilitates thinning of the pipe walls, which can eventually lead to leaks and pipe failure. The thinning of pipe walls also makes them more susceptible to external forces, such as soil movement, thereby causing further deterioration (Hu & Hubble, 2007). In the same vein, water with high alkalinity can cause scale formation inside pipes. Scale is a build-up of mineral deposits that adhere to the pipe's interior walls. This scale narrows the pipe's diameter, reducing the water flow and increasing the frictional loss of water (Arriba-Rodriguez et al., 2018). This often results in a decrease in water pressure and an increase in the energy required to pump water through the system (Taiwo et al., 2023).

5.2.4.3.   **External-related factors construct**

The initial conceptual model identifies eight external-related factors that contribute to pipe failure. However, only four of these factors were found to be critical for the failure of water pipes. The critical factors, along with their corresponding loadings, are as follows: ERF1 – "temperature" (0.905); ERF3 – "frost" (0.816); ERF6 – "microbiologically induced corrosion" (0.584); and ERF7 – "precipitation" (0.851).

According to the result, temperature was found as the most critical external-related factor influencing water pipe failure. This conforms with the extant literature investigating the impact of temperature on pipe failure (Tran et al., 2021; Zamenian et al., 2017). Extreme temperatures can cause the pipes' thermal expansion or contraction, resulting in stress and strain on the material, which ultimately causes the pipe to fail. For example, during cold winter months, when the temperature is low, water can freeze inside the pipes and expand, causing the pipes to rupture or burst. As per the finding by Bruaset & Sægrov (2018), a decrease in temperature from 23°C to -15°C brought about an 86% increment in pipe failures. Similarly, during hot summer, high temperatures can cause the pipes to expand, which puts excessive stress on the joints and results in failure (Gao, 2017). Moreover, temperature fluctuations may facilitate the expansion and contraction of the soil around the pipes, thereby leading to movement and displacement of the pipes.

Frost is another significant factor that has been found to contribute to water pipe failure. When water freezes inside a pipe, it expands and creates additional stresses on the pipe walls. Subsequently, as the water thaws, the pressure is released and often causes pipe leakage. Similarly, if the pipe is not properly insulated or buried deep enough underground, it can be more susceptible to freezing and subsequent damage from frost. The surrounding soil may also be impacted by frost action. The soil can shift and settle upon freezing and thawing, leading to uneven ground conditions and affecting pipe stability (Zywiec et al., 2019).

The finding also revealed that water pipe failure is exacerbated by microbiologically induced corrosion (MIC). MIC occurs in water pipes when certain bacteria, such as sulfate-reducing bacteria and acid-producing bacteria, form colonies on the interior surfaces of the pipes (Taiwo et al., 2023). These colonies produce organic and inorganic acids, hydrogen sulfide, and other

deleterious substances that can attack the metal surface of the pipes and cause corrosion. (San et al., 2012) found that the presence of two bacteria (i.e., Aeromonas salmonicida and Delftia acidovorans) contributed to a mass reduction of a steel pipe by 1.86 and 2.01 μg, respectively. The corrosion caused by MIC is often localized and leads to the formation of cracks and pits in the pipe walls, which facilitate the pipe burst and reduce water flow. To mitigate MIC, it is important to control the water chemistry and maintain a proper balance of disinfectants in the WDN.

After temperature, precipitation is the second-ranked (i.e., in terms of the outer loading) critical factor that influences water pipe failure. Precipitation, such as rainfall or snow, can increase the likelihood of pipe failure in several ways. It increases the level of groundwater, causing the water table to rise. When this happens, it exerts additional pressure on buried pipes and causes them to shift or move, leading to damage or failure. Precipitation also affects the water quality and chemistry within the pipes, which results in corrosion or other forms of deterioration. For instance, acidic precipitation can cause corrosion in metal pipes, while excessive minerals or salts in the water can lead to scaling or mineral build-up in the pipes, reducing their effectiveness over time. As reported by (Hekmati et al., 2020), the winter season had the highest failure rate of water pipes among all seasons, attributed to the substantial rainfall during this time.

5.2.4.4. Soil-related factors construct

The conceptual model initially identified six soil-related factors that could contribute to water pipe failure. However, after analysis, five of these factors were deemed critical. These critical factors, along with their corresponding loadings, are as follows: SRF1 - "bedding condition" (0.571), SRF2 - "soil moisture" (0.815), SRF3 - "Soil pH" (0.746), SRF4 - "soil resistivity" (0.811), and SRF5 – "soil type" (0.733).

Bedding condition is one of the identified critical factors influencing pipe failure. The bedding condition refers to the quality of the material used to support and protect the buried pipes. When the pipes are installed, they are usually placed in trenches that are then backfilled with a bedding material such as sand, gravel, or crushed stone. The quality of these materials often has a significant impact on the lifespan of the pipes. Inadequacy of bedding materials has been found in soils with low bearing capacity, such as organic soil, and soils with a high potential for shrinking/swelling and retaining water content, such as clay (Pritchard et al., 2013). Furthermore, pipes with bedding materials that are not properly compacted can become susceptible to external loads and stresses. Poor bedding conditions can also lead to misalignment or uneven support, leading to stress concentrations and resulting in pipe failure. Soil moisture is another critical factor contributing to the failure of water pipes. This is in agreement with previous studies, as it has been found that corrosivity increases with increasing moisture content until it reaches an optimal level and then decreases (Noor & Al-Moubaraki, 2014; Pritchard et al., 2013). This optimum level that is responsible for the highest corrosion rate in water pipes differs from one location to another, depending on the environmental conditions. For example, (Noor & Al-Moubaraki, 2014) found 10% as the optimum moisture content, while the finding of (Pritchard et al., 2013) indicated it to be between 50-65%.

The measurement model's results also indicate the criticality of soil pH. The pH of the soil is the measure of its acidity or alkalinity, with values below 7 indicating acidity and values above 7 indicating alkalinity. This factor influences the activity of microorganisms that live in the soil and can promote the growth of certain bacteria influencing corrosive activities in water pipes. In acidic soils, the presence of organic acid, hydrogen, and sulphide ions increases the susceptibility of the pipe to corrosion. On the other hand, alkaline soils can lead to water pipe failure by stimulating the growth of certain organisms that cause corrosion. For example,

alkaline soils are favorable for the growth of sulphate-reducing bacteria (SRB), which produce hydrogen sulphide gas as a by-product of their metabolic activity (Hou et al., 2016).

Soil resistivity and soil type are other significant factors causing water pipe failure. The soil's resistance to the flow of electric current is known as soil resistivity. High soil resistivity indicates that the soil is less conductive to electric current flow, resulting in a lower pipe corrosion rate. This is because the electrochemical reaction that causes corrosion requires a current to flow, and if the soil has high resistivity, the current flow will be reduced. On the other hand, when the soil resistivity is low, it means that the soil is more conductive and can facilitate the flow of electric current, leading to a higher corrosion rate (Najjaran et al., 2006; Taiwo et al., 2023). In the same vein, the type of soil in which a water pipe is buried can affect its durability and longevity. Different soil types have distinct properties that can impact the corrosion resistance of pipes. For instance, clay soil is associated with low permeability, which is a property that can promote soil corrosivity. Additionally, organic acid can be produced from clay soil with high levels of biological substances, which can corrode metal pipes (Doyle et al., 2003; Pritchard et al., 2013). Furthermore, sandy soils have high permeability and may allow water to drain easily, reducing the build-up of water pressure around the pipe. However, this property and the lower moisture retention capacity of sandy soil, thereby leading to dry conditions facilitating the development of cracks in the soil (Doyle et al., 2003). Hence, understanding the properties of various soil types and their potential impact on pipes can assist in selecting appropriate materials and installation techniques to minimize the risk of failure.

5.2.4.5. Failure modes construct

The failure modes construct is an important aspect of the model as it shows how water pipe failure factors affect the failure modes. Based on the measurement model results, three out of

five identified failure modes were found to be critical modes of water pipe failure. These modes are blown-out (FMD2), circumferential cracking (FMD3), and corrosion splitting (FMD4). Blown-out refers to the sudden rupture or breakage of a water pipe due to excessive internal pressure. This mode of failure is commonly associated with pipes with weak or defective construction, such as those made of poor-quality materials, or pipes damaged during installation or operation activities (Tang et al., 2019b). In addition, poor maintenance practices, including inadequate monitoring and repair, can increase internal pressure and cause the pipe to fail catastrophically. Blown-out was found as the major failure mode in a WDN in Australia (Rajeev et al., 2014). The measurement model result showed that blown-out has a significant loading of 0.829, indicating its criticality in water pipe failure.

Circumferential cracking occurs when a water pipe experiences a crack that runs around its circumference. This type of failure can result from several factors, including aging, fatigue, material defects, or external loads (Taiwo et al., 2023). Circumferential cracking can cause a reduction in the strength of the pipe and eventually lead to failure. This mode of failure is common in small-diameter pipes (Barton et al., 2019). The criticality of this failure mode is supported by its significant loading of 0.856.

Corrosion pitting refers to the creation of pits on water pipes due to corrosion. Metallic pipes are most susceptible to this failure mode, as they are more prone to corrosion compared to other materials (Rajeev et al., 2014). Corrosion can be caused by a variety of factors, including the chemistry of the water and the soil in which the pipe is buried. Corrosion splitting is considered a critical failure mode of water pipes, as indicated by its high loading of 0.907.

5.2.4.6.   **Hypothesis testing**

The result of the hypothesis that was tested in this study implied that the identified water pipe failure factors significantly influence the pipe's failure modes ($p \leq 0.05 \ and \ \beta = 0.567$). In

other words, the pipe-related, operation-related, external-related, and soil-related factors were found to be associated with how the water pipe failed. For instance, if a pipe has experienced significant corrosion due to aggressive chemicals or soil conditions, it may be more prone to failure due to corrosion pitting (Rajeev et al., 2014). Similarly, if the pipe's diameter is too small or its wall thickness is too thin, it is more likely to experience blown-out failures (Tang et al., 2019b).

5.4.   RESULT OF THE POF MODEL

This section presents and discusses the result of the experiments discussed in section 3.4. All the experiments were implemented using the Python environment. For experiments relating to hyperparameters and features optimization, "accuracy" and "f1 score" were used as fitness functions.

5.4.1.   Result and validation of optimizing logistic regression hyperparameters

As stated in the model development section (Section 3.4), the data was divided into two parts, where 80% was used for training, and the remaining 20% was used for model validation. The results reported in this section are based on the model output using the validation dataset. As GA is a stochastic process, the model was run 20 times, and the mean of the results was adopted. The confusion matrix for experiment 1 is presented in Table 5.10 using the validation dataset, and the five evaluation metrics described previously are presented in Table 5.11. As seen in Table 5.11, the accuracy score of the model increased from 0.819 to 0.856 when the data was transformed using standardization. This shows that data transformation is important to improve the accuracy of the model. Data standardization with the f1 score as the fitness function produced the highest values for all the evaluation metrics except for recall, which is produced by data normalization using the f1 score. Overall, models that use an f1 score as the fitness function outperformed those that use accuracy as the fitness function. This implies that the f1

score is a suitable fitness function for optimizing the hyperparameters of the model developed

in this experiment. Similarly, Figure 5.6 shows the ROC curves of the non-scaled data model

and those whose data were transformed and used the f1 score as the fitness function for the

optimization process. The AUC of the curves (see Table 5.11) are 0.849, 0.882, and 0.904 for

non-scaled, normalized, and standardized data, respectively. To demonstrate the effectiveness

of selecting the best hyperparameters in ML modeling, the data was fitted on LR using the

default hyperparameters of a Python library. The values of all the evaluation metrics were lower

than the ones achieved in experiment 1. The detailed results are provided in Appendix C.

Table 5. 10: Confusion matrix of experiment 1

	For non-scaled data using accuracy as the fitness function	
	Predicted Failure	Predicted Intact
True Failure	TF = 7523	FI = 2775
True Intact	FF = 1343	TI = 11204
	For non-scaled data using f1 score as the fitness function	
	Predicted Failure	Predicted Intact
True Failure	TF = 7681	FI = 2617
True Intact	FF = 1251	TI = 11296
	For normalized data using accuracy as the fitness function	
	Predicted Failure	Predicted Intact
True Failure	TF = 7882	FI = 2416
True Intact	FF = 1058	TI = 11489
	For normalized data using f1 score as the fitness function	
	Predicted Failure	Predicted Intact
True Failure	TF = 7938	FI = 2360
True Intact	FF = 1005	TI = 11542
	For standardized data using accuracy as the fitness function	
	Predicted Failure	Predicted Intact
True Failure	TF = 7933	FI = 2365
True Intact	FF = 954	TI = 11593
	For standardized data using f1 score as the fitness function	
	Predicted Failure	Predicted Intact
True Failure	TF = 7935	FI = 2363
True Intact	FF = 926	TI = 11621

Table 5. 11: Evaluation metrics of experiment 1

| Data transformation	Accuracy	Precision	Recall	F1 score	AUC
Non-scaled data (accuracy)	0.819	0.848	0.730	0.785	0.830
Non-scaled data (f1 score)	0.830	0.859	0.745	0.798	0.849
Normalized data (accuracy)	0.847	0.881	0.765	0.819	0.867
Normalized data (f1 score)	0.852	0.887	**0.771**	0.825	0.882
Standardized data (accuracy)	0.854	0.892	0.770	0.827	0.874
Standardized data (f1 score)	**0.856**	**0.895**	0.770	**0.828**	**0.904**
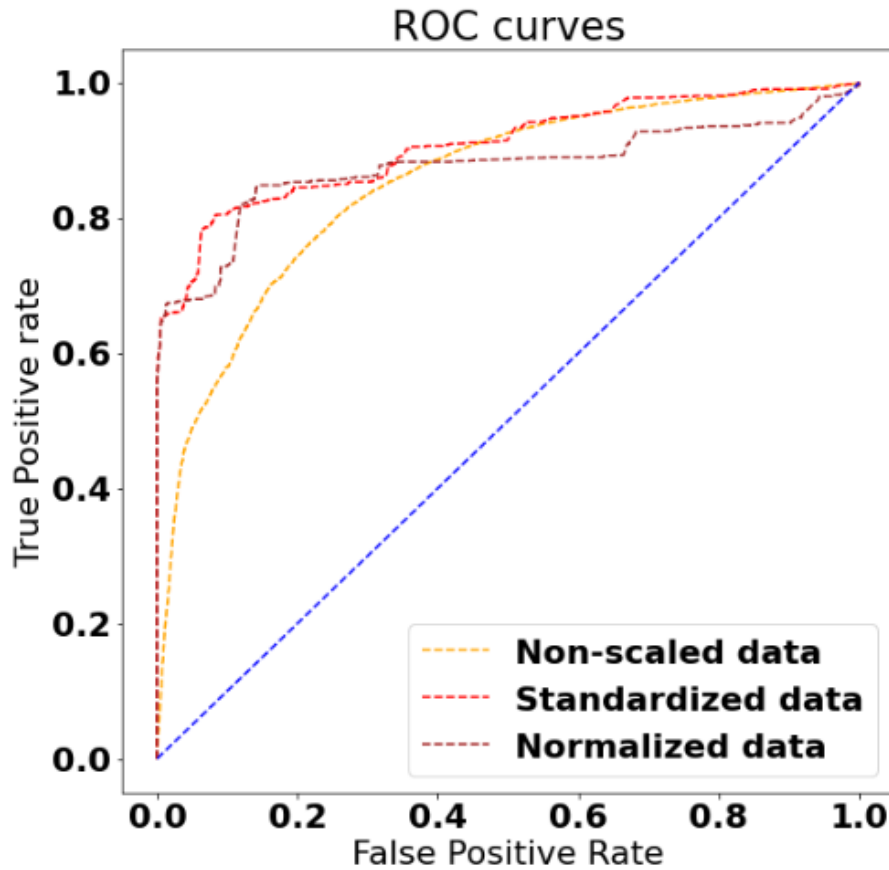
Figure 5. 6: The ROC curves in experiment 1

5.4.2. **Result and validation of optimizing logistic regression features**

This experiment is an extension of experiment 1 by adding a selection of the best features

needed to develop the LR model. After running the model several times, it was discovered that

different combinations of factors were selected by the GA. Hence, ten simulations were carried out, and their results were recorded. The 7th simulation achieved the highest performance in terms of the evaluation metrics, and 5 out of the 8 factors selected in this simulation occur in the other 9 simulations. The selected features are presented in Figure 5.7. The confusion matrix of experiment 2 can be found in Appendix C (see Table C3), while the evaluation metrics are depicted in Table 5.12.

It is noted that optimizing the hyperparameters and features using the f1 score as the fitness function produced the highest accuracy (0.889), precision (0.929), and AUC (0.944) using the validation dataset. However, the optimization using accuracy as the fitness function yielded the highest recall (0.816) and f1 score (0.868). This indicates that a modeler should explore these two fitness functions and determine the appropriate function for each evaluation metric. Although data standardization achieved the best results, the results are comparable with those of data normalization, especially for the AUC metric. This implies that both scaling methods should be applied to a dataset to determine the appropriate scaling method for each evaluation metric. The ROC curves of this experiment using f1 as the fitness function are plotted in Figure 5.8. The AUC values for these curves are 0.936, 0.943, and 0.944 for non-scaled, standardized, and normalized data. For the model evaluation, 88.9% (accuracy) of the pipes' status is correctly predicted, and 92.9% (precision) of all the failed pipes is correctly predicted. This means that a very low percentage of the pipes (7.1%) may be wrongly replaced. Similarly, the recall achieves a value of 81.5%, typifying that only 18.5% of the failed pipes are misclassified. Overall, the model developed in this experiment can efficiently assist WSD managers in predicting the status of a pipe with a very low false alarm rate, thereby contributing to the sustainable management of WDNs.

Table 5. 12: Evaluation metrics of experiment 2

Data transformation	Accuracy	Precision	Recall	F1 score	AUC
Non-scaled data (accuracy)	0.876	0.904	0.810	0.854	0.934
Non-scaled data (f1 score)	0.877	0.906	0.812	0.856	0.936
Normalized data (accuracy)	0.883	0.917	0.814	0.863	0.933
Normalized data (f1 score)	0.887	0.926	0.814	0.866	0.943
Standardized data (accuracy)	0.888	0.928	**0.816**	**0.868**	0.945
Standardized data (f1 score)	**0.889**	**0.929**	0.815	**0.868**	**0.944**

Feature 1	Age
Feature 2	Diameter
Feature 3	Length
Feature 4	Material
Feature 5	Precipitation
Feature 6	AADT
Feature 7	Soil corrosivity
Feature 8	Temperature

Figure 5. 7: The selected features in experiment 2
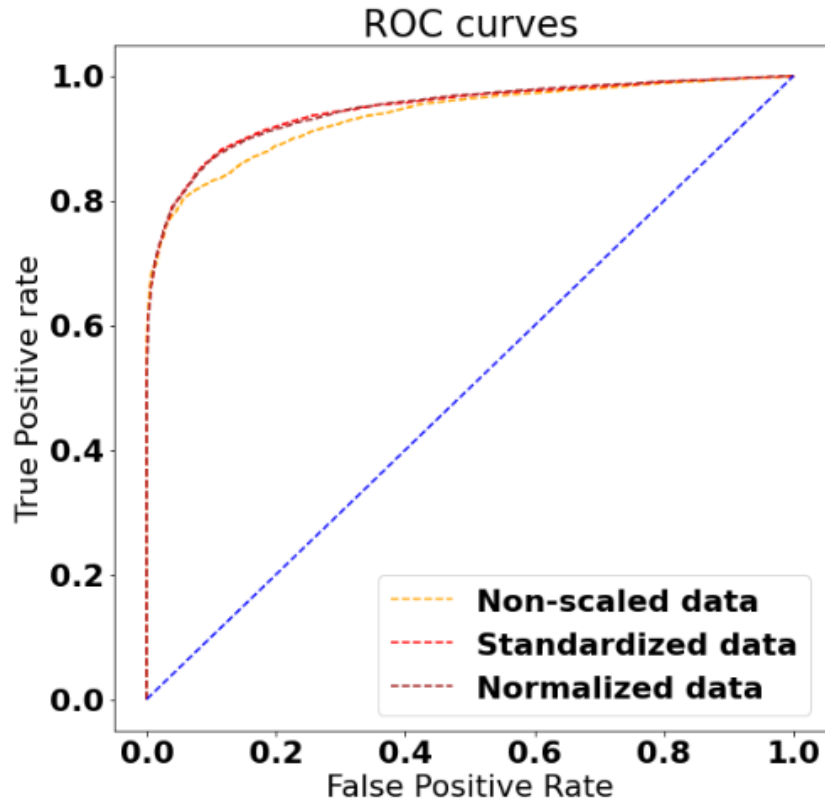
Figure 5. 8: The ROC curves in experiment 2

5.4.3. Model interpretability results

This section presents the coefficients of the model (the best model developed in experiment 2) and the results of the SHAP experiment. The coefficients of the LR model are shown in Table 8. By imputing the coefficients in Equation 3.6, water utility managers can determine the probability of failure of each pipe. This shows the advantage of explainable algorithms such as LR over other ML algorithms, often called black boxes. However, a limitation of LR in terms of interpretability is the fact that it cannot handle multicollinearity between the independent variables (Vaulet et al., 2022). It should be noted that this limitation only affects LR's interpretability and not its predictive capability. Hence, Figure 5.9 shows the contribution of each feature to the prediction using the SHAP algorithm.

Despite the limitation of LR's interpretability, the four most important features denoted by the magnitude of the LR's coefficients agree with the result of the SHAP experiment: age,

161

temperature, material_CI, and length (see Table 5.13). It is not surprising that pipe age contributes most to the model's prediction, as pipe failure probability increases as the age of the pipe increases. Additionally, older pipes tend to have a higher failure probability as the corrosion rate is a function of the pipe's age (Ji et al., 2020). Moreover, temperature high contribution to the failure prediction can be explained by the fact that a drop in the atmospheric temperature can facilitate the pipe's contraction, leading to stress development on the pipe and therefore contributing to its failure. Furthermore, CI material has the highest contribution to failure prediction compared to other materials in the network. This is due to the higher susceptibility of CI to corrosion, which thus increases its failure rate. This result agrees with the action taken by the HK WSD by discontinuing the usage of CI material for new pipe installation (Water Supplies Department HKSAR, 2021). Length ranks as the fourth important feature for failure prediction. This stems from the fact that the failure rate increases with pipe length as more pipe areas are exposed to environmental stress (Zamenian et al., 2017). On the order hand, AADT, soil corrosivity_none, and material_DI are the least important feature for the prediction. This shows that the differential loading resulting from the traffic has minimal effect on the structure of the pipes in HK WDN. Similarly, a lesser number of failed pipes are located in noncorrosive soil and manufactured with DI.

Since SHAP values are determined for each sample in the dataset, the direction of the contribution of each feature on the model's output can be determined. Figure 5.10 shows the distribution of the SHAP values on the model's outcome. The gradient color depicts the feature value for each sample, which spreads from red (highest) to blue (lowest). For continuous features, the value can be represented by the whole spectrum, while dummy variables take only two colors (i.e., red and blue). It can be seen from Figure 5.10 that diameter is associated with a high feature value and negative SHAP value. This means that diameter is negatively correlated with the failure probability of water pipes.

Table 5. 13: Coefficients of the optimized LR model

Feature	Coefficients
Length	0.874
Age	0.857
Temperature	-0.731
Material_CI	-0.468
Material_Steel	0.312
Diameter	-0.275
Precipitation	0.156
Material_AC	0.128
Soil corrosivity_Moderate	0.085
Material_Plastic	0.082
Soil corrosivity_High	0.034
Soil corrosivity_None	0.022
Material_DI	0.008
AADT	-0.005
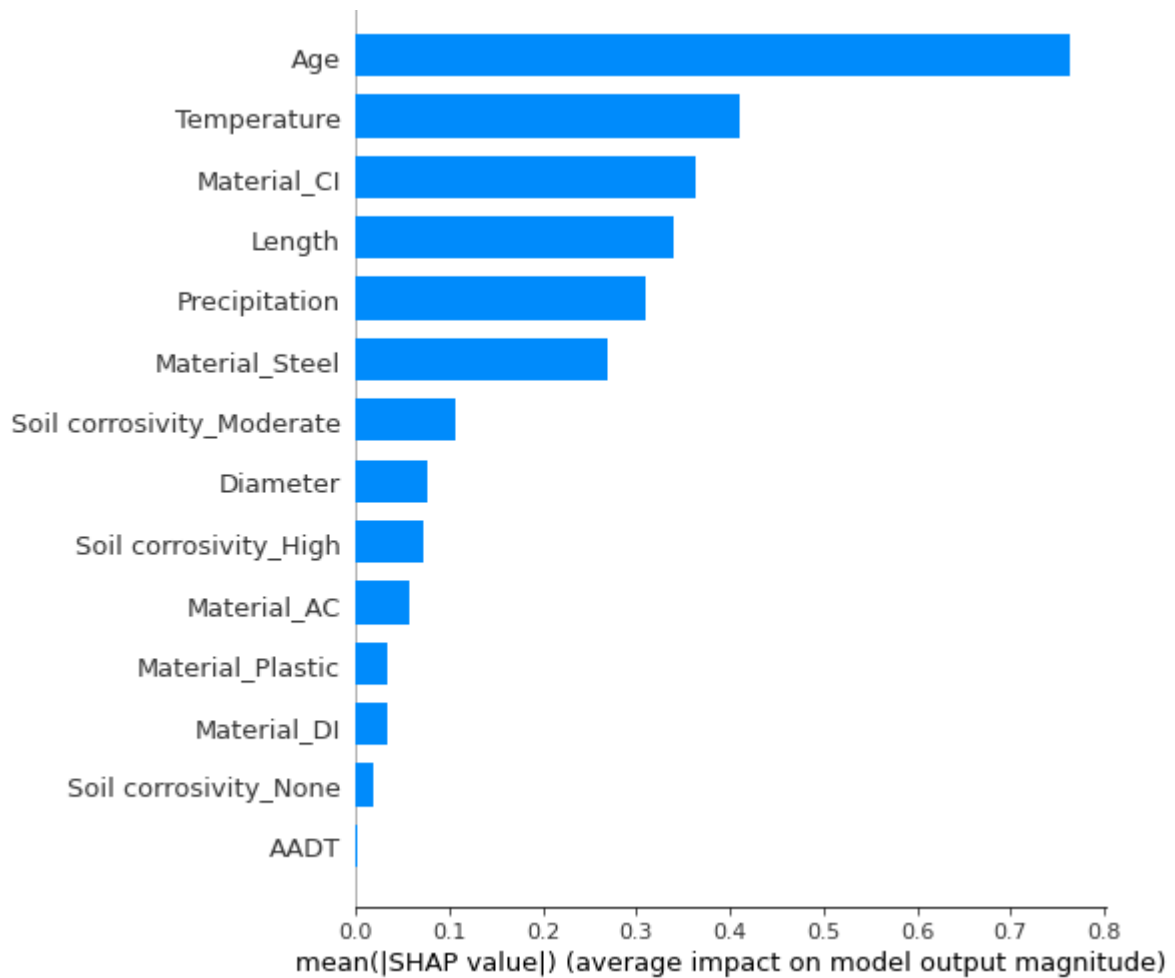
Figure 5. 9: The contribution of each feature to the model's prediction

This suggests that pipes with larger diameter experience lesser failures compared to those with smaller diameters. This is supported by the coefficient of diameter in the model (-0.275) and extant literature (Bruaset & Sægrov, 2018; Zywiec et al., 2019). Furthermore, length, precipitation, and among others have positive contributions to the failure probability of water pipes.
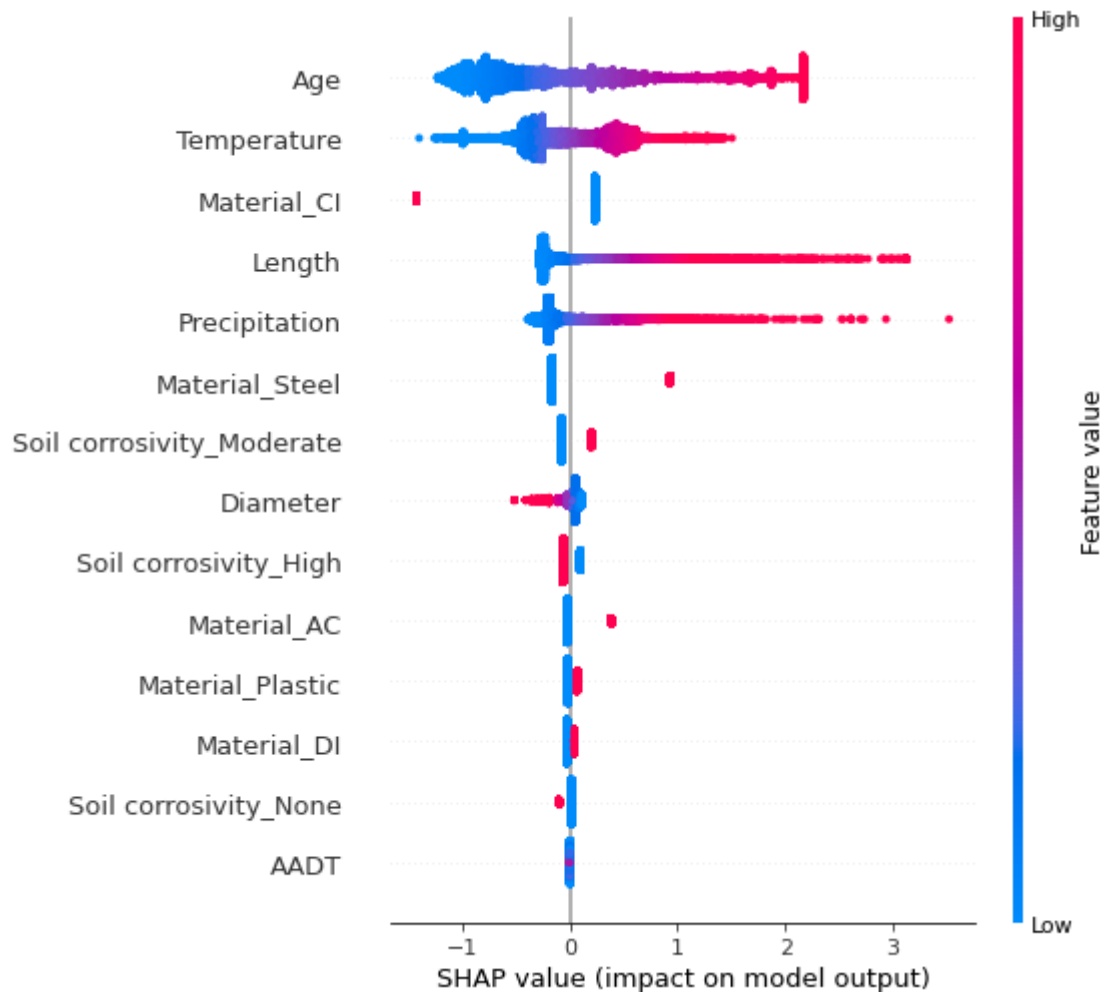
Figure 5. 10: Distribution of SHAP values for each feature on the model's output

5.5. RESULT OF THE POL AND POB MODELS

This section presents and discusses the results of the base-DL models, optimized-DL models, selection of the best DL model, and its interpretability in relation to POL and POB models.

5.5.1. Result and validation of the base-DL models

After training the DL models, their efficiency was evaluated using the validation dataset. Table 5.14, which was generated using the confusion matrix (see Table D1), displays the performance of these base-DL models, highlighting differences among scaling methods and models. Given the dataset's highly imbalanced nature, where leaks are relatively rare compared to non-leak instances, accuracy alone is not a sufficient measure of model performance. This imbalance also contributes to the generally lower precision across all models, as the number of true positive predictions is small relative to the number of false positive predictions. When comparing the scaling methods, it's evident that both normalization and standardization improve the performance across all models compared to non-scaled data. This improvement is particularly notable in the precision metric, which is critical in imbalanced datasets. For example, the precision for the DNN model increases from 0.584 with non-scaled data to 0.640 with normalized and 0.726 with standardized data, indicating that the likelihood of correctly identifying a leak has increased with proper scaling. Recall is robust across models, indicating a strong ability to identify actual leaks. The F1 score, which balances the precision and recall, is highest with the CNN model in standardized data and closely followed by CNN, underscoring its efficiency in managing the trade-off between identifying leaks and avoiding false alarms. MCC and Cohen's Kappa provide a more nuanced view of the models' performance by considering true negatives and the imbalance in the dataset. These metrics are particularly high with standardized data, reflecting a better true positive rate relative to the imbalance in the dataset. Similarly, Table 5.15 presents the evaluation metrics for the POB models (the confusion matrix is presented in Table D2). According to the table, recall values are high for all models, suggesting that the models are generally successful at identifying the majority of the actual burst events. In most cases, the performance of the model increases when the data is normalized or standardized. For instance, the F1 score of TabNet improved by 32.6%

when the data was normalized. Generally, the results show that CNN and TabNet outperformed DNN.

Table 5. 14: Evaluation metrics for base-DL models for predicting probability of leak

Data scaling	Models	Accuracy	Precision	Recall	F1 score	MCC	Cohen's Kappa
Non-scaled	DNN	0.974	0.584	0.861	0.696	0.696	0.682
	CNN	0.975	0.603	0.866	0.711	0.711	0.698
	TabNet	0.966	0.512	0.842	0.637	0.641	0.62
Normalized	DNN	0.978	0.640	0.882	0.742	0.741	0.731
	CNN	0.981	0.668	0.914	0.772	0.772	0.762
	TabNet	0.974	0.598	0.835	0.697	0.694	0.684
Standardized	DNN	**0.985**	0.726	0.915	0.809	0.807	0.801
	CNN	**0.985**	0.725	**0.924**	**0.812**	**0.811**	**0.805**
	TabNet	**0.985**	**0.727**	0.903	0.805	0.803	0.798

Table 5. 15: Evaluation metrics for base-DL models for predicting probability of burst

Data scaling	Models	Accuracy	Precision	Recall	F1 score	MCC	Cohen's Kappa
Non-scaled	DNN	0.997	0.247	0.914	0.388	0.474	0.387
	CNN	0.997	0.268	0.924	0.415	0.496	0.414
	TabNet	0.997	0.270	0.937	0.419	0.502	0.418
Normalized	DNN	0.998	0.332	0.947	0.492	0.56	0.491
	CNN	0.998	0.366	0.950	0.529	0.589	0.528
	TabNet	0.998	0.393	0.950	0.556	0.61	0.555
Standardized	DNN	0.998	0.388	0.950	0.551	0.607	0.550
	CNN	0.998	**0.433**	0.963	**0.597**	**0.645**	**0.597**
	TabNet	**0.999**	0.395	**0.967**	0.561	0.618	0.561

*number in bold represent the best result

5.5.2. Result and validation of the optimized-DL models

Table 5.16 presents the evaluation metrics for the optimized DL models using the validation dataset to predict the probability of leaks in water pipes (see Table D3 for the confusion matrix). Consistent with the base-DL models, the performance metrics of these optimized models demonstrate improvement with data normalization and standardization, reinforcing the significance of data scaling in DL applications. Markedly, the precision and recall of the

optimized DNN model showed increases of 12.7% and 8.4%, respectively, upon standardization. While normalization typically enhances model performance, the optimized TabNet model's slight decrease in precision and F1 score, suggesting its potential preferential alignment with non-scaled data. The standardized DNN and CNN models exhibit high F1 scores and MCC values, indicative of a well-calibrated balance between precision and recall, alongside a robust alignment between predicted outcomes and actual occurrences. When contrasting the base and optimized DL models, the superior performance of the latter becomes apparent, underscoring the effectiveness of hyperparameter tuning. For instance, the non-scaled DNN model's precision, recall, and F1 score surged by 23.8%, 8.7%, and 21.0%, respectively, post-optimization.

According to Table 5.17 (see Table D4 for the confusion matrix), which shows the evaluation metrics for the POB models, TabNet generally outperforms DNN and CNN in non-scaled and normalized datasets, particularly in precision and F1 scores. CNN shows a substantial increase in performance metrics with normalized and standardized data, indicating it may be sensitive to the scaling method applied. DNN shows less variation in performance across scaling methods but doesn't reach the precision value of CNN or TabNet with normalized data. A discerning examination of Tables 5.16 and 5.17 indicates that standardized data attained the highest performance for models predicting the POL, whereas normalized data showed optimal results for models predicting the POB. This distinction underscores the tailored impact of data scaling techniques on model efficacy, contingent on the specific predictive task at hand. Using the normalized data, Figure 5.11 and 5.12 visualizes the difference between the base and optimized DL models for the POL and POB models.

5.5.3. Selection of the best DL-model

The optimized models have shown high performance in predicting the probability of leaks and bursts. Yet, selecting the top-performing model necessitates a structured method since various

models show superiority in different evaluation metrics. For example, considering the POB in normalized datasets, the optimized TabNet surpasses its counterparts in precision, whereas the CNN stands out with the highest recall. Additionally, both CNN and TabNet share equivalent scores in terms of the F1 metric and Cohen's Kappa. Thus, employing a systematic evaluation approach that considers all metrics is crucial to ascertain the most effective model. Tables 5.18 and 5.19 show the results of the Copeland algorithm for ranking the optimized POL and POB models.

Table 5. 16: Evaluation metrics for optimized-DL models for predicting the probability of leak

Data scaling	Models	Accuracy	Precision	Recall	F1 score	MCC	Cohen's Kappa
Non-scaled	DNN	0.988	0.767	0.936	0.843	0.841	0.837
	CNN	0.991	0.826	0.950	0.883	0.881	0.879
	TabNet	0.993	0.865	0.957	0.908	0.906	0.905
Normalized	DNN	0.990	0.796	0.954	0.868	0.867	0.863
	CNN	0.990	0.801	0.946	0.867	0.865	0.862
	TabNet	0.989	0.793	0.942	0.861	0.859	0.856
Standardized	DNN	**0.994**	0.865	0.969	0.914	0.913	0.911
	CNN	**0.994**	**0.876**	**0.978**	**0.924**	**0.923**	**0.922**
	TabNet	0.990	0.799	0.938	0.863	0.861	0.858

Table 5. 17: Evaluation metrics for optimized-DL models for predicting the probability of burst

Data scaling	Models	Accuracy	Precision	Recall	F1 score	MCC	Cohen's Kappa
Non-scaled	DNN	0.999	0.586	0.973	0.732	0.755	0.731
	CNN	0.997	0.307	0.980	0.468	0.548	0.467
	TabNet	0.999	0.738	0.980	0.842	0.85	0.841
Normalized	DNN	0.999	0.733	0.977	0.838	0.846	0.837
	CNN	0.999	0.782	**0.987**	**0.872**	**0.878**	**0.872**
	TabNet	0.999	**0.783**	0.983	**0.872**	0.877	**0.872**
Standardized	DNN	0.999	0.608	0.983	0.751	0.773	0.751
	CNN	0.999	0.743	0.980	0.845	0.853	0.845
	TabNet	0.999	0.697	0.987	0.817	0.829	0.817
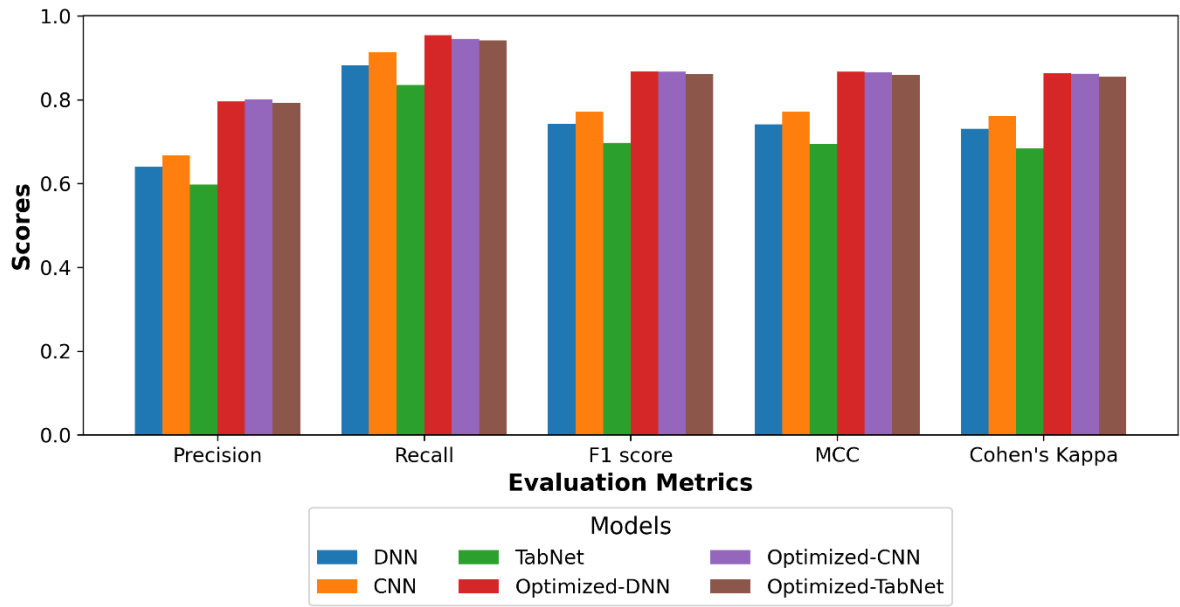
Figure 5. 11: Comparison of base and optimized-DL models for probability of leak models
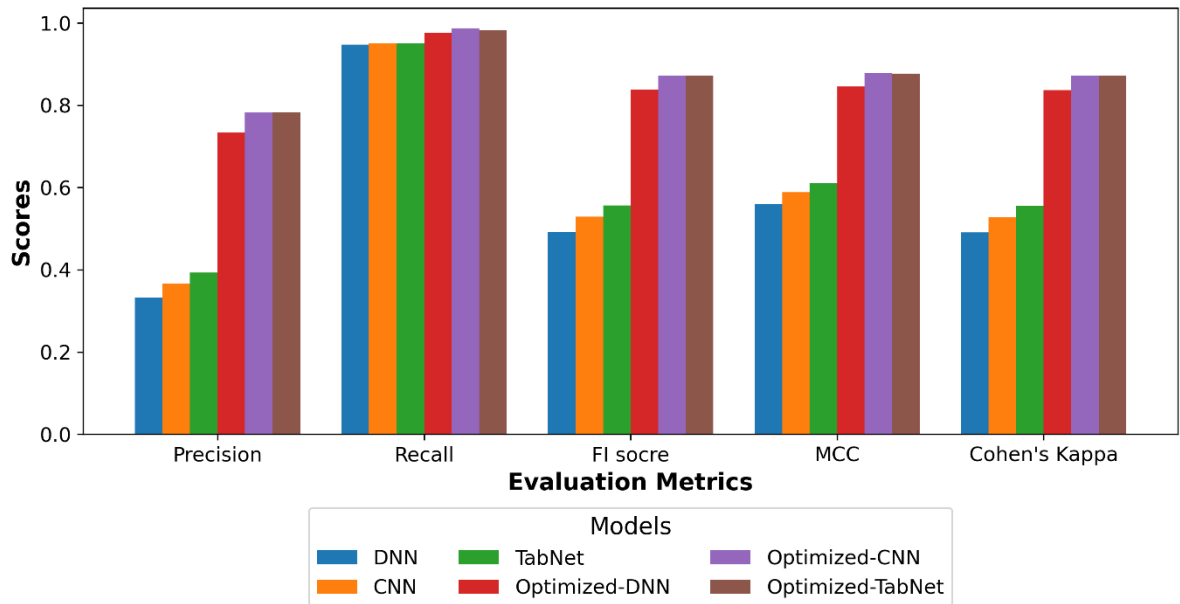

Figure 5. 12: Comparison of base and optimized-DL models for probability of burst models

Since data scaling has proved to improve the models' predictive capability, the comparison is made between normalized and standardized datasets. With normalized data, DNN performed the best with 2 wins, 0 losses, and a Copeland score of +2, giving it rank 1. For standardized

data, CNN emerged superior with 2 wins, 0 losses, and a score of +2, putting it in 1st place. Looking at the aggregate scores and ranks across both scaling methods, CNN performed the best overall with 14 Copeland points and a Copeland score of +2, securing it the top rank. DNN took 2nd place with an aggregate record of 10 Copeland points. TabNet came last with 0 wins, 4 losses, and a score of -4. The superior performance of CNN in this context suggests its potential as a reliable choice for predicting the probability of leak for the WDN.

Table 5.19 demonstrates consistency in model performance across both scaling methods. For the normalized data set, CNN takes the lead with the highest Copeland point, score, and rank, suggesting its better adaptability to normalized data, unlike DNN, which falls to the lowest rank. The standardized data set also sees CNN and TabNet performing well, but DNN lags behind. The aggregate scores across both data sets reveal CNN's superiority over others. The overall results highlight the importance of considering different data scaling methods when evaluating model performance. Each model's strengths and weaknesses become apparent under varying conditions, underscoring the necessity of choosing a model not only based on its overall accuracy but also on its adaptability to different data representations.

Table 5. 18: Results of the Copeland algorithm for ranking the probability of leak models

Data scaling	Model	Copeland Point	Wins	Losses	Copeland Score	Rank
Normalized	DNN	9	2	0	2	1
	CNN	3	1	1	0	2
	TabNet	-12	0	2	-2	3
Standardized	DNN	1	1	1	0	2
	CNN	11	2	0	2	1
	TabNet	-12	0	2	-2	3
Aggregate	DNN	10	3	1	2	2
	CNN	**14**	**3**	**1**	**2**	**1**
	TabNet	0	0	4	-4	3

Table 5. 19: Results of the Copeland algorithm for ranking the probability of burst models

Data scaling	Model	Copeland Point	Wins	Losses	Copeland Score	Rank
Normalized	DNN	-10	0	2	-2	3
	CNN	6	2	0	2	1
	TabNet	4	1	1	0	2
Standardized	DNN	-8	0	2	-2	3
	CNN	6	2	0	2	1
	TabNet	2	1	1	0	2
Aggregate	DNN	-18	0	4	-4	3
	CNN	**12**	**4**	**0**	**4**	**1**
	TabNet	6	2	2	0	2

5.5.4. Interpretability of the best DL-model

Following the selection result delineated in Section 5.5.3, the optimized CNN model utilizing standardized data for leak prediction and normalized data for burst prediction has been identified as the most performant. Figures 5.13 and 5.14 illustrate the feature importance as determined by SHAP values for models predicting the probability of leak and burst in water pipes, respectively. For the leak prediction model, 'Diameter', 'Material_Plastic', and 'Age' are identified as the leading features influencing the model's predictions. These attributes suggest that the model places significant emphasis on the physical properties and the material composition of the pipes, along with their operational lifespan. In contrast, the burst prediction model prioritizes 'Diameter' and 'Material_Plastic', similar to the leak model, but assigns greater importance to 'Corrosivity_Highly corrosive' conditions. This differentiation in feature importance highlights the distinct mechanisms and factors the model associates with the likelihood of burst incidents as opposed to leaks.

Figures 5.15 and 5.16 elucidate the relationship between feature values and their SHAP values, which quantify the impact on the model's predictions. The color gradient in these figures, transitioning from red to blue, represents the spectrum of feature values across the dataset, with

red signifying higher values and blue indicating lower ones. For continuous variables, this gradient reflects a range of values, while for categorical variables, represented as dummy variables, the visualization simplifies to red and blue, denoting the presence or absence of the feature, respectively. According to these figures, it is observed that smaller diameters contribute positively to the model's prediction of an event (leak or burst), as indicated by positive SHAP values (Taiwo et al., 2023b). Conversely, greater ages are similarly associated with positive SHAP values, suggesting that the probability of an event increases with the age of the pipe. These insights are consistent with extant literature and intuitive expectations where narrower pipes may be more susceptible to blockages leading to leaks, and older pipes may be more prone to failure due to material degradation over time (Barton et al., 2019; Farh et al., 2023). Furthermore, the figures reveal that the presence of saltwater (denoted by 'Water_type_SW') and higher pressure levels are positively correlated with the likelihood of both leaks and bursts. This positive correlation implies that the model recognizes these conditions as risk factors, with saltwater potentially accelerating corrosion and high pressure increasing stress on the pipe system (Barton et al., 2019). The model's sensitivity to these variables underscores their importance in the predictive framework and potentially guides targeted maintenance efforts where these factors are prevalent.

5.6. RESULT OF THE COF MODEL

This section focuses on the results of the EL-models and their performance in terms of the base and optimized models. The process of selecting the most effective model among the EL-models is discussed, along with an examination of its interpretability.

5.6.1. Result and validation of the base-EL models

The performance of the base-EL models was evaluated on the validation dataset, which comprised 20% of the original data and was not used during the training process.
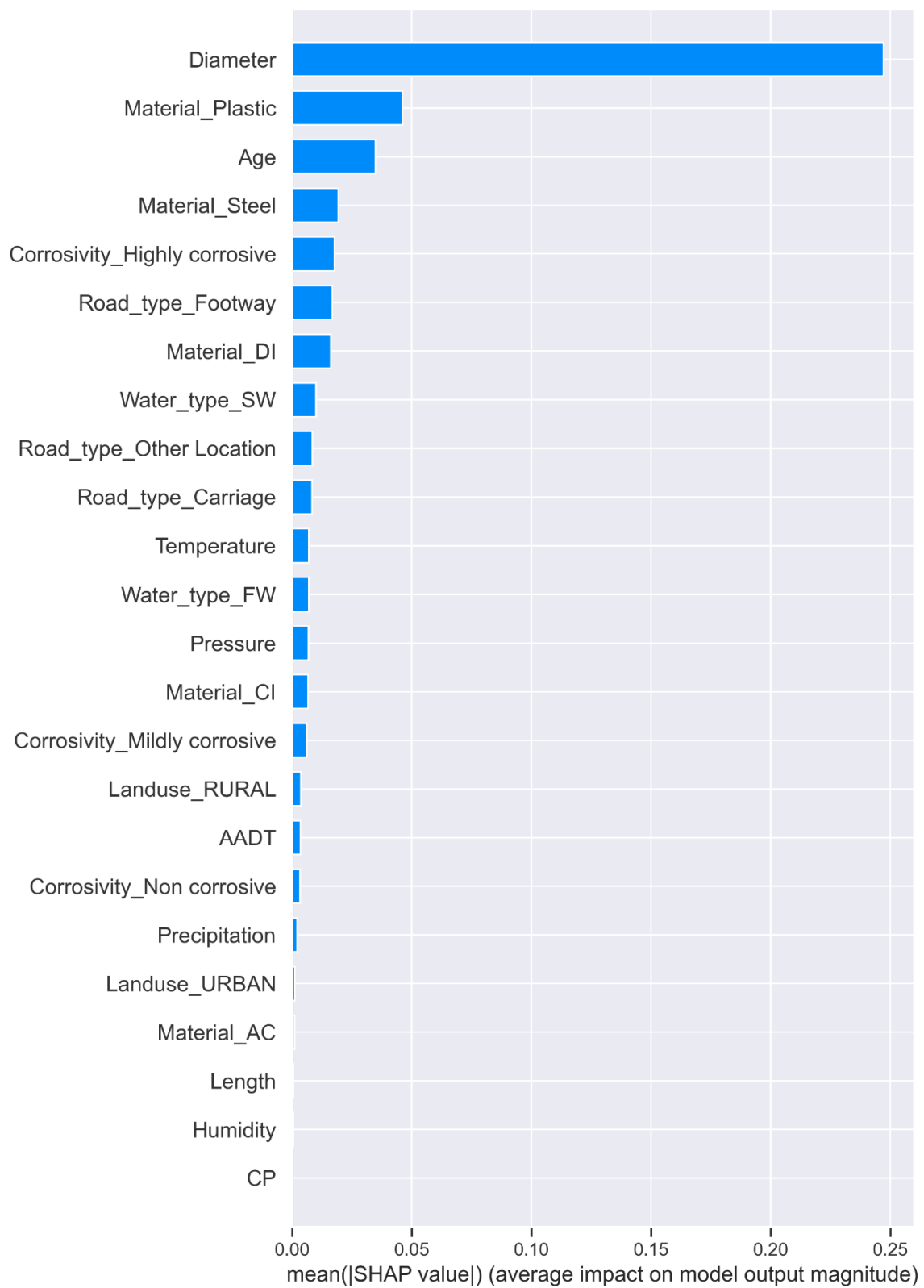
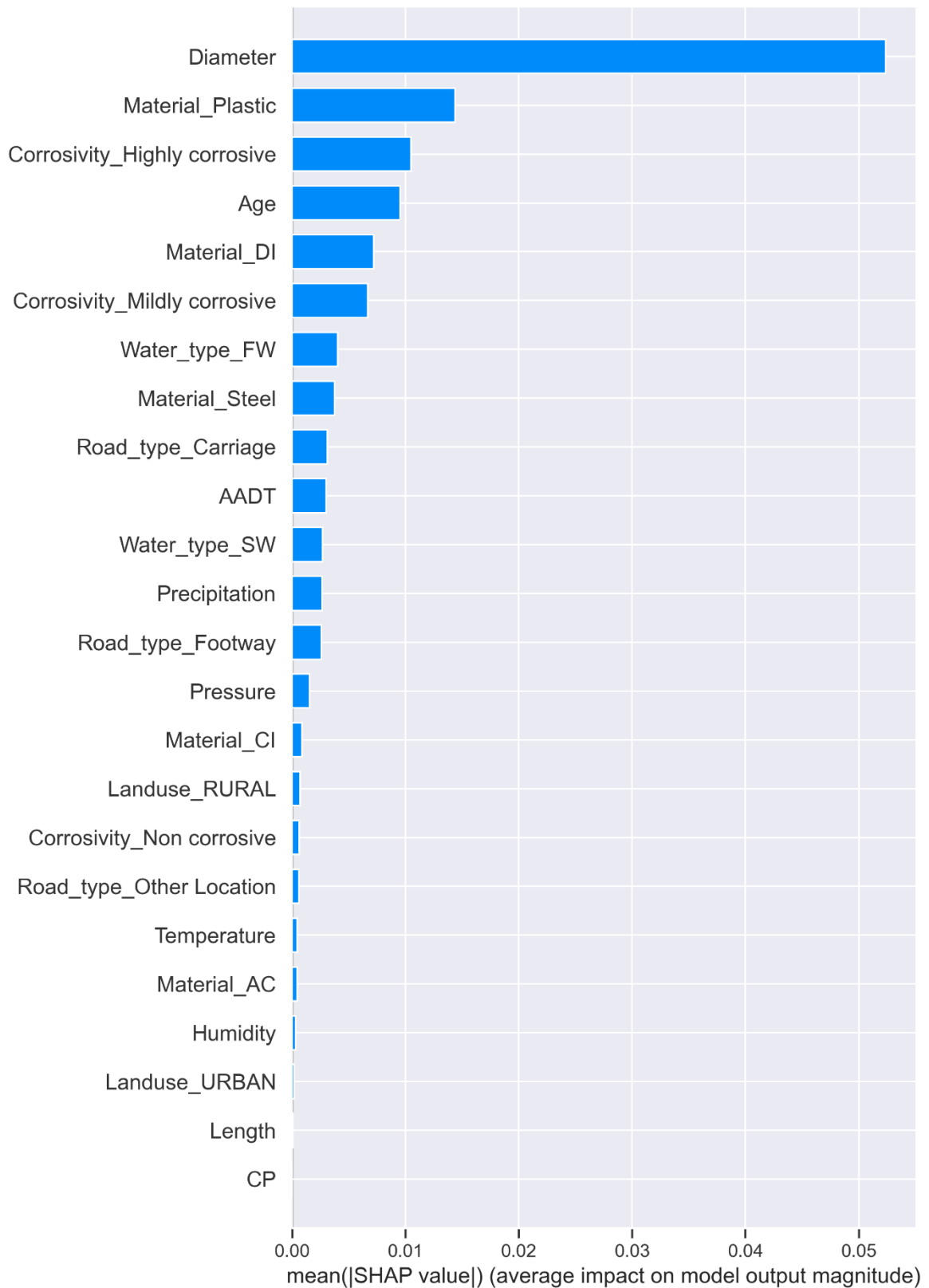Figure 5. 13: SHAP feature importance for the probability of leak model

Figure 5. 14: SHAP feature importance for the probability of burst model
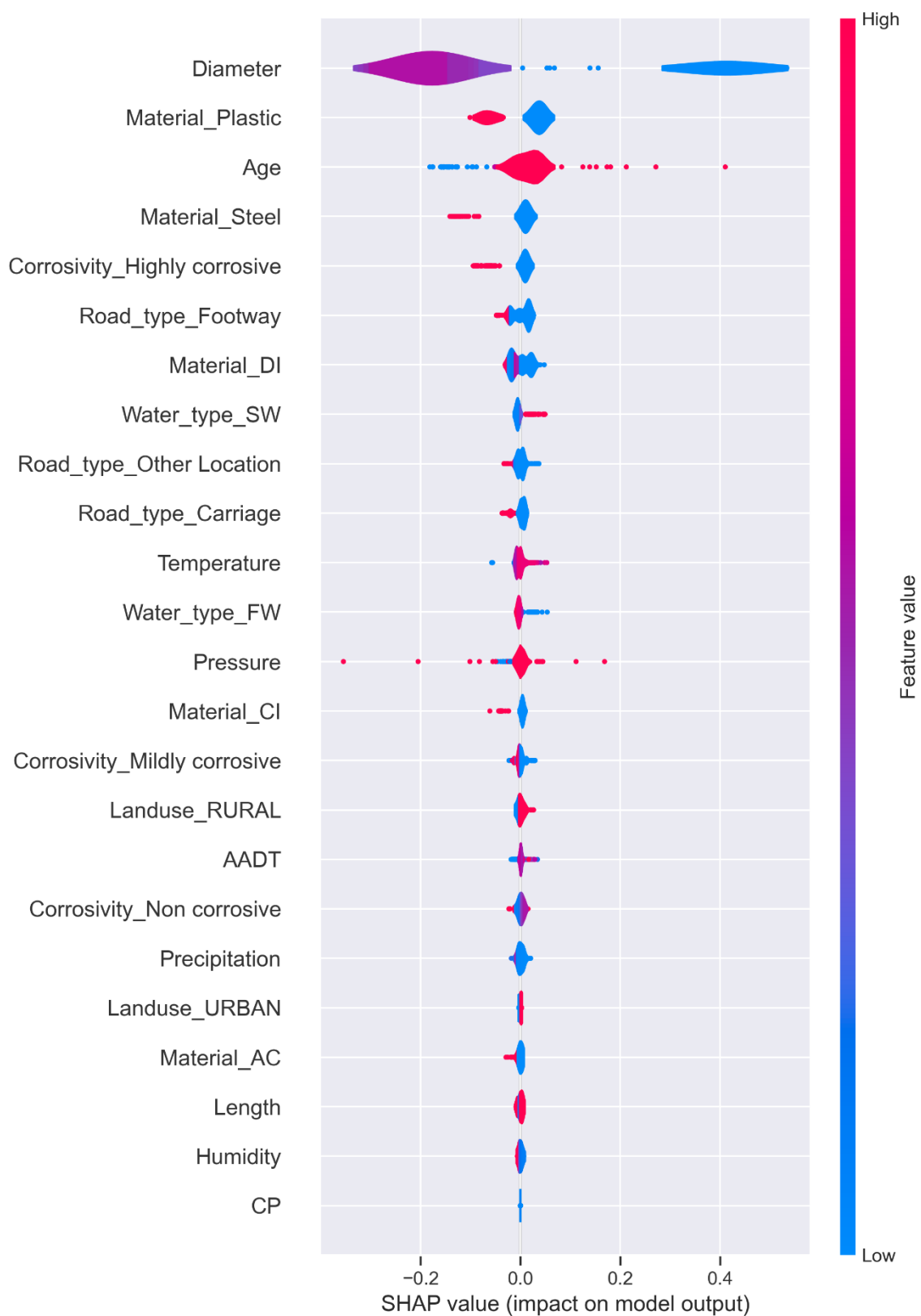
Figure 5. 15: Distribution of SHAP values for each feature for the probability of leak model
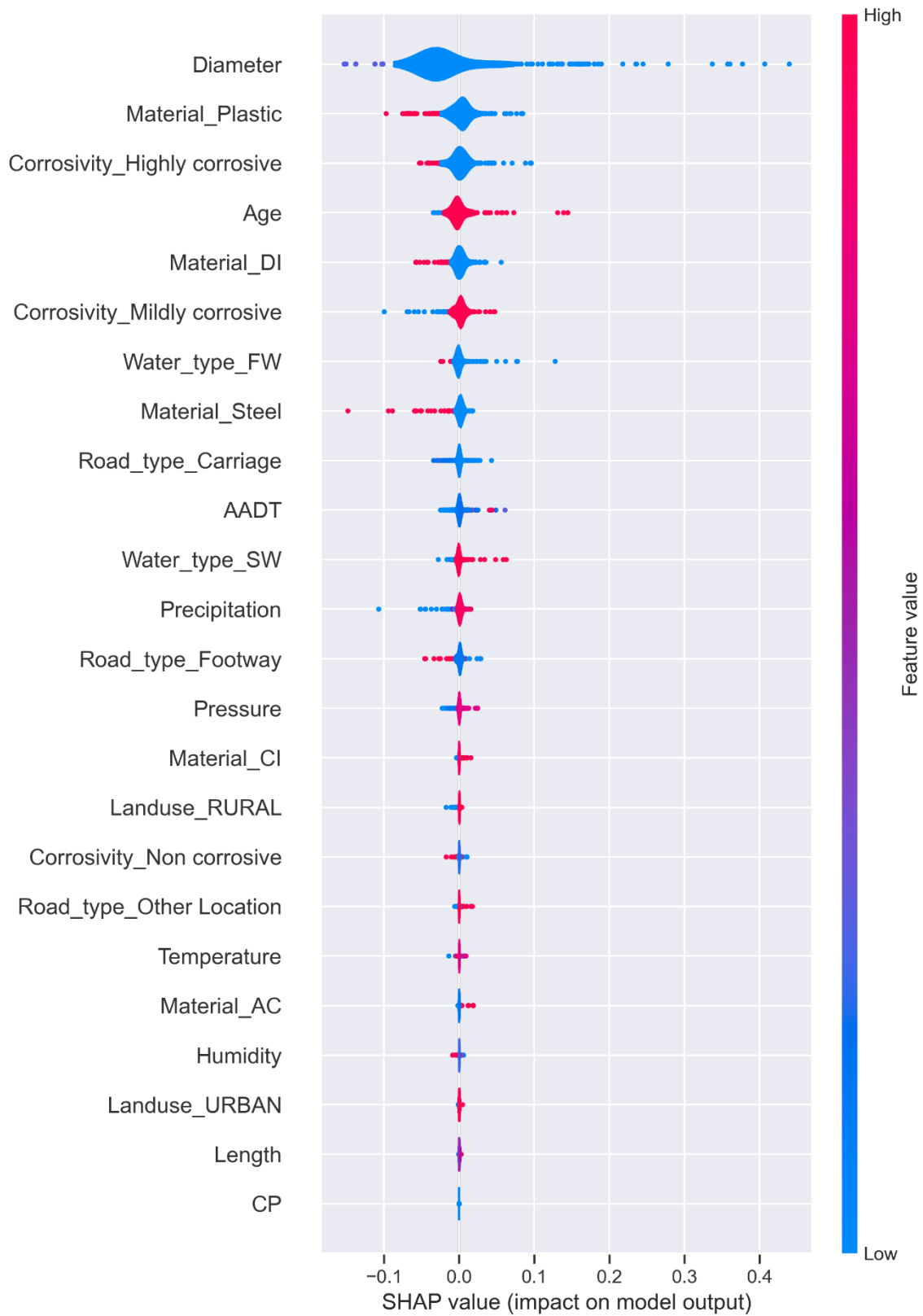
Figure 5. 16: Distribution of SHAP values for each feature for the probability of burst model

This approach ensures an unbiased assessment of the models' predictive capabilities on unseen data. Table 5.20 presents the performance metrics obtained by these models. The individual class performance metrics are presented in Table E1, while their confusion matrices are denoted in Figures E1 (a-e).

According to Table 5.20, LightGBM and AdaBoost achieved the highest accuracy of 0.78, followed by XGBoost and RF with an accuracy of 0.77, and CatBoost with an accuracy of 0.76. In terms of macro-averaged metrics, which treat all classes equally without considering class imbalance, LightGBM exhibited the best performance with a macro precision of 0.57, macro recall of 0.56, and macro F1 score of 0.56. This is closely followed by AdaBoost and XGBoost, with a macro F1 score of 0.55. With respect to weighted metrics, which account for class imbalance by weighting the respective metric for each class by the number of actual instances in that class, LightGBM again emerged as the top performer. It achieved the highest weighted precision, weighted recall, and weighted F1 score of 0.78.

Regarding the average AUC, which represents the mean of the individual AUC scores for each class, both XGBoost and LightGBM performed equally well, with an average AUC of 0.84. It is noteworthy that LightGBM exhibited the fastest training time of 51.88 seconds, significantly outperforming the other models in terms of computational efficiency. The base LightGBM model demonstrated superior performance across most evaluation metrics, making it a promising candidate for predicting the causes of water pipe failures. However, it is essential to optimize these models further to potentially enhance their predictive capabilities.

5.6.2. Result and validation of the optimized-EL models

The performance of the optimized models was evaluated following hyperparameter tuning using the TPE algorithm using the validation dataset. Table 5.21 presents the performance metrics obtained by the optimized EL models on the testing dataset. The performance metrics

for each class are shown in Appendix E under Table E2, and their corresponding confusion

matrices are depicted in Figures E2 (a-e).

Table 5. 20: Performance metrics for the base models for predicting causes of water pipe failure

Models	Accuracy	Macro Precision	Macro Recall	Macro F1 score	Weighted Precision	Weighted Recall	Weighted F1 score	Avg AUC	Training time (s)
XGBoost	0.77	0.55	0.54	0.55	0.77	0.77	0.77	**0.84**	591.04
AdaBoost	**0.78**	0.56	0.54	0.55	0.77	**0.78**	0.77	0.81	660.34
LightGBM	**0.78**	**0.57**	**0.56**	**0.56**	**0.78**	**0.78**	**0.78**	**0.84**	**51.88**
CatBoost	0.76	0.52	0.53	0.52	0.76	0.76	0.76	0.83	1077.89
RF	0.77	0.55	0.53	0.53	0.77	0.77	0.77	0.83	767.95

*the numbers in bold represent the best result

The results demonstrate that hyperparameter optimization led to consistent improvements in

model performance across all the evaluation metrics compared to the base models. Among the

optimized models, XGBoost + TPE attained the highest accuracy of 0.82, macro precision of

0.65, weighted precision of 0.80, weighted recall of 0.82, and weighted F1 score of 0.80. It also

achieved one of the best macro recalls of 0.58, macro F1 scores of 0.61, and average AUCs of

0.87. Compared to the base XGBoost model, optimization with TPE resulted in increases of

6.1% in accuracy, 15.4% in macro precision, 6.9% in macro recall, 9.8% in macro F1 score,

and 3.8% in weighted F1 score.

LightGBM + TPE emerged as the model with one of the highest average AUC of 0.87 post-

optimization. It also attained the second-best accuracy of 0.81 and a weighted F1 score of 0.80.

Optimization enhanced LightGBM's performance considerably, increasing its accuracy by

3.7%, macro precision by 9.5%, macro F1 score by 6.6%, and weighted F1 score by 2.5%

compared to the base model. The other optimized models, AdaBoost + TPE, CatBoost + TPE

and RF + TPE, also showed gains over their respective base versions across the evaluation

metrics. For instance, RF + TPE achieved 4.9% higher accuracy, 12.7% higher macro precision,

7.0% higher macro recall, and 3.8% higher weighted F1 score compared to the base RF model.

These results highlight the effectiveness of hyperparameter tuning in extracting the maximum potential from the models. Figures 5.17 and 5.18 visualize the differences between the base and optimized models, highlighting the effectiveness of TPE algorithm in finding near-optimal hyperparameter values to improve the models' predictive capabilities. Figures 5.19 to 5.23 also show the AUC curves of each model, highlighting the difference between the optimized and base models across the four predicted classes.

Table 5. 21: Performance metrics for the optimized models for predicting causes of water pipe failure

Models	Accuracy	Macro Precision	Macro Recall	Macro F1 score	Weighted Precision	Weighted Recall	Weighted F1 score	Avg AUC	Training time (s)
XGBoost + TPE	**0.82**	**0.65**	**0.58**	**0.61**	**0.80**	**0.82**	**0.80**	**0.87**	8468.94
AdaBoost + TPE	0.79	0.58	0.54	0.56	0.77	0.79	0.78	0.82	73235.69
LightGBM + TPE	0.81	0.63	0.57	0.60	0.79	0.81	**0.80**	**0.87**	**985.16**
CatBoost + TPE	0.81	0.63	**0.58**	0.60	0.79	0.81	**0.80**	**0.87**	9924.26
RF + TPE	0.81	0.63	0.57	0.59	0.79	0.81	**0.80**	0.86	6694.57

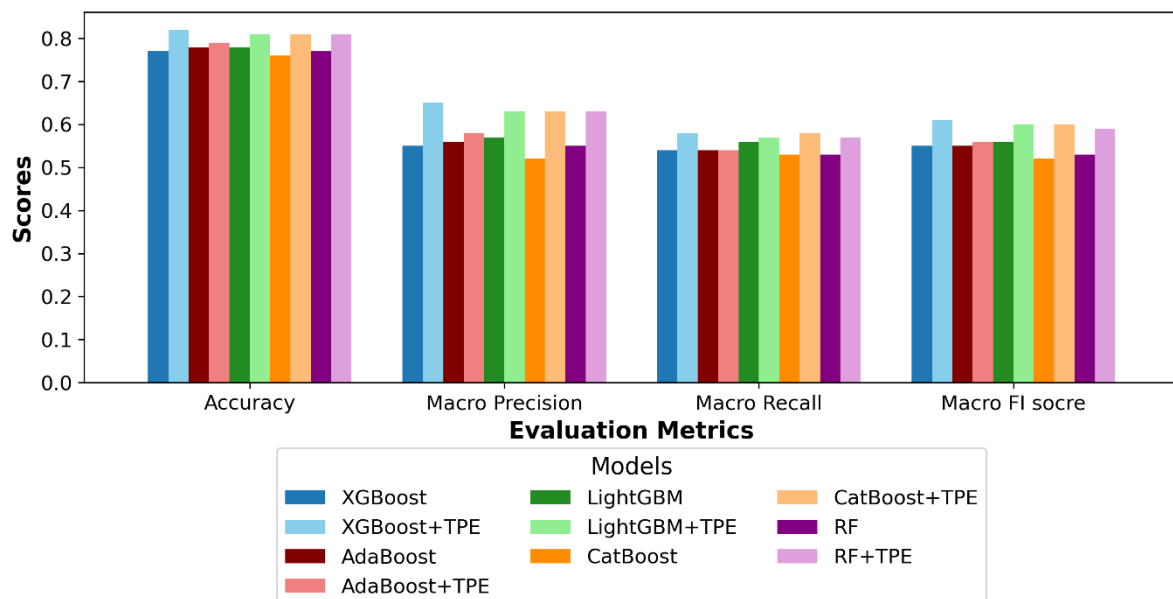*the numbers in bold represent the best result

Figure 5. 17: Comparison of the base and optimized models using accuracy, macro precision, macro recall, and macro F1 score
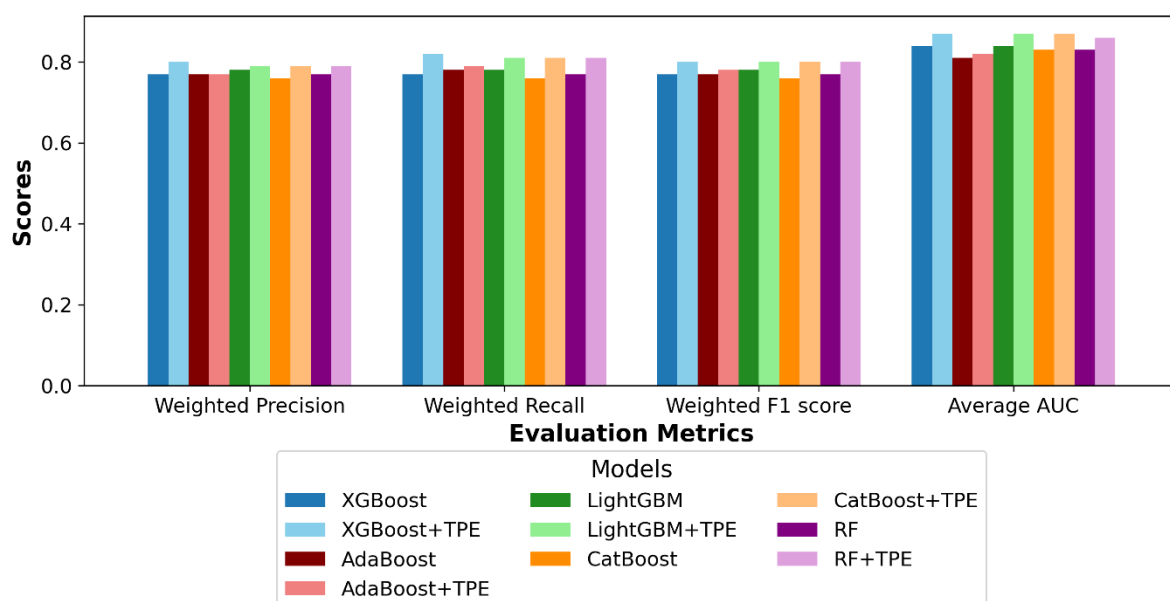
Figure 5. 18: Comparison of the base and optimized models using weighted precision, weighted recall, weighted F1 score, and average AUC
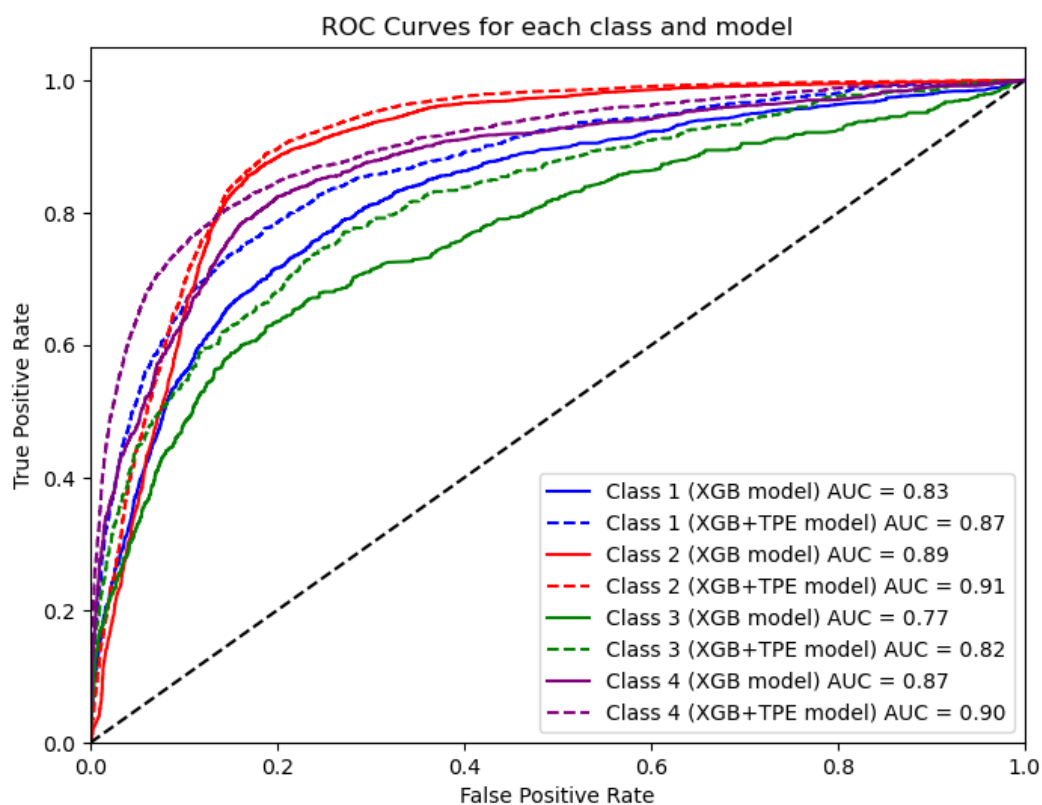
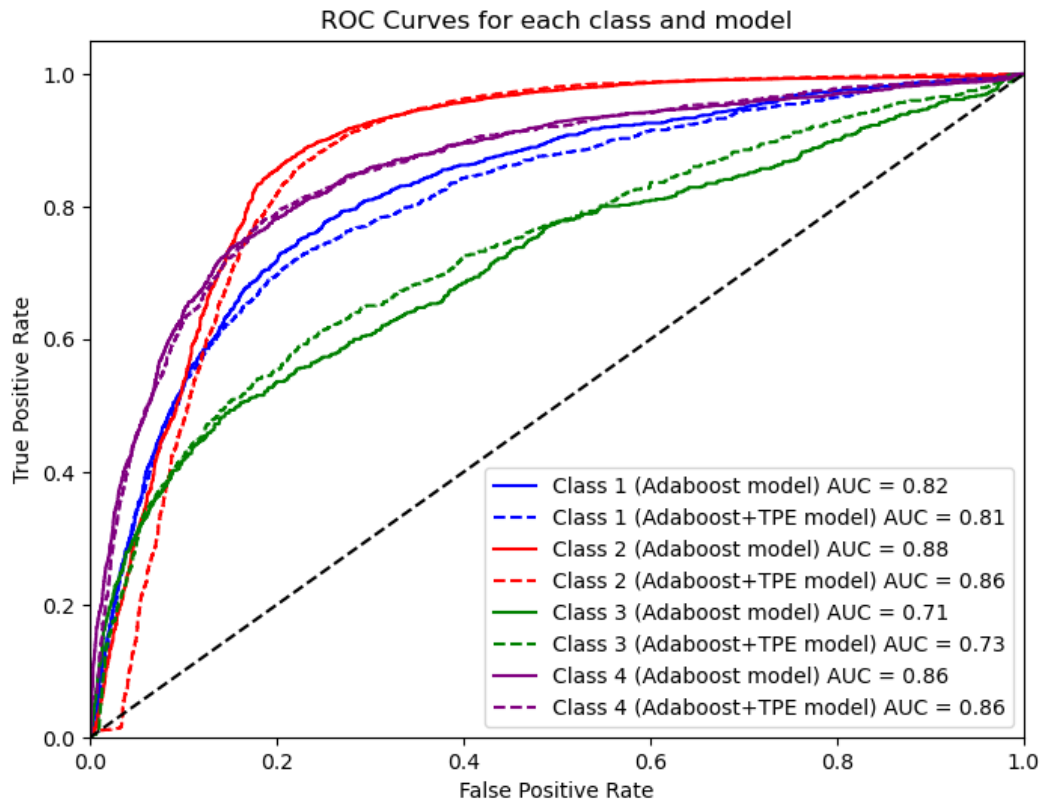

Figure 5. 19: ROC Curves for XGBoost-based models

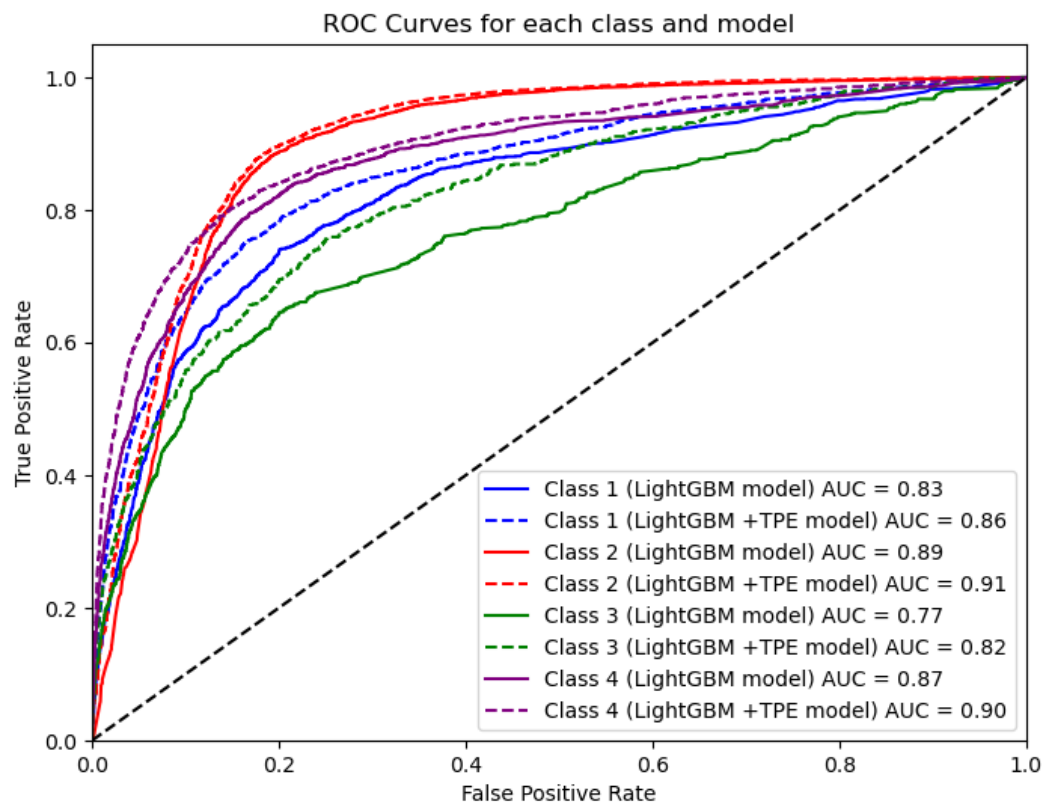Figure 5. 20: ROC Curves for AdaBoost-based models

Figure 5. 21: ROC Curves for LightGBM-based models
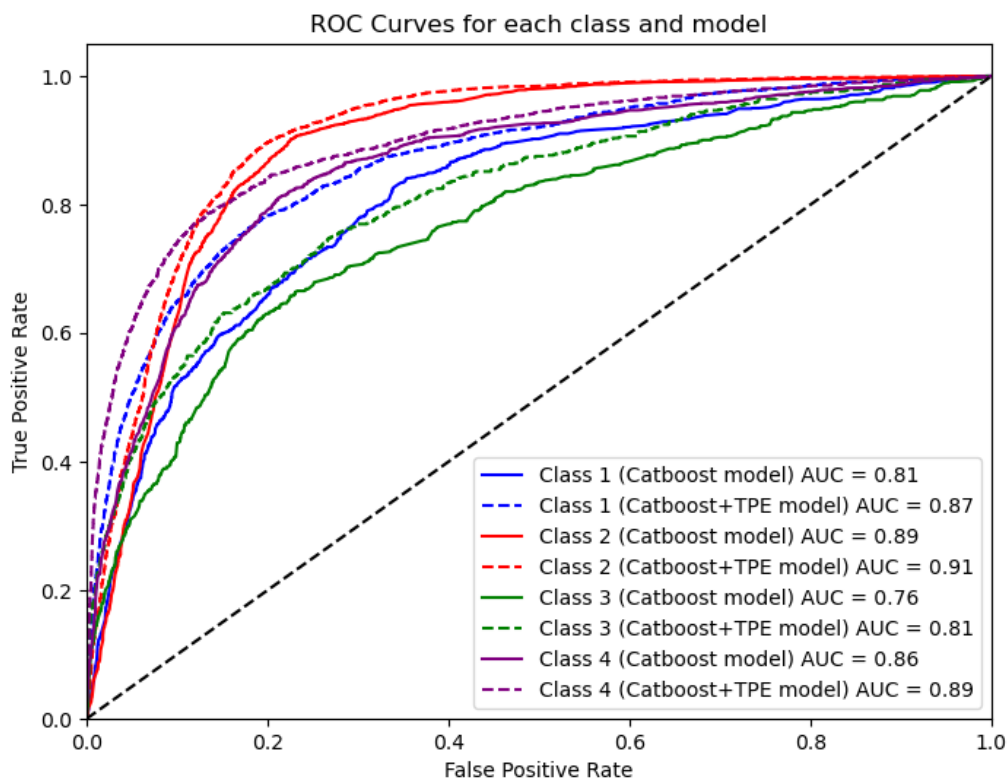
Figure 5. 22: ROC Curves for CatBoost-based models

5.6.3. Selection of the best-EL model

Model selection is an essential step in developing and deploying reliable predictive models. Relying solely on individual performance metrics may lead to biased or inconsistent conclusions, as different metrics often highlight varying aspects of a model's performance. To overcome this, the Copeland algorithm was implemented on the result presented in Table 5.21 Table 5.22 presents the results of the Copeland algorithm, which evaluates each model's performance against every other model based on the 9 metrics. The algorithm assigns Copeland points to each model based on its relative performance in pairwise comparisons, with the model having the highest Copeland points being ranked as the best-performing model.

As shown in Table 5.22, the optimized XGBoost model emerged as the top-ranked model, accumulating 26 Copeland points and achieving 4 wins without any losses.
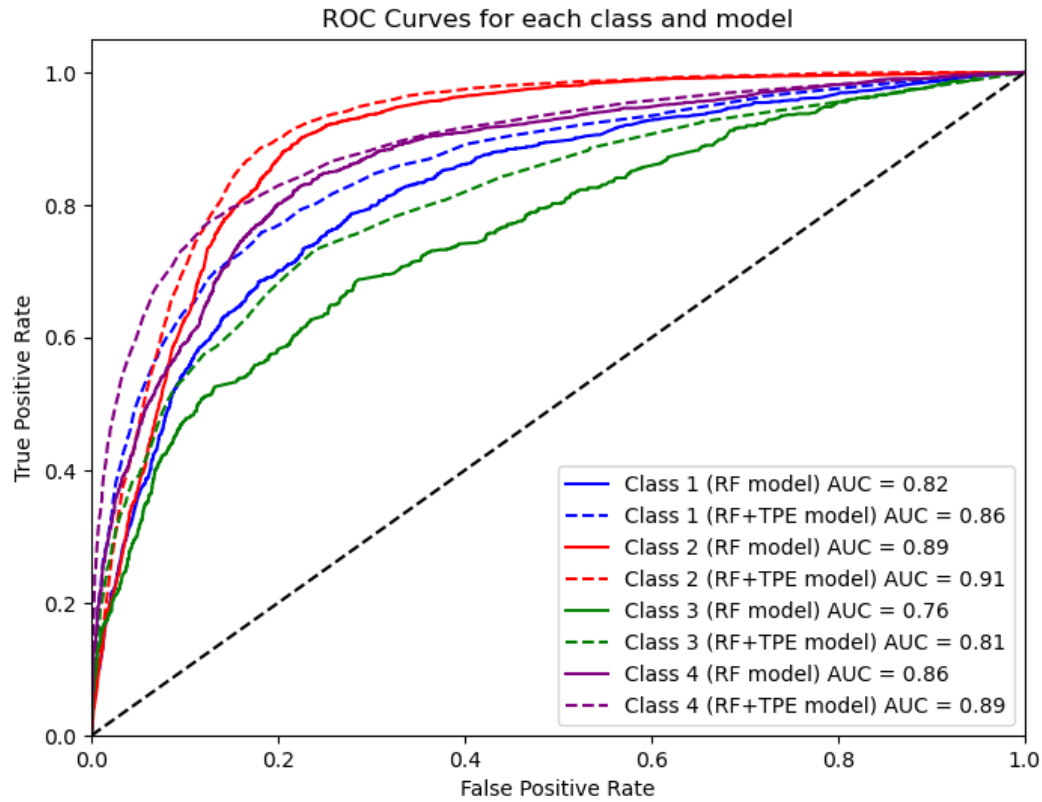
Figure 5. 23: ROC Curves for RF-based models

This indicates that the XGBoost model outperformed all other models in the pairwise comparisons, consistently demonstrating superior performance across the evaluated metrics. The LightGBM and CatBoost models secured the second and third ranks, respectively, with Copeland points of 7 and 5, and equal wins and losses. The RF model was ranked fourth, with -2 Copeland points and 1 win against 3 losses. Similarly, the AdaBoost model ranked last, with -36 Copeland points and no wins against 4 losses, suggesting its relative underperformance compared to the other models.

This systematic ranking approach will allow water utilities to reliably select the best performing model, optimized XGBoost for this case, for model deployment, ensuring accurate prediction of water pipe failure causes and enabling informed decision-making processes for proactive maintenance and replacement strategies.

Table 5. 22: Result of the Copeland algorithm

Models	Copeland point	Wins	Losses	Rank
XGBoost + TPE	26	4	0	1
AdaBoost + TPE	-36	0	4	5
LightGBM + TPE	7	2	1	2
CatBoost + TPE	5	2	1	3
RF + TPE	-2	1	3	4

5.6.4. Interpretability of the best-EL model

The interpretability of ML models is essential for practical applications, as it provides trust and actionable insights into the predictions. The SHAP framework was utilized to enhance the interpretability of the optimized XGBoost model. Figure 5.24 shows the mean SHAP values, representing the global feature importance averaged across all instances. Water type (saltwater), material (GIL), age, traffic, and diameter emerged as the five most important features affecting the prediction of causes of water pipe failure. The prominence of saltwater aligns with domain knowledge, as saltwater is more corrosive than freshwater and can accelerate pipe deterioration (Mohammed Abdelkader et al., 2024). GIL pipes are also susceptible to corrosion failures over time. The importance of pipe age is justified, as older pipes experience more wear and tear, increasing failure risk (Robles-velasco et al., 2020). Higher traffic levels suggest that external loading from nearby vehicular movement can contribute to pipe failures. Larger diameter mains are typically placed under carriageways with heavy traffic, exposing them to additional dynamic loads. The influence of these five key features highlights the multifaceted nature of pipe failure mechanisms. It can also be observed from Figure 5.24 that class 2 (corrosion) contributes most to the feature importance of the variables. This is justified as corrosion is known to be a major cause of water pipe failures (Taiwo et al., 2023b).

Figure 5.25 depicts the inherent feature importance from the optimized XGBoost model, showing water type (saltwater), material (GIL), age, material (S), and diameter as the top features, while material (AC), relative humidity, precipitation, landuse (rural) and material (SS) as the least contributing factors. Comparing this with the SHAP feature importance in Figure 16, there is a significant overlap - 4 of the top 5 features (water type, material GIL, age, diameter) and 3 of the bottom 5 features (material AC, landuse rural, material SS) are common. This substantial agreement between the two feature importance estimations provides robust confirmation of the most and least influential variables. The collective prominence of age, diameter, GIL material, and saltwater type implies that older, larger-diameter GIL pipes carrying saltwater are at the highest risk of failure. In contrast, AC pipes in rural areas with low humidity and precipitation are least likely to fail. These explanatory insights empower asset managers to focus inspection and maintenance efforts on vulnerable subsets of pipes based on critical features like age and material.

Figure 5.26 displays the distribution of SHAP values across different features for class 2, providing insight into how each feature influences the model's prediction of corrosion. The features are listed on the y-axis, and the SHAP values are shown on the x-axis. Each violin represents the distribution of SHAP values for a feature, with wider sections indicating a higher density of data points. A SHAP value of zero means the feature does not influence the prediction for that particular data point; positive values indicate an increased likelihood of the model predicting corrosion, while negative values suggest a decreased likelihood. The color transition in the figure from red to blue depicts the range of feature values within the data, where red highlights higher feature values and blue signifies lower ones. It can be discerned that smaller diameters, as represented by the blue shades, align with positive SHAP values. This association indicates that smaller pipe diameters are likely to have a greater influence on the model's

prediction of corrosion, potentially due to factors such as increased water velocity or pressure that may exacerbate the corrosive effects on the pipe material (Farh et al., 2023).

From the top of the plot, the 'Diameter', 'Material_GIL', 'Pressure', and 'Water_type_Saltwater' features show a wider distribution towards the higher SHAP value range, signifying a strong positive impact on predicting corrosion. This suggests that pipe diameters, the use of GIL, higher pressure within the pipes, and exposure to saltwater are influential factors in the model's decision-making process for corrosion. The presence of these conditions may be indicative of environments or operational states that are more conducive to corrosive processes, thus making them critical points of consideration for predictive maintenance and failure prevention strategies.
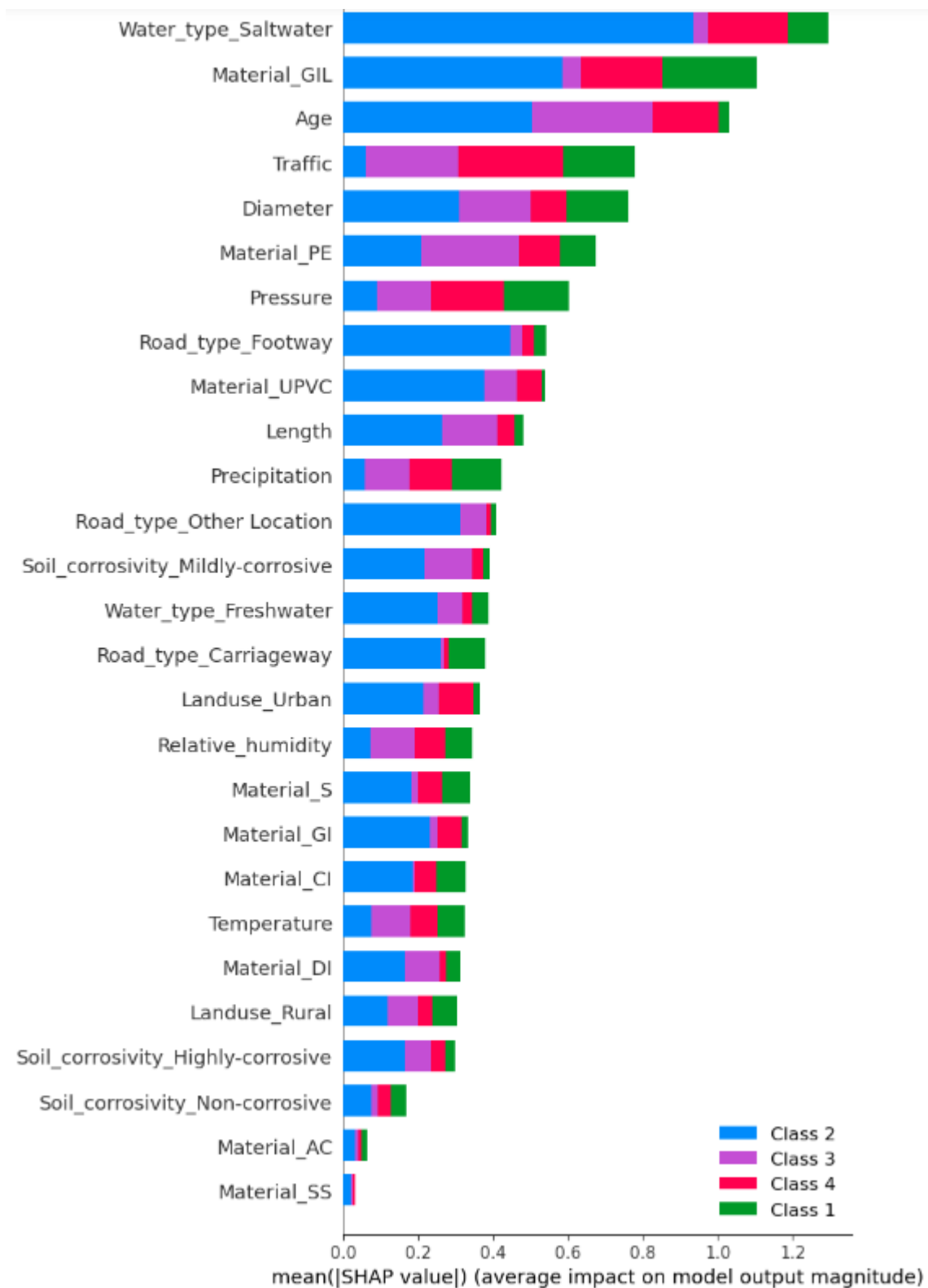
Figure 5. 24: Feature importance of the selected model using SHAP
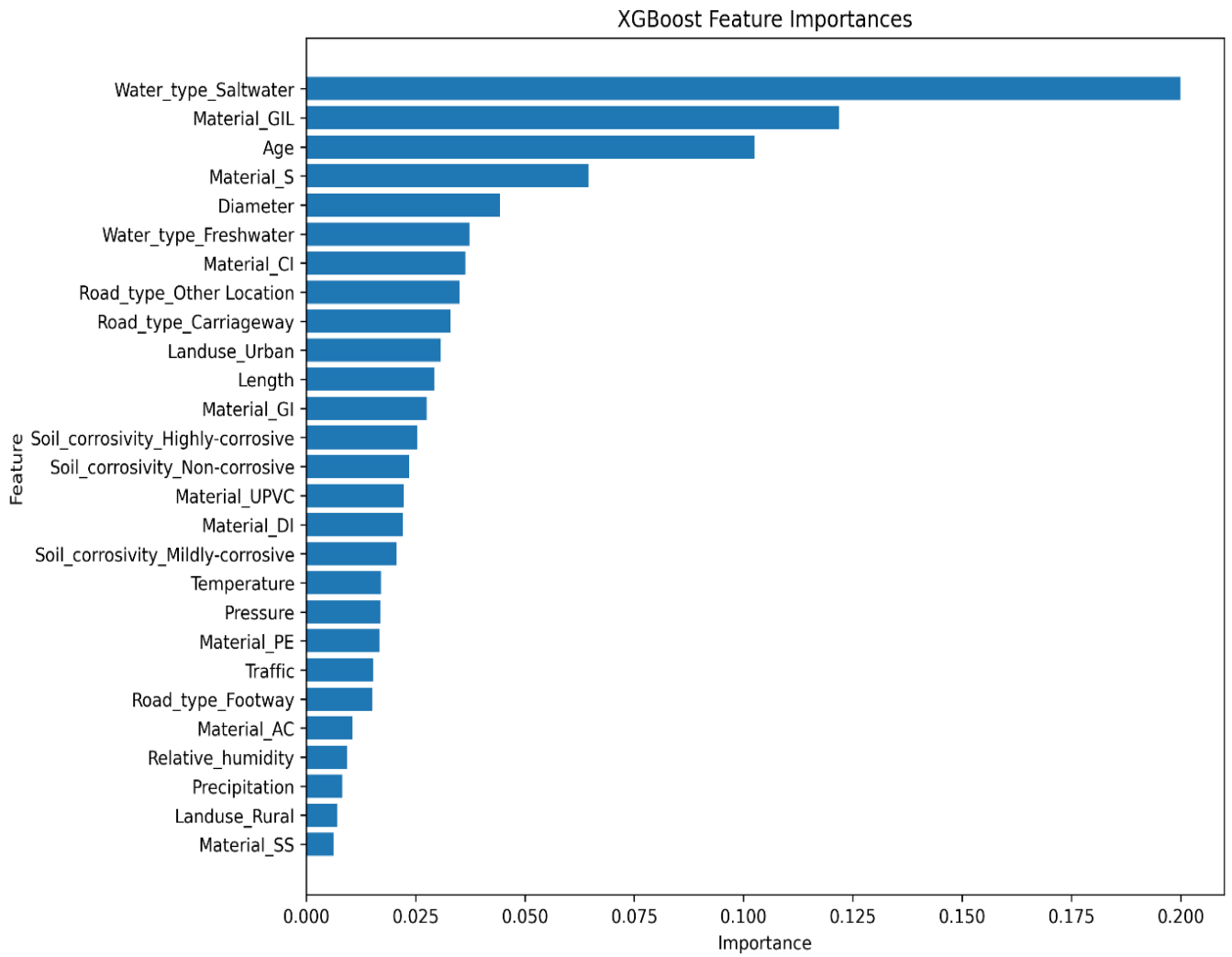
Figure 5. 25: Inherent feature importance of the selected model

Figure 5. 26: Distribution of SHAP values for the selected model for Class 2

5.7.   SUMMARY

This chapter presents the findings from the systematic review and the models developed within

this study. The systematic review highlights 30 factors that influence water pipe failures and

identifies 5 distinct failure modes. These factors and modes are depicted using FTL to enhance

visualization and understanding. The chapter then details the results from the implementation

of the PLS-SEM model, covering both measurement and structural models. Additionally, the results of the IPMA are discussed to demonstrate the significance and effectiveness of the included constructs. Furthermore, the chapter presents and discusses the result of the predictive models for calculating the probability of failure of individual pipes in HK WDN. The results of the POL and POB models are also presented, including the outcome of the base and optimized DL models. The chapter also includes the results of the base and optimized EL models for predicting COF. The selection and interpretability of the best performing EL model are discussed.

Chapter 6

AUTOMATED APPLICATIONS[5]

6.1. INTRODUCTION

This chapter details the development of four web-based applications designed to facilitate the straightforward implementation of the methodologies proposed in this study. The applications are specifically tailored for the POF, POL, POB, and COF models. These applications are developed using Python and its associated libraries for backend operations and HTML and CSS for frontend design, ensuring both functionality and user-friendly interfaces. This chapter discusses the development process of each application, illustrating the steps taken from initial design to final implementation. Additionally, the chapter provides an overview of the graphic user interface (GUI) for each model, showcasing how users can interact with the functionalities provided to efficiently apply the predictive models in real-world scenarios.

---

[5] This chapter is largely based upon:

Taiwo, R., Zayed, T. & Ben Seghier, M. E. A. (2024). " Integrated intelligent models for predicting water pipe failure probability". Alexandria Engineering Journal, 86, 243-257, https://doi.org/10.1016/j.aej.2023.11.047

Taiwo, R., Yussif, A., Ben Seghier, M. E. A., & Zayed, T. (2024). "Explainable Ensemble Models for Predicting Wall Thickness Loss of Water Pipes". Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2024.102630

Taiwo, R., Zayed, T. & Adey, B.T. "Explainable deep learning models for predicting water pipe failure." Journal of Environmental Management (IF = 8.7, Q1). (Under review – 1st cycle)

Taiwo, R., Zayed, T. & Adey, B.T. "Interpretable ensemble models for predicting causes of water pipe failure." Reliability Engineering and System Safety (IF = 8.1, Q1). (Under review – 1st cycle)

Taiwo, R., Zayed, T. Elshaboury, N., & Abdelkader, E. M. "Promoting Sustainable Water Distribution Networks: Modeling of Water Pipe Failure Factors and Modes." Cleaner Engineering and Technology (IF = 5.3, Q1). (Under review – 2nd cycle)

6.2. POF MODEL

6.2.1. Deployment process of the POF model

The deployment process of the POF model is detailed in this section, following the evaluation of the model's performance outlined in Section 5.4. From the results presented, the most effective model emerged from Experiment 2, where hyperparameters and input variables were optimized to enhance performance. This model, demonstrating superior accuracy and efficiency, was selected for deployment.

The deployment began with the encapsulation of the ML model into a backend service. This service was developed using the Python Flask library, which is adept at handling both data processing and model inference tasks. This backend component is crucial as it processes input data and computes the model's predictions. To create a seamless user experience, a frontend interface was developed using HTML and CSS. This interface is designed to be intuitive, allowing users to easily interact with the model by inputting data and receiving predictions. The frontend connects to the backend, ensuring that data flows smoothly between the user interface and the processing service.

The deployment infrastructure was streamlined by hosting the model on the Render Cloud server (Render, 2023). This cloud-based approach eliminates the need for a dedicated database on our premises, simplifying the management of user inference on the model. Testing was carried out to ensure the application's robustness, confirming its reliability across various user scenarios. This included handling simultaneous access from multiple users and accommodating variations in data input, thus guaranteeing consistent performance under diverse conditions.

This systematic approach to deployment ensures that the POF model is not only effective in its predictive capabilities but also robust and user-friendly in practical application scenarios.

6.2.2. **POF Model Inputs**

Figure 6.1 shows the Graphic User Interface (GUI) for the POF model inputs, which is designed for user-friendliness and efficient interaction. The screen of the GUI is structured to facilitate easy and accurate data entry, specifically tailored to gather necessary information for predicting water pipe failure probabilities. The GUI is divided into two main sections for data input: Pipe-related factors and Environment-related factors.

(I) Pipe-related Factors:

Length (m): Users can enter the length of the pipe.

Diameter (mm): Users can specify the pipe's diameter.

Age (years): The age of the pipe can be entered directly.

Material: A dropdown menu allows users to choose the material of the pipe, such as PE (Polyethylene).

(II) Environment-related Factors:

Temperature (°C): Users input the ambient temperature.

Precipitation (mm): This field is for entering the amount of precipitation.

Corrosivity: Users can select the level of corrosivity from options like 'Mildly corrosive'.

Traffic (AADT): The average annual daily traffic can be entered, affecting the stress on the pipe.

6.2.3. **POF Model Output**

At the bottom of these input fields, a "Predict" button is placed (see Figure 6.2). When clicked, this button initiates the calculation, feeding the input data into the backend model to compute the probability of failure.

Figure 6. 1: GUI of POF Model Inputs

The results are then displayed to the user, allowing them to understand the risk associated with the specified conditions. The layout is designed with clarity and simplicity in mind, ensuring that users of all technical levels can easily navigate and utilize the application without confusion. While Figure 6.2 doesn't show actual prediction results (i.e., values) due to data confidentiality, the output includes the pipe's failure probability and reliability, which sum to 1, representing the complete spectrum of the pipe's condition.
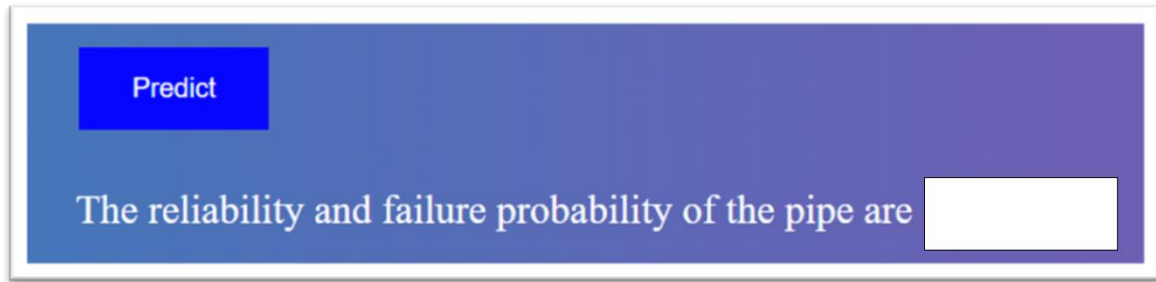
Figure 6. 2: GUI of POF Model Output

6.3. POL AND POB MODELS

6.3.1. Deployment process of the POL and POB models

After determining the optimized CNN as the most effective model based on the evaluations in Section 5.5.3, the next step was to deploy this model for predicting both POL and POB. The CNN model was configured to use standardized data for leak predictions and normalized data for burst predictions.

To ensure a user-friendly and accessible deployment, the Streamlit framework was employed (Mhadbi, 2021). Streamlit allows for the development of interactive web applications with ease, providing a seamless platform for users to input data and receive predictions. The deployment process involved several key steps. First, the trained CNN model was integrated into the Streamlit application, ensuring efficient data flow from user input to model predictions. The application architecture was designed to handle user queries reliably, minimizing any potential downtime.

Next, the web application was split into two separate pages, one dedicated to POL predictions and the other to POB predictions (see Figure 6.3). This separation enhanced the clarity and usability of the application, allowing users to focus on their specific prediction requirements without unnecessary clutter.

195

To make the application accessible to HK WSD, it was deployed to the Streamlit Cloud platform. This cloud deployment eliminated the need for users to run the application locally, providing a convenient web-based interface accessible from any device with an internet connection. Extensive testing was conducted to verify the application's performance, reliability, and responsiveness under various conditions. The deployment strategy prioritized high availability and a seamless user experience, ensuring that the application could handle multiple concurrent users without compromising prediction accuracy or speed.

6.3.2. **POL and POB Model Inputs**

The GUI for the POL and POB models' inputs is carefully designed to ensure ease of use and effective interaction for both models (see Figures 6.3). This unified GUI facilitates user engagement and data input for both leak and burst predictions, providing a coherent experience across both applications. The interface offers two distinct methods for data input: individual prediction via the sidebar and batch prediction through file upload.

For individual predictions, the GUI is structured with a sidebar dedicated to inputting a wide range of water pipe attributes and environmental factors. Users can input parameters such as the material type of the pipe, diameter, length, and additional environmental conditions including temperature, humidity, and precipitation. These inputs are essential as they directly influence the predictive outcomes of the models. Interactive elements like sliders and dropdown menus allow for precise and hassle-free entry of these variables. For batch predictions, the system provides a file upload functionality that enables users to process multiple pipe cases simultaneously. Users can upload their own dataset in either CSV or Excel format, containing the same parameters as the individual input method. The upload interface requires specific column naming conventions and data formats to ensure compatibility with the model. This batch processing capability is particularly valuable for utilities and organizations

that need to analyze multiple pipe segments efficiently. The system accommodates two options for batch processing: users can either utilize a provided test dataset or upload their own structured data file, making the tool adaptable to various organizational needs and data management practices.

6.3.3. **POL and POB Model Output**

Figures 6.4 and 6.5 show the GUI for POL and POB model output. At the upper part of the GUI is an informative display that not only enhances the interface aesthetically but also educates users about the deep learning processes involved in the leakage and burst predictions. This part includes detailed lists of the features used for model training, which are crucial for users to understand the variables impacting the model's predictions. Features such as pipe age, traffic levels, and soil corrosivity are among the inputs that help predict the likelihood of a leak or burst.

The model output is presented in two distinct formats, corresponding to the two input methods available. For individual predictions, users can initiate the model's prediction by clicking the "Predict Probability of Leakage" or "Predict Probability of Burst" button, depending on the model being accessed. The results are displayed immediately at the lower part of the interface, providing instant feedback on the risk assessment based on the provided data. For batch predictions, the output is presented in a tabular format that includes all the input parameters alongside their corresponding predictions. Users can preview these results directly within the interface, where the original dataset is augmented with an additional column containing the model's predictions. To facilitate further analysis and record-keeping, the system provides a download option, allowing users to export the complete results as a CSV file. This feature is particularly valuable for organizations conducting large-scale assessments of their pipeline networks.

Water Pipe Attributes

Select the **Cathodic Protection** Status of the Pipe

Absent

Enter the **Length** of the Pipe in Meters

10.00

Enter the **Diameter** of the Pipe in Millimeters

300.00

Enter the **Pressure** exerted on the Pipe in Bars

7.00

Enter the **Age** of the Pipe in Years

25.00

Enter the **Traffic** associated with the Pipe in AADT

1500.00

Enter the **Temperature** associated with the Pipe in °C

25.00

Enter the **Relative Humidity** associated with the Pipe

40.00

Enter the **Precipitation** associated with the Pipe

10.00

Select **Material Type** of the Pipe

DI

Select **Soil Corrosivity** associated with the Pipe

Highly corrosive

Select **Road type** associated with the Pipe

Carriage

Select **Water Type** carried by the Pipe

Freshwater

Select **Land use** associated with the Pipe

Rural

Predict Probability of Leakage

Water Pipe Attributes

Select the **Cathodic Protection** Status of the Pipe

Present

Enter the **Length** of the Pipe in Meters

15.00

Enter the **Diameter** of the Pipe in Millimeters

400.00

Enter the **Pressure** exerted on the Pipe in Bars

5.00

Enter the **Age** of the Pipe in Years

40.00

Enter the **Traffic** associated with the Pipe in AADT

789.00

Enter the **Temperature** associated with the Pipe in °C

16.00

Enter the **Relative Humidity** associated with the Pipe

68.00

Enter the **Precipitation** associated with the Pipe

13.00

Select **Material Type** of the Pipe

PE

Select **Soil Corrosivity** associated with the Pipe

Mildly corrosive

Select **Road type** associated with the Pipe

Footway

Select **Water Type** carried by the Pipe

Saltwater

Select **Land use** associated with the Pipe

Urban

Predict Probability of Burst

Figure 6. 3: GUI of POL Model Inputs (Left) and POB Model Inputs (Right)

198

It is important to note that the actual prediction results in Figures 6.4 and 6.5 are blurred due to the confidential nature of the network data. This measure ensures the protection of sensitive information while still allowing the functionality of the models and web applications to be demonstrated through the GUI. The same confidentiality measures are applied to the batch prediction outputs, maintaining data security while demonstrating the applications' capability to global stakeholders in WDN. This approach provides a comprehensive yet secure way to utilize the predictive capabilities of the models, whether for single pipe segments or extensive network analyses.

6.4. COF MODEL

6.4.1. Deployment process of the COF model

The deployment process for the COF model followed a similar approach to that of the POL and POB models. The optimized ensemble learning model -XGBoost + TPE, identified as the best-performing model in Section 5.6.3, was selected for deployment. This model excelled in predicting the COF based on various pipe attributes and environmental factors.

The Streamlit framework was once again utilized to create an interactive web application for the COF model. The application was designed to provide users with a user-friendly interface for inputting pipe characteristics and obtaining real-time predictions of the most likely COF.

The deployment process involved integrating the trained EL model into the Streamlit application, ensuring seamless data processing and prediction generation. The application architecture was optimized for performance, scalability, and reliability, allowing it to handle multiple user requests simultaneously.

Water Pipe Leakage Prediction App

Deep learning-based predictor for water pipe leakage

Deep Learning For Leakage Prediction

Single Prediction   Batch Prediction

Features Used For Model Training

1. **Cathodic Protection:** Present, Not-present
2. **Length:** Measured in Meters
3. **Diameter:** Measured in Millimeters
4. **Pressure:** Measured in Bars
5. **Age:** Measured in Years
6. **Traffic:** Measured in Annual Average Daily Traffic (AADT)
7. **Temperature:** Measured in °C
8. **Relative Humidity:** It has no unit
9. **Precipitation:** Measured in Millimeters
10. **Material:** Asbestos Cement (AC), Cast Iron (CI), Ductile Iron (DI), Galvanized Iron (GI), Lined Galvanized Iron (GIL), Polyethylene (PE), Steel (S), SS (Stainless Steel), UPVC (Unplasticized Polyvinyl Chloride)
11. **Corrosivity:** Highly corrosive, Mildly corrosive, Non corrosive
12. **Road type:** Carriage, Footway, Other Location
13. **Water type:** Freshwater (FW), Saltwater (SW)
14. **Land use:** Urban, Rural

The predicted probability of leakage is:

(a)

Batch Prediction from File

Choose your data source:

🔘 Use Test Dataset
⚪ Upload Own File

Preview of test data:

	CP	Length	Diameter	Pressure	Age	AADT	Temperature	Humidity	Precipitation	Mate
0	Absent	10	300	7	25	1,500	25	40	10	DI
1	Present	50	600	10	40	3,000	34	60	19	GI
2	Absent	200	450	8	44	5,000	28	45	76	PE

Predict for Test Data

Download Results as CSV

Results Preview:

	Precipitation	Material	Corrosivity	Road_type	Water_type	Landuse	Predicted_Leakage_
0	10	DI	Highly corrosive	Footway	Freshwater	Rural	
1	19	GI	Mildly corrosive	Carriage	Saltwater	Urrban	
2	76	PE	Non corrosive	Other Location	Freshwater	Urban	

See required column format   ⌄

(b)

Figure 6. 4: GUI of POL Model Output (a) – single prediction (b) batch prediction

Water Pipe Burst Prediction App

Deep learning-based predictor for water pipe burst

Deep Learning For Burst Prediction

Single Prediction   Batch Prediction

Features Used For Model Training

1. **Cathodic Protection:** Present, Not-present
2. **Length:** Measured in Meters
3. **Diameter:** Measured in Millimeters
4. **Pressure:** Measured in Bars
5. **Age:** Measured in Years
6. **Traffic:** Measured in Annual Average Daily Traffic (AADT)
7. **Temperature:** Measured in °C
8. **Relative Humidity:** It has no unit
9. **Precipitation:** Measured in Millimeters
10. **Material:** Asbestos Cement (AC), Cast Iron (CI), Ductile Iron (DI), Galvanized Iron (GI), Lined Galvanized Iron (GIL), Polyethylene (PE), Steel (S), SS (Stainless Steel), UPVC (Unplasticized Polyvinyl Chloride)
11. **Corrosivity:** Highly corrosive, Mildly corrosive, Non corrosive
12. **Road type:** Carriage, Footway, Other Location
13. **Water type:** Freshwater (FW), Saltwater (SW)
14. **Land use:** Urban, Rural

The predicted probability of burst is: [   ]

Figure 6. 5: GUI of POB Model Output (a) – single prediction (b) batch prediction

To enhance the user experience and maintain consistency with the POL and POB applications, the COF application was also deployed to the Streamlit Cloud. This cloud deployment provided easy access to the application through a web browser, eliminating the need for local installation or setup. Testing was conducted to validate the application's accuracy, responsiveness, and

stability under various usage scenarios. The deployment strategy aimed to deliver a high-quality, reliable tool for water utilities to identify and prioritize potential failure causes, enabling proactive maintenance and risk mitigation strategies.

6.4.2. COF Model Inputs

Similar to the GUI of POL and POB model inputs, the GUI of the COF model inputs was designed with user-friendliness and clarity in mind. It has two distinct methods for data input. For individual predictions (see Figure 6.6), the interface consists of a sidebar with input fields for each relevant pipe attribute, such as age, length, diameter, pressure, traffic, soil type, and others. Users can easily select or enter the appropriate values for their specific pipe using dropdown menus, sliders, or text input fields.

The interface also provides a batch prediction capability through its file upload functionality. Users can choose between utilizing a provided test dataset or uploading their own customized file in either CSV or Excel format. For user-uploaded files, the system requires specific column formatting and naming conventions to ensure proper data processing. This batch input feature is particularly beneficial for organizations needing to analyze multiple pipe segments simultaneously, streamlining the assessment process for larger datasets. Both input methods maintain consistent data requirements and validation checks, ensuring reliable model predictions regardless of the chosen input method.

6.4.3. COF Model Output

The GUI of the COF Model output includes clear labels and explanations for each attribute, ensuring that users understand the information used for the prediction. The output presentation varies based on the chosen input method. For individual predictions, users can initiate the prediction by clicking the "Predict the Cause of Failure" button, with results displayed immediately within the GUI as shown in Figure 6.7.

Figure 6. 6: GUI of COF Model Inputs

205

For batch predictions, the system generates a table that includes all input parameters alongside their corresponding predicted causes of failure. Users can preview these results directly in the interface and have the option to download the complete analysis as a CSV file. This feature facilitates efficient documentation and further analysis of multiple pipe segments simultaneously. The tabular output includes the original input parameters and adds a new column containing the predicted cause of failure for each pipe segment.

It is important to note that Figure 6.7 does not show the predicted COF due to the confidential nature of the network data. Instead, it illustrates the functionality and design of the GUI for both individual and batch predictions, ensuring the confidentiality of sensitive information while demonstrating the capabilities of the COF model effectively. This dual-output approach provides a versatile solution for both targeted analyses of individual pipes and comprehensive assessments of larger pipeline networks.

It should be noted that the development of separate web applications for POF, POL, POB, and COF prediction models was driven by their distinct operational requirements and input variables. While a unified interface might seem preferable, the technical architecture reflects the specialized nature of each prediction task. For instance, the POF model requires 8 optimized input variables, including pipe characteristics and environmental factors, while POL and POB models need 14 input variables, and the COF model requires 13 specific input parameters based on the optimization performed to achieve the best predictive capability. This separation enhances model accuracy and allows water utilities to utilize specific applications based on their immediate needs.

These applications serve distinct purposes in the decision-making process, with POF application providing overall failure probability assessment, POL and POB applications offering detailed prediction of specific failure types, and COF application assisting in root

cause analysis. While operating independently, they form a comprehensive suite of tools for pipe failure management. The individual deployment approach enables water utilities to perform focused analyses on specific pipes, supporting targeted maintenance decisions. This granular analysis at the individual pipe level provides the foundation for network-wide strategic planning, as utilities can aggregate individual pipe assessments to develop broader maintenance strategies.

Causes of Water Pipe Failure Prediction App

Ensemble learning-based predictor for causes of water pipe failure

Machine Learning For Causes of Pipe Failure

Features Used For Model Training

1. **Age:** Measured in Years
2. **Length:** Measured in Meters
3. **Diameter:** Measured in Millimeters
4. **Pressure:** Measured in Bars
5. **Traffic:** Measured in Annual Average Daily Traffic (AADT)
6. **Temperature:** Measured in °C
7. **Relative Humidity:** It has no unit
8. **Precipitation:** Measured in Millimeters
9. **Material:** Asbestos Cement (AC), Cast Iron (CI), Ductile Iron (DI), Galvanized Iron (GI), Lined Galvanized Iron (GIL), Polyethylene (PE), Steel (S), SS (Stainless Steel), UPVC (Unplasticized Polyvinyl Chloride)
10. **Corrosivity:** Highly corrosive, Mildly corrosive, Non corrosive
11. **Road type:** Carriage, Footway, Other Location
12. **Water type:** Freshwater (FW), Saltwater (SW)
13. **Land use:** Urban, Rural

The predicted class is: [   ]

Single Prediction   Batch Prediction

Batch Prediction from File

Choose your data source:

⦿ Use Test Dataset
◯ Upload Own File

Preview of test data:

	Age	Length	Diameter	Pressure	Traffic	Temperature	Relative_humidity	Precipitation	Mater
0	25	10	300	7	1,500	25	40	10	DI
1	40	50	600	10	3,000	34	60	19	GI
2	44	200	450	8	5,000	28	45	76	PE

Predict for Test Data

Download Results as CSV

Results Preview:

	Relative_humidity	Precipitation	Material	Corrosivity	Road_type	Water_type	Landuse	
0	40	10	DI	Highly corrosive	Footway	Freshwater	Rural	
1	60	19	GI	Mildly corrosive	Carriage	Saltwater	Urrban	
2	45	76	PE	Non corrosive	Other Location	Freshwater	Urban	

See required column format   ⌄

Figure 6. 7: GUI of COF Model Output (a) – single prediction (b) batch prediction

6.5.   SUMMARY

This chapter presents the development and deployment of four web-based applications designed to facilitate the implementation of the predictive models proposed in this study.

The POF application, built using the Python Flask framework for the backend and HTML and CSS for the frontend, provides a user-friendly interface to input pipe attributes and obtain predictions of the likelihood of pipe failure. The POL and POB applications, deployed

separately on the Streamlit Cloud, allow users to predict the probability of leaks and bursts, respectively. These applications leverage the optimized CNN model, which demonstrated superior performance in the evaluation phase. The COF application, also deployed on the Streamlit Cloud, enables users to identify the most likely causes of water pipe failures based on various pipe attributes and environmental factors. The application utilizes the best-performing EL model identified in the study.

Chapter 7

CONCLUSIONS AND FUTURE WORK[6]

7.1. INTRODUCTION

The primary aim of this research was to deepen the understanding of the various factors that contribute to water pipe failures, delineate the different failure modes, and develop predictive models to significantly improve the management of WDNs. Effective management of these networks entails a series of complex and critical tasks, demanding robust analytical strategies and innovative solutions. To methodically tackle these challenges, the study was designed

---

[6] This chapter is largely based upon:

Taiwo, R., Zayed, T. & Ben Seghier, M. E. A. (2024). " Integrated intelligent models for predicting water pipe failure probability". Alexandria Engineering Journal, 86, 243-257, https://doi.org/10.1016/j.aej.2023.11.047

Taiwo, R., Yussif, A., Ben Seghier, M. E. A., & Zayed, T. (2024). "Explainable Ensemble Models for Predicting Wall Thickness Loss of Water Pipes". Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2024.102630

Taiwo, R., Shaban, I. A., & Zayed, T. (2023). "Development of sustainable water infrastructure: A proper understanding of water pipe failure". Journal of Cleaner Production, 398: 136653 https://doi.org/https://doi.org/10.1016/j.jclepro.2023

Taiwo, R., Ben Seghier, M. E. A., & Zayed, T. (2023). "Towards sustainable water infrastructure: The state-of-the-art for modeling the failure probability of water pipes". Water Resources Research. e2022WR033256. https://doi.org/10.1029/2022WR033256

Taiwo, R., Zayed, T. & Adey, B.T. "Explainable deep learning models for predicting water pipe failure." Journal of Environmental Management (IF = 8.7, Q1). (Under review – 1st cycle)

Taiwo, R., Zayed, T. & Adey, B.T. "Interpretable ensemble models for predicting causes of water pipe failure." Reliability Engineering and System Safety (IF = 8.1, Q1). (Under review – 1st cycle)

Taiwo, R., Zayed, T. Elshaboury, N., & Abdelkader, E. M. "Promoting Sustainable Water Distribution Networks: Modeling of Water Pipe Failure Factors and Modes." Cleaner Engineering and Technology (IF = 5.3, Q1). (Under review – 2nd cycle)

around four key objectives that collectively outline the comprehensive scope of this research. The objectives are as follows:

a) Identify comprehensive failure factors and failure modes of water pipes;

b) Model, rank, and investigate the relationship between water pipe failure factors and failure modes;

c) Develop and automate optimized models to predict the probability of failure, leaks, and bursts of water pipes; and

d) Develop and automate an optimized model to predict the causes of water pipe failure.

The preceding chapters have comprehensively detailed the methodology employed, the data used for model development and testing, the obtained results, and the automated applications developed to facilitate the implementation of the research findings. Hence, this concluding chapter summarizes the main findings, highlights the significance and contributions of the study, acknowledges the research limitations, and provides recommendations for future work.

7.2.   SUMMARY OF THE KEY FINDINGS

Objective 1: Identify comprehensive failure factors and failure modes of water pipes

As per this objective, a systematic review of previous studies in identifying the failure factors and failure modes of water pipes was conducted. Thirty failure factors were identified and were broadly classified into four: pipe-related, soil-related, external-related, and operation-related factors. The review also identified five failure modes. Apart from the systematic review, other methodologies employed to achieve this objective include scientometric analysis and fault tree logic.

From the systematic review, it could be observed, from a general perspective, that a direct relationship exists between water pipe failure and pipe age, pipe length, soil moisture, frost,

traffic, precipitation, aggressive chemical substances, internal pressure, number of leaks, and water temperature. On the other hand, an inverse relationship seems to exist between water pipe failure and pipe diameter, wall thickness, efficient protection methods, buried depth, atmospheric temperature (seasonality), soil and water pH, and soil resistivity. In addition, the methods and materials used in manufacturing water pipes have witnessed advancement and improvement over time. A good example of this is the improvement in the way of manufacturing CI pipes. The introduction of the centrifugal method in casting makes the uniform wall thickness of CI achievable, coupled with an increase in the pipe's strength. A general way to conclude would probably be that each type of pipe has its advantages and limitations. For example, PVC pipe is more corrosion resistant than steel pipe, while steel pipe can withstand more external load than PVC pipe. It is worth mentioning that water pipes' failure mechanism is a complex one. Hence, the direct and inverse relationships mentioned above in relation to water pipe failure and its contributing factors may not be entirely applicable in some cases, as the level of dominance amongst the factors may differ in a system.

Objective 2: Model, rank, and investigate the relationship between water pipe failure factors and failure modes

Water pipe failure is a significant issue that affects both developed and developing countries across the globe. It leads to several negative impacts on the environment, economy, and social conditions of a nation. Hence, it is important to understand the critical factors influencing the pipe failure. In order to identify the criticality of these factors and failure modes, questionnaire data were collected and analyzed using the PLS-SEM algorithm.

The algorithm consists of a measurement and structural model. The result of the measurement model identified pipe age, diameter, length, material, and wall thickness as the most critical pipe-related factors affecting water pipe failure. Similarly, internal pressure, maintenance practices, number of leaks, water acidity, and water alkalinity were found as the most critical

213

operation-related factors. Temperature, frost, microbiologically induced corrosion, and precipitation were identified as the most influential external-related factors, while bedding condition, soil moisture, soil pH, soil resistivity, and soil type were evidenced to be the most critical soil-related factors. In the same vein, blown-out, circumferential cracking, and corrosion splitting were identified as the critical failure modes. The results of the structural model provide confirmation of the hypothesis: factors influencing water pipe failure significantly impact failure modes. This is evidenced by p-values less than 0.05 and a path coefficient (β) of 0.567. Consequently, this study enhances the understanding of the intricate relationships between various factors that influence pipe failure and how these interactions determine failure modes.

Objective 3 (first task): Develop and automate an optimized model to predict failure probability of water pipes

To fulfil this objective, this study gathered data from three different sources: HK Water Supply Department, HK Observatory, and HK Transportation Department. The data were grouped into pipe-related, environment-related, and operation-related factors affecting the failure probability of the pipes. Subsequently, the data was pre-processed by outlier removal and imputation of missing data using the appropriate descriptive statistics. Afterward, the data was transformed via normalization and standardization. Two levels of optimization have been experimented. The first experiment used GA to optimize the hyperparameters of LR models, and their results outperformed the models without hyperparameter optimization. The second experiment selects the best hyperparameters and features using GA. The result outperforms that of the first experiment, which emphasizes the importance of feature selection. The best model achieved an accuracy, precision, recall, f1 score, and AUC of 0.889, 0.929, 0.815, 0.867, and 0.944, respectively. This implies that the optimized model can effectively predict the failure probability of water pipes with a very low false alarm rate. This would help the water utility

management make informed decisions about when and what pipes need to be repaired based on their probability of failure. Additionally, the result of the SHAP experiment shows that "age," "temperature," "cast iron material," and "length" contribute most to the prediction. This result also agrees with the interpretation of the LR coefficients and scholarly literature. Moreover, a web-based application has been developed as part of Experiment 2 to provide a user-friendly interface for the implemented model. This application has been delivered to HK Water Supply Management to enhance their decision-making processes in effectively managing their WDN. The web application serves as a valuable tool to assist in analyzing and interpreting the model's outputs, enabling the management team to make informed decisions and take appropriate actions to optimize the performance of their WDN.

Objective 3 (second task): Develop and automate an optimized model to predict the probability of leak and burst in water pipes

This task has leveraged deep learning (DL) architectures, specifically deep neural networks (DNN), convolutional neural networks (CNN), and TabNet, to develop predictive models for the likelihood of leaks and bursts in water pipes. The study enhanced these base DL models by optimizing their hyperparameters through Bayesian Optimization and then interpreting the optimal model with SHapley Additive exPlanations (SHAP).

Data compiled from the Hong Kong Water Supply Department, the Hong Kong Observatory, and the Hong Kong Transportation Department were categorized into variables associated with pipe characteristics, environmental conditions, and operational factors. Data processing involved removing outliers and imputing missing values through established statistical methods, followed by data normalization and standardization. It was observed that transforming the data remarkably augmented the models' predictive power. For example, in the context of leak prediction, the precision of the TabNet model increased from 0.512 to 0.598

with normalization and further to 0.727 following standardization. Additionally, the refinement of the models through optimization led to enhanced performance compared to the base-DL models. This is exemplified by the increase in recall for the DNN model in burst prediction, which improved from 0.914 to 0.947 with normalization and further to 0.950 after standardization.

The Copeland algorithm identified the CNN as the most effective model for predicting the probability of leaks and bursts. For the POL, the optimal model demonstrated high performance metrics: an accuracy of 0.994, precision of 0.876, recall of 0.978, F1 score of 0.924, MCC of 0.923, and Cohen's Kappa of 0.922. Similarly, the optimal model for the POB achieved an accuracy of 0.999, precision of 0.782, recall of 0.987, F1 score of 0.872, MCC of 0.878, and Cohen's Kappa of 0.872. The SHAP analysis highlighted the predominance of 'diameter' and 'material composition, particularly plastic, in influencing model predictions. The insights gained from this investigation are invaluable for proactive management of WDNs. The predictive models developed can help utility companies mitigate pipe failures and bolster the reliability of their supply infrastructure.

Utilities can implement targeted condition monitoring and preventive maintenance for pipes with high POF and POL predictions, like pressure management and water quality management, to reduce internal corrosion. The cost-effectiveness of these interventions can be evaluated by comparing the implementation costs against potential savings from prevented leaks and associated water losses. For pipes showing high POB probability, urgent rehabilitation or replacement may be warranted, given the severe consequences of bursts. The prioritization can be optimized by considering both the probability and potential impact of failure, allowing utilities to allocate their limited resources to highest-risk assets first. When the COF model indicates material degradation as the primary cause, utilities can develop material-specific intervention programs. For example, implementing corrosion protection for metallic pipes or

adjusting operational parameters for plastic pipes. The feasibility of different interventions can be assessed based on access constraints, service disruption, and available resources. This data-driven approach enables utilities to move from reactive maintenance to proactive asset management, optimizing their maintenance budgets while improving service reliability. The cost-effectiveness of various strategies can be evaluated by analyzing historical maintenance costs against predicted failure probabilities.

Objective 4: Develop and automate an optimized model to predict the causes of water pipe failure

This objective aimed to develop accurate and interpretable ML models for predicting the causes of water pipe failures, systematically selecting the best-performing model, and generating meaningful explanations of the model predictions. The study employed five state-of-the-art EL algorithms: AdaBoost, Random Forest, XGBoost, LightGBM, and CatBoost. The TPE algorithm was utilized for hyperparameter optimization, significantly improving the performance of the base models. The Copeland algorithm was then applied to rank the optimized models, identifying the XGBoost model as the top performer. Finally, the SHAP framework was used to enhance the interpretability of the selected model.

The results demonstrate that the optimized XGBoost model achieved the highest accuracy (0.82), macro precision (0.65), weighted precision (0.80), and weighted recall (082), among the evaluated models. LightGBM was found to have one of the highest values for macro recalls (0.58), weighted F1 scores (0.80), and AUCs (0.87). It was also found to be the most efficient in terms of computational cost (i.e., least training time). The utilization of the Copeland algorithm for systematic ranking facilitated the selection of the optimized XGBoost as the best-performing model for deployment. The SHAP feature importance analysis and the inherent feature importance from the optimized XGBoost model provided deep insights into the critical

factors influencing the prediction of water pipe failure causes. Both analyses highlighted water type (specifically saltwater), material (GIL), age, and diameter as key determinants, demonstrating a significant overlap in their results. Specifically, the SHAP analysis identified traffic as another crucial factor, underscoring its role in affecting pipe integrity. Conversely, the XGBoost model pinpointed another material type, steel (S), as a significant predictor. This slight variance in the two analyses enriches our understanding, suggesting that both traffic conditions around the water pipes and the specific materials used for their construction are pivotal in predicting failures. This consistency in the results across different analytical methods confirms the reliability of these features as predictive indicators. The identification of these factors enables targeted interventions and more precise preventative measures, improving the management and maintenance strategies within water distribution networks.

To clarify the model's capabilities and limitations, the optimized XGBoost model predicts the most probable cause of failure from a predefined set of failure causes based on the input features. While the model has been trained on historical data with known failure causes, its predictions are limited to identifying patterns similar to those in the training dataset. The term 'automate' in this context refers to the implementation of the model in a web-based application that automatically processes user inputs and generates predictions without manual intervention. The model's optimization specifically involved tuning its hyperparameters using the TPE algorithm to maximize its performance. The model's reliability in predicting failure causes is evidenced by its consistent identification of key contributing factors (water type, material type, age, diameter, and traffic) through both SHAP analysis and inherent feature importance, which align with established engineering knowledge and previous studies in the field.

7.3. SIGNIFICANCE AND CONTRIBUTION OF THE RESEARCH

This research makes theoretical, practical, and broader contributions to the field of WDN management by advancing the understanding of water pipe failure mechanisms and improving predictive modeling techniques.

The following theoretical contributions are supported by the findings of this study:

a) Comprehensive identification of failure factors and modes: This research has systematically identified and categorized a comprehensive set of failure factors and failure modes associated with water pipes. This comprehensive identification of failure factors and modes provides a solid foundation for understanding the complex nature of water pipe failures and serves as a valuable resource for researchers and practitioners in the field.

b) Establishing the relationship between failure factors and modes: This study has developed and validated a conceptual model to establish the relationships between failure factors and modes. This contribution enhances the theoretical knowledge of water pipe failure mechanisms and enables a more holistic approach to addressing this complex problem.

c) New application of advanced analytical techniques: The study has successfully applied state-of-the-art analytical techniques, such as Partial Least Square Structural Equation Modeling (PLS-SEM), deep learning architectures (DNN, CNN, and TabNet), and ensemble learning algorithms (AdaBoost, Random Forest, XGBoost, LightGBM, and CatBoost) to predict water pipe failures. The innovative use of these techniques in the context of water infrastructure management has expanded the theoretical boundaries and demonstrated their potential for addressing complex problems in this field. This contribution showcases the value of applying advanced analytical methods to tackle

real-world challenges and encourages further exploration of these techniques in related domains.

d) Enhanced interpretability of predictive models: By employing advanced interpretation techniques, such as SHapley Additive exPlanations (SHAP) and inherent feature importance analysis, this research clarifies the relative influence of different factors on pipe failure predictions. The transparency of the model is achieved through SHAP analysis, which quantifies how each feature contributes to individual predictions, showing exactly how much factors like age, diameter, or pressure affect the predicted outcome. This transparent quantification allows utilities to understand why the model makes specific predictions - for example, showing that a high failure probability might be driven 60% by pipe age, 25% by pressure conditions, and 15% by soil characteristics (this is for illustration purposes).

The practical contributions of this study are as follows:

a) Identification of critical failure factors and modes: The research has identified the most critical factors influencing water pipe failures and critical failure modes. This practical knowledge empowers water utilities to focus their efforts on monitoring and mitigating these key factors, leading to more effective and targeted maintenance strategies. By prioritizing the most influential factors, water utilities can optimize their resource allocation and improve the overall reliability and resilience of their distribution networks.

b) Development of enhanced predictive models: The research has developed optimized models for predicting the probability of failure, leaks, bursts, and causes of water pipe failures. These models serve as practical tools for water utilities to assess the condition of their infrastructure, prioritize maintenance and rehabilitation efforts, and optimize resource allocation. The models are implemented through a web-based interface where

utilities can input their latest network data and receive immediate predictions. This implementation enables utilities to maintain current assessments of their infrastructure as new data becomes available through routine inspections, maintenance activities, or operational monitoring. For instance, when utilities gather new condition assessment data or update their maintenance records, they can readily input this information to obtain updated failure predictions, supporting dynamic decision-making based on the most recent network information. This continuous update capability ensures that maintenance and rehabilitation decisions are based on current network conditions rather than outdated assessments.

c) Creation of user-friendly web applications: The study has developed web-based applications that provide a user-friendly interface for the implemented prediction models. These applications allow utility managers to input their pipe data through a simple form interface and receive failure outcomes. The current implementation focuses on core functionality: data input, prediction generation, and basic result display. The applications make the complex prediction models accessible to users without requiring programming knowledge or direct interaction with the underlying algorithms. While the current version provides essential functionality, future enhancements could include more advanced visualization features, integration with GIS systems, and automated reporting capabilities. It is acknowledged that fuller integration into daily operations would require additional features and customization based on specific utility needs and workflows.

d) Improved Risk Management: The prediction models support risk management by identifying pipes with high failure probabilities, enabling utilities to prioritize specific interventions. For example, when the model predicts high failure probability driven primarily by pressure factors, utilities can implement targeted pressure management in

that zone. Similarly, when predictions indicate clusters of high-risk pipes in areas with aggressive soil conditions, utilities can prioritize these zones for cathodic protection or protective coating applications. For pipes where age and material type are the primary risk factors, the predictions help justify pipe replacement or rehabilitation in capital planning. The model outputs can also guide the frequency and intensity of condition monitoring. Pipes with moderate risk levels might be scheduled for more frequent acoustic leak detection surveys, while those with the highest risk predictions might warrant continuous pressure monitoring. This risk-based approach helps utilities allocate their limited inspection and maintenance resources more effectively.

The broader contributions of this research are highlighted below:

a) Advancement of sustainable water infrastructure management: By deepening the understanding of water pipe failures and providing predictive tools, this research contributes to the broader goal of developing sustainable water infrastructure. The findings and models developed in this study can help water utilities enhance the reliability, resilience, and longevity of their distribution networks, ensuring the safe and efficient delivery of water to communities. This contribution aligns with the global efforts to achieve sustainable development and promotes the responsible management of vital water resources.

b) Promotion of data-driven decision-making: The research demonstrates the value of data-driven approaches in addressing complex challenges in water infrastructure management. By leveraging advanced analytical techniques and integrating data from multiple sources, this study promotes a culture of evidence-based decision-making in the water sector, encouraging the adoption of similar approaches in other domains. This contribution has the potential to transform the way decisions are made across various industries, leading to more informed, efficient, and effective outcomes.

222

c) Contribution to the global sustainability agenda: The effective management of WDNs is crucial for achieving the United Nations' Sustainable Development Goals (SDGs), particularly SDG 6, which aims to ensure access to clean water and sanitation for all. This research contributes to the global sustainability agenda by providing tools and insights that can help water utilities worldwide improve their infrastructure management practices, ultimately contributing to the achievement of the SDGs. The findings and models developed in this study have the potential to be adapted and applied in different contexts, supporting the global efforts to ensure sustainable access to safe drinking water.

7.4. RESEARCH LIMITATIONS

While this research has made significant contributions to the understanding and prediction of water pipe failures, it is essential to acknowledge the limitations of the study. These limitations may serve as opportunities for future research and improvements in the field.

a) Scope of Failure Factors and Failure Modes: While the research has systematically identified a broad range of failure factors and modes, the complexity of WDN means that not all possible influencing factors may have been captured in the literature reviewed. Unforeseen interactions or rare events that were not sufficiently represented in the literature could lead to unforeseen failures.

b) Assumptions in modeling techniques: The research employed various modeling techniques, including PLS-SEM, deep learning architectures, and ensemble learning algorithms. These techniques involve certain assumptions and simplifications to represent the complex reality of water pipe failures. For example, the PLS-SEM model assumes that the relationships between variables are linear, which may oversimplify the actual complex and potentially nonlinear interactions involved.

c) Data Dependency: The study utilized global expert knowledge for the questionnaire survey. While core failure principles remain consistent, regional variations in infrastructure characteristics, operational practices, and environmental conditions may affect the relative importance of different factors. Future work could benefit from targeted local expert surveys to refine the findings' applicability to Hong Kong's specific context. The predictive models developed, and conclusions drawn are heavily reliant on the quality and extent of the data available from the Hong Kong Water Supply Department, the Hong Kong Observatory, and the Hong Kong Transportation Department. While comprehensive, these data sources may not fully represent conditions in other geographical locations or environments with different infrastructure dynamics. This could limit the models' applicability in regions with vastly different climatic, geological, or operational characteristics.

d) Model Specificity: The predictive models, particularly the deep learning and ensemble methods, were optimized and validated based on specific datasets. Although these models show high accuracy and reliability within the scope of the available data, their performance might vary when applied to data with different attributes or under different operational conditions.

e) Technological Limitations: The advanced analytical techniques employed—such as deep learning algorithms, ensemble learning algorithms, and SHAP feature importance analysis—require substantial computational resources due to the complexity and size of the datasets used. These resource requirements might constrain the practical deployment of the models in settings with limited computational capabilities or in scenarios requiring rapid data processing.

7.5. RECOMMENDATIONS FOR FUTURE WORK

This section outlines potential areas for improvement, aiming to build upon the foundation laid by this research. These recommendations are categorized into enhancements of the existing research and extensions to broaden its scope and applicability.

Enhancement for the existing research:

a) Incorporate additional failure factors: Future research could explore the inclusion of additional factors that may contribute to water pipe failures, such as the quality of installation workmanship, the presence of nearby underground utilities, and the impact of seismic activities. Incorporating these factors, subject to data availability, could enhance the predictive power of the models and provide a more comprehensive understanding of the failure mechanisms.

b) Investigate alternative modeling techniques and non-linear relationships: While this study employed PLS-SEM, deep learning architectures, and ensemble learning algorithms, future research could explore advanced modeling techniques capable of capturing complex non-linear relationships between failure factors. This includes investigating non-linear structural equation modeling, advanced time-series analysis with non-linear components, survival analysis with time-varying coefficients, and sophisticated hybrid models that can handle both linear and non-linear interactions. Additionally, exploring deep learning architectures specifically designed for complex temporal and spatial dependencies could provide better insights into the intricate relationships between failure factors and modes. These advanced modeling approaches may lead to improved predictive performance and a deeper understanding of the complex failure factors in WDNs.

c) Enhance model interpretability: Although this study employed techniques like SHAP and inherent feature importance analysis to interpret the predictive models, future

research could further enhance the interpretability of the models. This could involve developing more intuitive visualizations and automating the generation of human-readable explanations. Enhancing model interpretability would facilitate better communication of the findings to stakeholders and support more informed decision-making.

d) Incorporate real-time data: Future research could explore the integration of real-time data, such as sensor readings or monitoring data, into the predictive models. Incorporating real-time data would enable the models to adapt to changing conditions and provide more timely predictions of impending failures. This could support proactive maintenance interventions and improve the overall reliability of the WDNs.

e) Evaluate different management strategies: Future development could focus on utilizing the predictions from the web applications (POF, POL, POB, and COF) to evaluate different management strategies. By aggregating these model outputs across the pipe network, utilities can assess various maintenance scenarios, such as prioritizing repairs based on failure probabilities, optimizing rehabilitation schedules using predicted leakage risks, or developing targeted maintenance strategies based on predicted break patterns. This enhancement will transform the current pipe-level predictions into network-wide strategic planning tools.

Extension for the existing research:

a) Validate models in different contexts: Future research could validate the developed models using data from other regions or countries with different geographical, climatic, and infrastructural characteristics. This would help assess the generalizability of the findings and identify any context-specific factors that may influence water pipe failures. Validating the models in diverse contexts would strengthen their robustness and applicability to a wider range of WDNs.

b) Investigate effective intervention strategies: While this study focused on predicting water pipe failures, future research could extend the work by investigating effective intervention strategies to prevent or mitigate failures. This could involve evaluating the impact of different rehabilitation techniques, materials, or management practices on reducing the risk of failures. Conducting cost-benefit analyses of various intervention strategies would support informed decision-making and optimize the allocation of resources for maintaining and improving WDNs.

c) Develop decision support systems: Future research could aim to develop comprehensive decision support systems that integrate the predictive models, intervention strategies, and other relevant information. These systems could assist water utility managers in making data-driven decisions regarding maintenance planning, resource allocation, and infrastructure upgrades. Developing user-friendly interfaces and incorporating visualization tools would enhance the usability and adoption of these decision support systems in practice.

d) Explore the impact of climate change: Future research could investigate the potential impact of climate change on water pipe failures. As climate patterns evolve, the frequency and severity of extreme weather events, such as droughts, floods, or temperature fluctuations, may affect the integrity and performance of WDNs. Studying the long-term effects of climate change on water pipe failures and developing adaptive strategies could help water utilities build resilience and prepare for future challenges.

7.6. SUMMARY

This chapter recaps the objectives of this research, followed by a summary of the key findings derived from the study. It then outlines the significant contributions and implications of the research. Additionally, the chapter acknowledges the limitations of the study. Based on these discussions, suggestions for potential future research directions are explained.

APPENDIX A
Further details on the literature review

(I) Further literature on Limit state equations based on structural reliability

As known in the limit state reliability theory, the results from the study (Mahmoodian & Aryai, 2017) show that consideration of multiple limit states is essential in determining the failure probability of water pipes since different failure mechanisms contribute to overall pipe failure. For instance, the service life of a pipe was estimated to be 140 years when only the leakage limit was considered, whereas it was estimated to be 45 years when six limit states were considered (Mahmoodian & Aryai, 2017). This shows that consideration of only one limit state can lead to inappropriate service life estimation, which may facilitate wrong decision-making in relation to pipe management.

While developing a risk index for water pipes located at Fairfield, California, USA, Mazumder et al. (2021) adopted a physical probabilistic approach to generate a fragility curve (i.e., the conditional probability of failure). MCS (10,000 samples) was used to deal with the uncertainty associated with stress calculation. Lognormal and normal distributions were used for the probabilistic analysis. The pipe capacity was estimated using the model of (Ji et al., 2017), while the equation proposed by Robert et al. (2016) on exerted stress was adopted. The actual ages of the pipes in the network were assumed to range between 70 to 100 years. This may be inaccurate, as cast iron (CI) pipes have been used in the USA since the late 19th century. Similarly, the model does not consider any operation and environment-related factors apart from internal pressure and traffic loads, respectively. No sensitivity analysis was conducted on the input variables.

The First/Second Order Reliability Methods (FORM/SORM) are two analytical reliability methods that use the linear and quadratic approximations of the LSF to calculate the reliability index, which is then converted to failure probability. The efficiency of these approaches is determined by the complexity of the LSF (e.g., more applicable to less expensive and linear

LSF) and the distribution of the random variables (e.g., FORM is applied when the variables follow a normal distribution) (Ben Seghier et al., 2020, 2022). Apart from FORM/SORM, First Order Second Moment Approximation (FOSM) is another approach that has been used to model the failure probability of water pipes. De Leon & Macías (2005) examined the effect of spatial correlation on the failure probability of corroded pipes using the FOSM method. Their results indicated that pipe segments adjacent to each other have a high correlation coefficient in terms of failure probability due to corrosion depths, while pipe segments far away, having two or more segments between them, had zero degree of correlation. However, the study assumes that failure occurs only due to internal pressure, serving as a limitation of the study. A similar assumption was made in the study of De-Silva et al. (2006).

While other research studies focused on the failure probability of CI pipes, the study conducted by Davis et al. (2008) focused on AC pipes. Unlike CI and other metallic pipes that undergo electrochemical corrosion, AC pipes' main degradation process is due to leaching corrosion as a result of contact with soft water. The uncertainty in the degradation mechanism of water pipes is complex and dynamic in nature. For example, the leaching rate in an AC pipe can differ across the pipe length due to changes in the water pH or other influencing factors. One way to handle this uncertainty is by defining the input parameters in the form of a probability distribution. This promotes better understanding and facilitates accurate failure prediction.

Furthermore, it is essential to note that the spatial-temporal analysis of corroding pipes has received limited attention from researchers. Aryai & Mahmoodian (2017) used random field theory to determine the correlation length of corroded water pipes, which was used in the estimation of failure probability. The correlation length is the length at which a corroded surface can be said to exhibit uniform corrosion. That is, the lesser the correlation length, the more fluctuated the surface is in terms of corrosion. Although the prediction accuracy of their model is 84.4%, which is higher compared to other studies that do not consider the spatial

relationship of corroding pipes, the accuracy can be improved if the probability distribution of corrosion depth is not assumed to be constant over time. A similar investigation was conducted by Wang et al. (2021). Their research considered five failure modes: ring deflection failure, leakage failure, burst failure, collapse failure, and bending failure. They found that each of the failure modes exhibited a different probability of failure due to the differences in their failure mechanism and contributing factors causing the failure. Although five failure modes were considered in the study, the addition of other failure stress, such as thermal stress, can increase the robustness of the models.

Moreover, the maximum stress a pipe is subjected to can be determined using finite element modeling (FEM) (Aryai et al., 2020; Li et al., 2021). Li et al. (2021) developed 243 FEA models to forecast the failure probability based on failure pressure. The reason for developing a large amount of FEA models (i.e., 243) is to fit the stress distribution to an appropriate statistical model. The maximum stress of the 243 FEA models was fitted into four distributions: Weibull, lognormal, normal, and Gumbel distributions. Normal distribution was found most appropriate for the stress distribution and was employed for MCS. Sensitivity analysis was conducted, and thickness was found as the most sensitive parameter, followed by internal pressure and traffic. The study only focused on large-diameter pipes. Furthermore, only localized corrosion was considered in the development of the FEA models, while uniform corrosion was ignored.

(II) Further literature on statistical-based models
Parametric models
Parametric models involve a finite number of parameters that describe the historical data under investigation. Parametric models make assumptions regarding the distribution of the data and the form (i.e., linear, exponential, etc.) of the relationship between the variables being studied.

This indicates that the shape of the model is fixed and determined by the chosen parameters. Parametric models are relatively easier to interpret since their parameters are known.

Weibull distribution models

Different researchers adopted the Weibull distribution model to predict the failure probability of water pipes (Vladeanu & Koo, 2015; Ward et al., 2017). Equation A.1 represents the failure probability $F(t)$ of a pipe at time t using the Weibull distribution.

$$F(t) = 1 - e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}}$$
(A.1)

where β, γ, and η are the shape, scale, and location parameters. The parameters can be determined by conducting a regression analysis of the historical failure data.

In the process of developing a risk model for water distribution networks, Phan et al. (2019) employed three-parameter Weibull distribution to model the probability of water pipe failure. The model showed that CI pipes exhibited a 100% probability of failing at 60 years, while that of DI pipes was at 90 years. Thus, CI pipes are more prone to failure. However, limited historical data was used to develop the model; hence, the model shows low accuracy. Vladeanu & Koo (2015) applied two-parameter Weibull distribution to achieve the same objective. The network comprises AC, CI, DI, PVC, and concrete pipes. Due to the unavailability of complete historical data, the authors assumed that the first breakage experienced by each of the pipes happened after 74 years of installation. This assumption may either underestimate or overestimate the network failure probability. Similarly, Ward et al. (2017) applied a 3-parameter Weibull distribution and found that the distribution was suitable for incomplete historical data left-truncated. The model was applied to two case studies. The accuracy of their model was investigated by plotting the predicted failure count against the observed failure count. All the pipe materials had an R^2 greater than 0.96. Although the main output of the model

is the failure probability of each pipe material, the failure count was used to identify the pipe age range that experiences the most failure.

Poisson distribution models

Historical failure data can also be fitted to a Poisson distribution to determine the failure probability of a system. Using the Poisson distribution, the probability of occurrence of an event within a specified interval is given by Equation A.2.

$$p\,(\lambda, n) = \frac{e^{-\lambda} \times \lambda^n}{n!} \tag{A.2}$$

where n is the number of event occurrences, and λ is the mean value (i.e., failure rate) at a specific interval. Therefore, the failure probability of a water pipe can be expressed by Equation A.3 (Singh & Adachi, 2012). Based on the failure rate (λ), Poisson distribution can either be homogenous or non-homogeneous.

$$P_f = 1 - \frac{e^{-\lambda} \times \lambda^0}{0!} = 1 - \frac{1}{e^{-\lambda}} \tag{A.3}$$

Homogenous Poisson distribution models

A Poisson distribution model is classified as homogenous when it has a constant failure rate. In the process of developing a risk model for a municipality in Canada, Elsawah et al. (2016) developed a homogenous Poisson model for predicting the failure probability of water pipes. The failure rate was determined based on historical failure data for the last five years. The results indicated that galvanized steel pipe exhibited the highest failure probability compared to other pipes, while the failure probability of PVC, CI, and DI are similar. No attempt to validate the model with other historical or field data was made; hence, the model's prediction accuracy may not be generalized. Furthermore, Singh & Adachi (2012) assumed homogenous Poisson distribution to fit the historical data of a water utility in Honolulu, USA. The annual failure rate from 1988 to 2008 was determined from the historical data and employed in the

development of the model. Concrete pipes were found to have the lowest average probability of failure, followed by DI pipes. As one would expect, CI has the highest average failure probability. It should be noted that the homogenous Poisson model may be inadequate for failure prediction, as the failure rate of water pipes is not always constant due to the complex mechanism of pipe deterioration.

Non-homogeneous Poisson distribution models (NHPP)

The failure rate varies with time in the non-homogenous Poisson distribution model (NHPP). Rogers & Grigg (2007) employed NHPP in predicting the water pipe failure probability of a utility in the USA. However, the NHPP model requires a minimum of three break records for a pipe before it can be applied since the model is governed by the time-varying failure rate, which is an integral function (i.e., two break intervals are needed); else, the governing equation will be unsolvable. Economou et al. (2007) modified the traditional NHPP model to deal with zero inflation in the historical data of a utility in New Zealand. Excess numbers of zero points characterize the historical failure data. The data comprised 532 AC pipes, where only 81 of them had exhibited one or more failures. Their model suggested allowing for zero inflation in failure prediction increases the predictive capacity of such a model. A major limitation of NHPP models is that they are memoryless, since the effect of previous failure is not taken into consideration while determining the subsequent failure rate.

Proportional hazard models (PHM)

While Cox originally developed the Proportional Hazard Model (PHM) for medical applications, it has been applied for pipe deterioration mechanisms; thus, it has been used to predict pipe failure probability (Debón et al., 2010; Mailhot et al., 2000). Generally, the PHM models are used in determining the hazard function of a system. From the hazard function, the reliability of such a system could be determined through its failure probability. The general form of the Cox proportional hazard model is represented by Equation A.4.

234

$$\lambda(t, x_1 \dots x_p) = \lambda_0(t) \times \exp(\beta_1 x_1 + \dots + \beta_p x_p) \tag{A.4}$$

where $\lambda(t, x_1 \dots x_p)$ is the hazard function, $\lambda_0(t)$ is the baseline hazard function at time "t", which describes the hazard when no covariate is considered. The coefficient of the covariate, x, (i.e., explanatory variables) influencing the risk of a system is denoted by β. There are two approaches to developing PHM models: semi-parametric and parametric methods. In the semi-parametric approach, the baseline model is left undefined (non-parametric), which eliminates the bias of assuming the shape of the hazard function. However, an assumption about the beta parameter is made (parametric). This is the case for the Cox PHM model. On the other hand, an assumption about the hazard function and the beta parameter is made in the case of parametric PHM models such as Weibull and exponential PHM models. An interesting fact about PHM models is their ability to estimate the relative importance of explanatory variables on water pipe failure, for example, without knowing the form of the hazard function.

The failure probability of water pipes in a medium-sized water utility company in Spain was modeled by Debón et al. (2010) using the Cox PHM. The pipes in the network have been installed since 1941; however, the failure data from 2000 onwards to the time of the model's development is only available. Due to this, 98% of the data used in the model development was left-censored. Moreover, only the failure year is recorded in the limited data; hence, failure times are assumed. Based on the interpretation of their results, it could be inferred that the hazard rate of shorter pipes is lower compared to longer ones. This shows that a longer pipe has a higher failure probability. Similar trends have been noticed in pipes with higher pressure. However, an inverse relationship between pipe diameter and failure probability was reported. Although the AUC of the model is 0.76, the accuracy may increase if the data quality is improved. Furthermore, with the aim of identifying the most influencing factors on water pipe failure, Vanrenterghem-Raven (2007) employed Weibull and Cox PHM. Right censorship was

applied to the pipes since only 6.5% of the pipes had experienced at least one breakage. The Cox PHM was used to identify the most significant factors and their interdependencies, while Weibull PHM was used for failure prediction. The results indicated that pipe age, diameter, and previous break are the most significant factors. However, a main disadvantage of the PHM model is that the hazard ratio is constant and does not change over time. This is not true for water pipes, as the deterioration mechanism depends on various factors that are dynamic and time dependent.

Logistic regression models

Logistic regression (LR) is a regression form used to solve classification problems. In this type of modeling, the historical failure data is categorized into two, where a value of 1 is usually assigned to failed pipes and 0 to non-failed pipes. The probability of water pipe failure, for instance, using the logistic regression approach, can be represented by Equation A.5, and the value of z is explained by Equation A.6.

$$P_f = \frac{1}{1+e^{-z}} \tag{A.5}$$

and $z = b_o + b_1 X_1 + b_2 X_2 + \cdots b_p X_P$ (A.6)

where b_o is the intercept of the regression line, X_1 to X_P are the explanatory variables and b_1 to b_p are their coefficients, respectively.

Al-Ali et al. (2020) employed LR to analyze the failure of 8 different types of pipes. Out of the 43 independent variables used for the predictive model, 16 variables were statistically significant and formed the basis of the regression equation. Although the prediction accuracy of the model is more than 70%, it can be improved if more comprehensive velocity data is used to develop the model, as only 20% of the velocity data was available, while an absolute value was assumed for the rest. Additionally, more explanatory variables such as temperature,

bedding factor, and water hammer, amongst others, could be added to make the model more robust.

Furthermore, the logistic model of Cooper et al. (2000) used input parameters such as pipe diameter, soil corrosiveness, the proximity of pipes to each other, traffic load, and urban development year. It should be noted that pipe age was not used in the model, as it is absent in the historical data, contributing to the limitations of the model. Their results indicated that the "proximity of pipes to each other" is a significant variable in failure prediction. This shows that this parameter needs to be explored in further research, as many studies have not paid attention to it.

Nonparametric models

Nonparametric models are models whose parameters are infinite and not predefined before the development of the model. Unlike parametric models, no assumption is made regarding the distribution of the data or the functional form of the relationship between the variables. This makes nonparametric models to be more flexible than the parametric ones, but it also indicates that nonparametric models can be relatively difficult to interpret.

Bayesian-based models

The Bayesian-based model is a type of statistical model where the concept of probability is used to deal with uncertainty associated with both the inputs and output of the model. Bayesian models stand on the concept of Bayes' theorem presented in Equation A.7.

$$P\left(A|B\right) = \frac{P(B|A)\,P(A)}{P(B)} \tag{A.7}$$

where the probability of A occurring, given B, is denoted by $P(A|B)$ while the probability of B occurring given A is denoted by $P\left(B|A\right)$. P(A) and P(B) are the probability of A and B occurring, which are referred to as the prior and marginal likelihood, respectively. Furthermore, $P\left(A|B\right)$ is termed the posterior probability.

Singh (2011) leveraged Bayes' theorem to determine the failure probability of seven pipe types based on various factors causing failures. The factors considered are break cause, pipe age, pipe diameter, and soil type. The first process was to determine the prior probability of pipe failure based on any of the considered factors given its pipe material. The second stage involved the calculation of posterior probability based on the result of the prior probability. The prior probability will allow the utility managers to know the dominant cause of water pipe failure, while the posterior probability will assist the utility managers in deciding the best choice of pipe material for a particular location. It should be noted that this method is suitable for determining the failure probability in the current year of assessment. An attempt to use the model for predicting future failure probability might give inaccurate results unless the database is updated; otherwise, various assumptions will need to be made.

Chik et al. (2017) proposed a Bayesian simple model (BSM) involving four steps for estimating the failure probability. The steps include 1) grouping the pipes based on the number of failures until the year before the assessment year; 2) grouping the pipes based on the number of failures in the assessment year 3) counting the number of pipes in each group; and 4) estimating the failure probability of the pipes in each group based on the assumption that failure probability of water pipes follow Bernoulli distribution. The BSM model was compared with HBP and the Nonhomogeneous model, which showed comparable accuracy. However, their results found that the BSM model's accuracy for long-term prediction depends on the availability of new data. For instance, if the available data used for training the model is up till 2020, the estimated probability of failure for the same group of pipes will be the same for 2021, 2022, and so on, unless new data is available to update the model. The grouping of pipes referred to in the study by (Chik et al., 2017) is a result of the unavailability of data relating to the network, which is a typical issue in many WDNs.

Three common characteristics of WDN data are right censoring, left truncation, and exclusion of replaced pipe data (Scheidegger et al., 2013). Right censored data are pipes that are yet to experience any failure since installation or last replacement. Data related to pipes that have been installed before failure records are systematically observed are referred to as left-truncated data. Hence, the number and time of failures of these pipes are unknown. The third characteristic refers to the absence of replaced pipe data, which may be because of deleting a replaced pipe record from the database or the pipe replacement was done before establishing a record database.

Furthermore, Tchórzewska-Cieślak et al. (2019) used the Bayesian model to determine water pipes' total and conditional failure probability based on the length of the pipe and the number of failures. Their investigation includes distribution and transmission pipes. It was found that the failure probability of the distribution water pipes was higher than transmission pipes.

Shin et al. (2016) used a competing hazard model to estimate the failure probability of water pipes in a city in South Korea. The estimation was done using a Bayesian Inference based on Markov Chain Monte Carlo Method. They argued that it is important to investigate the effect of a competing event on a pipe's failure. Hence, pipe failure due to a burst in the pipe's body was defined as the main failure event (termed as B-burst), while a burst in the connection part of the pipe was defined as the competing event (termed as C-burst). The result of the pipe material tested (ductile cast iron) showed that the pipe exhibits a lower B-burst than C-burst.

Furthermore, the failure probability of individual pipes belonging to a certain homogeneous group of pipes was investigated using the likelihood ratio obtained via Bayes' theorem (Kleiner & Rajani, 2012). The pipes were grouped based on the previous number of failures. The results were compared with those obtained from a logistic regression model and an ordered list (a

239

heuristic-based technique) model. In terms of evaluation metrics, no model was superior to the other as the performance of each model was different for various datasets.

Hierarchical Beta Process (HBP) models

The hierarchical beta process (HBP) is a model that is capable of predicting the failure probability of water pipes by regrouping pipes with similar features. It usually consists of three phases, including the beta, the Bernoulli, and the hierarchical processes. The HPB model can be referred to as a non-parametric Bayesian model as the beta and Bernoulli processes can be employed as prior distribution for the hierarchical process. The beta and Bernoulli processes are conjugates of each other.

Luo et al. (2017) analyzed pipe failure using two algorithms based on the Bayesian framework. The first algorithm employed the Infinite Gamma-Poisson Mixture Model, a representation of the Dirichlet process, to assign an index to pipe groups based on similar features. Afterward, the output of the first process was used as input for developing the HBP model. Prior to the development of the HBP model, the beta-Bernoulli process was generated to estimate the prior failure probability of each pipe group, which was subsequently regenerated and ranked using the HBP model. The prediction accuracy of the proposed model increased by more than 10% compared to the conventional HBP model that uses the knowledge of domain experts rather than the Infinite Gamma-Poisson Mixture Model for pipe grouping. However, the proposed model could be more realistic by considering factors related to the weather (e.g., temperature) and operation (e.g., internal pressure) of the network.

Similarly, the model updating capability of the nonparametric Bayesian method was explored by Lin et al. (2015) for the failure probability prediction. Like the study of Luo et al. (2017), two processes were involved in the development of the model. However, the first process used the Chinese restaurant process (CRP) for generating the group index. As CRP is employed as a representation of the Dirichlet process, it mimics the influx of customers (i.e., data points)

into a restaurant and getting assigned to different tables (i.e., clusters). The second process of developing the HBP model follows the same methodology as that of the study by Luo et al. (2017). It should be noted that more influencing factors affecting the failure probability are considered in the study of Lin et al. (2015) compared to that of Luo et al. (2017). These factors include soil expansiveness, soil map, soil geology, availability of coating, and distance to a traffic intersection. Additionally, the model developed by Lin et al. (2015) had higher prediction accuracy compared to that of Luo et al. (2017), probably because more influencing factors were considered in the former.

Other statistical-related models

In addition to the probability theory, Ismaeel & Zayed (2018) employed multicriteria decision methods such as fuzzy analytical network process (FANP), preference ranking organization method for enrichment evaluation (PROMETHEE), multi-attribute utility theory (MAUT) for calculating performance indices for both water pipes and their associated accessories such as valves. The failure probability of each component was determined from the performance index (PI). The model was built on expert opinions, which was subsequently validated by applying it to a case study. On average, the model achieved a prediction accuracy of 94.4%, using the validation factor proposed by Zayed & Halpin (2004). Similarly, Karamouz et al. (2012) developed an algorithm to determine the vulnerability of water pipes. Their definition of vulnerability centers on the failure probability of water pipes. Out of six factors determined through the literature review, three influencing factors relating to water pipe failure were selected as the most representative factors by Minimum Redundancy-Maximum Relevance (MRMR) feature selection method, shown in Figure A1. Subsequently, Analytical Hierarchical Process (AHP) was used to give weights to these three factors. Finally, the probability of any pipe failing in a distribution network was estimated from the aggregation of the AHP results and the values of each factor determined from the MRMR. However, the developed model was

applied to a hypothetical case study, which raised a question on the real-world applicability of such a model in addition to the fact that the determination of weights for the factors is subjective.
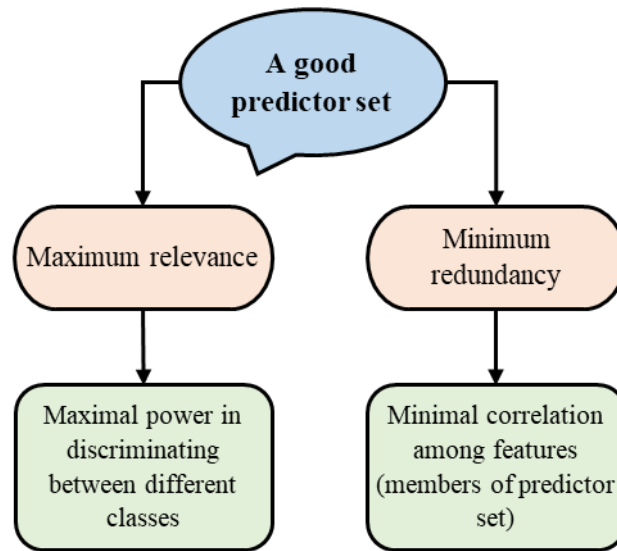

Figure A1: Specifications of a good predictor set.

Moreover, Tchorzewska-Cieslak (2012) estimated the failure probability of water pipes by making some assumptions based on the failure rate (number of failures/km/year). They assumed low, medium, and large failure probability when the failure rate is less than 0.5, equals to 0.5, and greater than 1, respectively (i.e., weighting method). This assumption is subjective and may not be applicable to other WDNs.

(III) Further literature on AI-based models

Fuzzy based models

Fuzzy-based models are used to tackle complexity, vagueness, and uncertainty in different systems. In the case of water pipes, the failure mechanism is complex and not properly understood since different factors interact ambiguously to cause pipe failure. Hence, fuzzy-based models can be employed to solve this problem. In the fuzzy approach, a crisp, imprecise variable is fuzzified using defined membership functions. Afterward, a set of fuzzy rules is

applied to the fuzzified variable to modify it. Subsequently, the fuzzified variables are defuzzified to get an unbiased crisp variable.

Al-Zahrani et al. (2016) developed a fuzzy-based model that uses 13 input factors relating to the structural integrity of the pipes, water quality, and operation of the network. These factors were arranged in a hierarchy, and fuzzy membership functions were defined for each of the factors based on the characteristics of each factor obtained in the literature (AWWA, 2002; Sarbatly & Krishnaiah, 2007). Subsequently, AHP was employed to determine the relative weights of each fuzzy set. After aggregating the fuzzy sets and AHP, the fuzzy variables were defuzzified to determine the failure probability based on the historical data. Although this approach quickly estimates the failure probability, high precision may not be achieved as it was used as an approximate method. Moreover, it is not clear how the authors assumed some values for the weight of the fuzzy sets.

In a related study, fuzzy rule-based, and fuzzy-synthetic evaluations were employed for prioritizing water pipes based on their risk index. The risk index is a multiplication of failure probability and consequences of water pipes. Similar to the approach employed by Al-Zahrani et al. (2016), the failure probability index for each water pipe was determined by aggregating the fuzzy sets with the weights of each factor obtained from the AHP analysis. Afterward, the risk indices were developed, and the most probable pipes to fail were visualized using the GIS map. The model did not capture critical influencing factors of water pipe failure, such as pipe diameter and internal pressure. The ability of fuzzy-based models to be built on imprecise, distorted, and vague data is a major advantage over other techniques. However, a major limitation is its reliance on expert knowledge to formulate fuzzy rules.

Machine learning-based models

Machine learning (ML) models can learn or recognize specific patterns from a set of data. ML-based models make future predictions based on the learned patterns when new data are fed into

them. It should be noted that some statistical models such as Bayesian Network, Logistic Regression, Naïve Bayes, and so on are adapted as an ML technique to overcome the challenge of computational intensiveness and inability to handle high dimensional data of statistical models. Therefore, some statistical models that are used as an ML method appear in this section. ML models can be broadly classified into supervised, unsupervised, and reinforcement learning models (Abdi, 2016). The input and output variables are clearly defined in supervised learning. In contrast, the machine is left to discover the input and output variables on its own from the data in the case of unsupervised learning. In reinforcement learning, the machine learns from its own experience by using a feedback approach. In this study, only supervised-based models relating to the failure probability of water pipes are reviewed since no studies are found on the other two classes of ML. These models include artificial neural network (ANN), support vector machine (SVM), k-nearest neighbors (k-NN), gradient boosting-based tree (GBT), extreme gradient boosting (XGBoost), and random forest (RF). k-NN can be used as a supervised or unsupervised machine learning algorithm; however, it has been used as a supervised learning algorithm in the studies reviewed in this section. It should be noted that these algorithms can be used to solve regression and classification problems (Abdelmageed et al., 2022). However, the reviewed studies have approached the failure probability of water pipes as a classification problem using historical data.

Fan et al. (2022) used five ML algorithms to classify each pipe in a network as either broken or intact. These ML algorithms are lightGBM, ANN, k-NN, SVM, and LR. The five models produced output for each pipe which ranged from 0 to 1, denoting the failure probability of a pipe. 13 factors were considered in the modeling, including 11 continuous variables, while the remaining two variables are categorical. While most of the considered factors are correlated with each other's and with the failure of water pipes, none of them was dominant. This strengthens the hypothesis that the failure of water pipes is a complex mechanism and does not

depend on a single factor. Based on the performance indicators such as prediction accuracy and computational efficiency, the lightGBM algorithm was selected as the best model, followed by the ANN model. An ML interpreter, Shapley Additive exPlanations (SHAP), was adopted to investigate the relative importance of each factor. "Interval to the last break," "Cold days," "Pipe length," "Hot days," and "Pipe age" were found as the topmost important factors for pipe failure prediction.

Moreover, (Rifaai et al., 2022) employed LR to predict the failure probability and mean time to water pipe failure in a WDN in Austin, USA. The dataset included 244,830 pipes, which are made of AC, CI, DI, PVC, and others. Their model achieved an accuracy of 80%, AUC of 0.69, and Mathew Coefficient Correlation (MCC) of 0.53. However, the accuracy of the model could be improved by selecting the best hyperparameters that could fit the data well and performing feature (i.e., variable) selection prior to the modeling.

In order to solve the problem of limited historical failure data experienced in some water utilities, Chen et al. (2022) combined the historical data of six utilities to make failure probability predictions. Three algorithms were used: RF, GBT, and XGBoost. For each of the six utilities, four datasets were prepared. The first data set consists of half-historical data of a reference utility, while the second data set represents the full historical data of such utility. The third data set, which is a union of all the utilities' historical data, was prepared in such a way that the pipe material distribution of each of the utilities matched one another. Meanwhile, the fourth dataset is a union of all the utilities' historical data without considering their material distribution consistency. Their results indicated that using data from other utilities for failure probability prediction of a reference utility does not improve the prediction accuracy of such a model. Additionally, for the first and second datasets, the prediction accuracy of pipe failure is not associated with the quantity of the data but rather the quality of the explanatory variables. Overall, RF had the highest prediction accuracy. Similarly, although not absolutely enough, the

study of Raspati et al. (2022), which used RF for failure prediction, noted that an advantage of RF over other "black box" machine learning algorithms is its interpretability and simplicity. The interpretability here means RF models could be visualized, and the relationship between the explanatory variables with respect to the prediction could be seen.

Furthermore, GBT, SVM, ANN, and Bayes were employed to predict the failure probability of water pipes (Giraldo-González & Rodríguez, 2020). The models incorporated 11 input variables. The performance of the algorithms was assessed using the prediction accuracy formula (Equation A.8), recall value on the confusion matrix, and AUC of receiver operating characteristic (ROC). It was observed that the accuracy metric does not give reliable performance evaluations since the data consists of non-failed pipes than failed ones (i.e., imbalanced data). Hence, the classifiers, especially ANN, were majorly able to predict non-failed pipes correctly. Therefore, it is better to assess the performance of ML classifier-based models using the cofunction matrix and AUC. Overall, GBT and SVM achieved the best performance in terms of failure prediction.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$ (A.8)

where TP (true positive) and TN (true negative) represent the number of correctly classified pipes as failed and non-failed, respectively. On the other hand, FP and FN represent the number of pipes that are wrongly classified as failed and none-failed, respectively.

APPENDIX B
Questionnaire design for implementing objective two

Questionnaire Survey

RELATIONSHIP BETWEEN FACTORS INFLUENCING WATER PIPES FAILURE

Do you consent to participate in this survey?

☐   Yes, I consent

☐   No, I do not consent

SECTION 1: DEMOGRAPHICS OF THE RESPONDENTS

What is your current job role?

☐   Academic

☐   Consultant

☐   Manager

☐   Engineer

☐   Site supervisor

☐   Government official

How many years of experience do you have in water infrastructure industry?

☐   < 5 years

☐   5-10 years

☐   11-15 years

☐   16-20 years

☐   > 20 years

What is your highest academic qualification?

☐   Diploma

☐   Bachelor

☐   Master

☐   Doctorate

What is your degree major?

☐   Civil Engineering

☐   Mechanical Engineering

☐   Project Management

☐   Production and Operation

☐   Environmental Engineering

☐   Others

Kindly state the country where you are currently employed

SECTION 2: ASSESSMENT OF WATER PIPE FAILURE FACTORS

Pipe-related factors influencing water pipe failure

What is the degree to which the following factors affect water pipe failure? (1 - Very low influence, 2 - Low influence, 3 - Moderate influence, 4 - High influence, 5 - Very high influence)

	Very low influence	Low influence	Moderate influence	High influence	Very high influence
Pipe age	☐	☐	☐	☐	☐
Buried depth	☐	☐	☐	☐	☐
Pipe diameter	☐	☐	☐	☐	☐
Pipe length	☐	☐	☐	☐	☐
Pipe material	☐	☐	☐	☐	☐
Protection efficiency	☐	☐	☐	☐	☐
Pipe wall thickness	☐	☐	☐	☐	☐

External-related factors influencing water pipe failure

What is the degree to which the following factors affect water pipe failure? (1 - Very low influence, 2 - Low influence, 3 - Moderate influence, 4 - High influence, 5 - Very high influence)

	Very low influence	Low influence	Moderate influence	High influence	Very high influence
Temperature	☐	☐	☐	☐	☐
Chemical substance	☐	☐	☐	☐	☐
Frost	☐	☐	☐	☐	☐
Land use	☐	☐	☐	☐	☐
Lime leaching	☐	☐	☐	☐	☐
Microbiologically induced corrosion	☐	☐	☐	☐	☐
Precipitation	☐	☐	☐	☐	☐
Traffic	☐	☐	☐	☐	☐

Soil-related factors influencing water pipe failure

What is the degree to which the following factors affect water pipe failure? (1 - Very low influence, 2 - Low influence, 3 - Moderate influence, 4 - High influence, 5 - Very high influence)

	Very low influence	Low influence	Moderate influence	High influence	Very high influence
Bedding condition	☐	☐	☐	☐	☐
Soil moisture	☐	☐	☐	☐	☐
Soil pH	☐	☐	☐	☐	☐
Soil resistivity	☐	☐	☐	☐	☐
Soil type	☐	☐	☐	☐	☐
Soil aeration	☐	☐	☐	☐	☐

Operation-related factors influencing water pipe failure

What is the degree to which the following factors affect water pipe failure? (1 - Very low influence, 2 - Low influence, 3 - Moderate influence, 4 - High influence, 5 - Very high influence)

	Very low influence	Low influence	Moderate influence	High influence	Very high influence
Installation and pump operation	☐	☐	☐	☐	☐
Internal pressure	☐	☐	☐	☐	☐
Maintenance practices	☐	☐	☐	☐	☐
Number of leaks	☐	☐	☐	☐	☐
Water acidity	☐	☐	☐	☐	☐
Water alkalinity	☐	☐	☐	☐	☐
Water hammer	☐	☐	☐	☐	☐
Water temperature	☐	☐	☐	☐	☐
Water velocity	☐	☐	☐	☐	☐

SECTION 3: ASSESSMENT OF WATER PIPE FAILURE modes

Failure modes in water pipes

What is the degree of severity of the following failure modes in water pipes? (1 - Low severity, 2- Moderate severity, 3- High severity)

	Low severity	Moderate severity	High severity
Circumferential cracking	☐	☐	☐
Longitudinal cracking	☐	☐	☐
Bell splitting	☐	☐	☐
Corrosion splitting	☐	☐	☐
Blown-out hole	☐	☐	☐

APPENDIX C
Supplementary results of POF model

Experiment A was conducted to compare its result with that of Experiment 1. In Experiment A, the default hyperparameter for LR in a Python's library was employed. Details of Experiment 1 can be found in section 3.4.2.

Table C1: Confusion matrix of experiment A

For non-scaled data		
	Predicted Failure	Predicted Intact
True Failure	TF = 6567	FI = 3731
True Intact	FF = 2226	TI = 10321
For normalized data		
	Predicted Failure	Predicted Intact
True Failure	TF = 6897	FI = 3401
True Intact	FF = 1561	TI = 10986
For standardized data		
	Predicted Failure	Predicted Intact
True Failure	TF = 6972	FI = 3326
True Intact	FF = 1439	TI = 11108

Table C2: Evaluation metrics of experiment A

Data transformation	Accuracy	Precision	Recall	F1 score	AUC
Non-scaled data	0.739	0.746	0.637	0.687	0.832
Normalized data	0.782	0.815	0.669	0.734	0.855
Standardized data	0.791	0.828	0.677	0.745	0.861

Table C3 shows the confusion matrix of experiment 2 explained in section 4.7.

Table C3: Confusion matrix of experiment 2

For non-scaled data using accuracy as the fitness function		
	Predicted Failure	Predicted Intact
True Failure	TF = 8349	FI = 1949
True Intact	FF = 883	TI = 11664
For non-scaled data using f1 score as the fitness function		
	Predicted Failure	Predicted Intact
True Failure	TF = 8367	FI = 1931
True Intact	FF = 862	TI = 11685
For normalized data using accuracy as the fitness function		
	Predicted Failure	Predicted Intact
True Failure	TF = 8392	FI = 1906
True Intact	FF = 758	TI = 11789
For normalized data using f1 score as the fitness function		
	Predicted Failure	Predicted Intact

True Failure	TF = 8386	FI = 1912
True Intact	FF = 664	TI = 11883

For standardized data using accuracy as the fitness function

	Predicted Failure	Predicted Intact
True Failure	TF = 8401	FI = 1897
True Intact	FF = 651	TI = 11896

For standardized data using f1 score as the fitness function

	Predicted Failure	Predicted Intact
True Failure	TF = 8392	FI = 1906
True Intact	FF = 636	TI = 11911

APPENDIX D

Supplementary results of POL and POB models

Tables D1 and D2 present the confusion matrix for the base-DL models for POL and POB, respectively, while Tables D3 and D4 show the confusion matrix for the optimized-DL models for POL and POB, respectively.

Table D1: Confusion matrix of base-DL models for predicting probability of leak

Data scaling	Models	TL/B	TI	FL/B	FI
Non-scaled	DNN	8005	249582	5711	1297
	CNN	8054	249985	5308	1248
	TabNet	7830	247832	7461	1472
Normalized	DNN	8206	250670	4617	1096
	CNN	8498	251075	4218	804
	TabNet	7764	250082	5211	1538
Standardized	DNN	8508	252075	3218	794
	CNN	8593	252032	3261	709
	TabNet	8398	252140	32153	904

Table D2: Confusion matrix of base-DL models for predicting probability of burst

Data scaling	Models	TL/B	TI	FL/B	FI
Non-scaled	DNN	275	254350	840	26
	CNN	278	254430	760	23
	TabNet	282	254426	764	19
Normalized	DNN	285	254617	573	16
	CNN	286	254695	495	15
	TabNet	286	254548	442	15
Standardized	DNN	286	254539	451	15
	CNN	290	254810	380	11
	TabNet	291	2544744	446	10

Table D3: Confusion matrix of optimized-DL models for predicting probability of leak

Data scaling	Models	TL/B	TI	FL/B	FI
Non-scaled	DNN	8709	252644	2649	593
	CNN	8834	253426	1867	468
	TabNet	8898	253899	1394	404
Normalized	DNN	8876	253021	2272	426
	CNN	8799	253101	2192	503
	TabNet	8762	253005	2288	540
Standardized	DNN	9018	253886	1407	284
	CNN	9097	254011	1282	205
	TabNet	8725	253100	2193	577

Table D4: Confusion matrix of optimized-DL models for predicting probability of burst

Data scaling	Models	TL/B	TI	FL/B	FI
Non-scaled	DNN	293	254983	207	8
	CNN	295	254525	665	6
	TabNet	295	255085	105	6
Normalized	DNN	294	255083	107	7
	CNN	297	255107	83	4
	TabNet	296	255108	82	5
Standardized	DNN	296	254999	191	5
	CNN	295	255088	102	6
	TabNet	297	255061	129	4

APPENDIX E

Supplementary results of COF models

Table E1: Individual class performance metrics for the base models

| Models	Class	Precision	Recall	F1 score	AUC
XGBoost	C01	0.45	0.43	0.44	0.83
	C02	0.90	0.91	0.90	0.89
	C03	0.28	0.31	0.30	0.77
	C04	0.57	0.52	0.55	0.87
Adaboost	C01	0.44	0.44	0.44	0.82
	C02	0.89	0.91	0.90	0.88
	C03	0.32	0.27	0.29	0.71
	C04	0.58	0.53	0.55	0.86
LightGBM	C01	0.46	0.50	0.47	0.83
	C02	0.90	0.91	0.90	0.89
	C03	0.29	0.35	0.32	0.77
	C04	0.62	0.50	0.56	0.87
CatBoost	C01	0.40	0.44	0.42	0.81
	C02	0.90	0.90	0.90	0.89
	C03	0.26	0.30	0.28	0.76
	C04	0.54	0.46	0.49	0.89
RF	C01	0.42	0.45	0.43	0.82
	C02	0.90	0.92	0.91	0.89
	C03	0.26	0.34	0.29	0.76
	C04	0.61	0.42	0.50	0.86
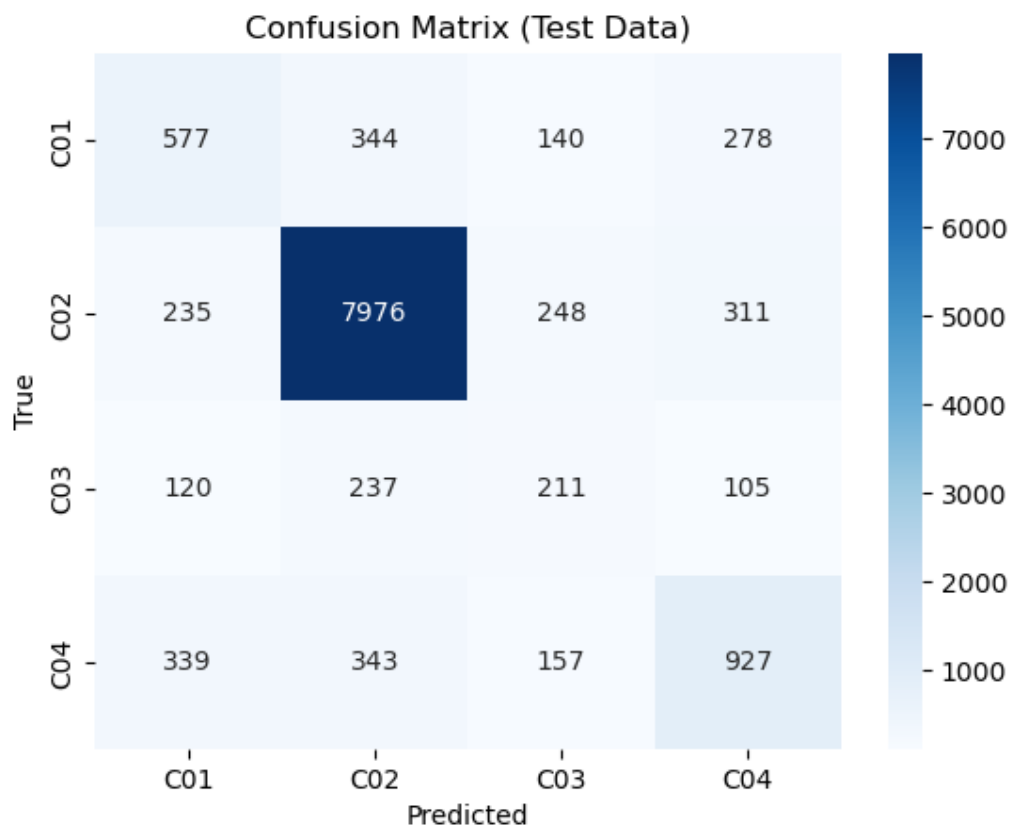
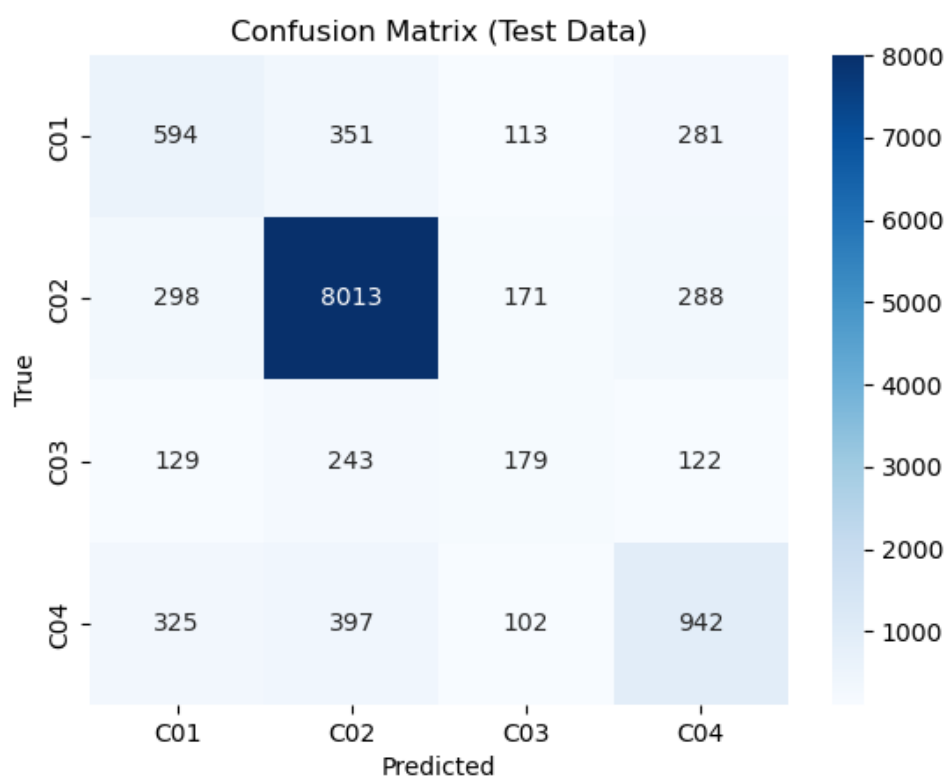Figure E1(a): Confusion matrix for XGBoost model

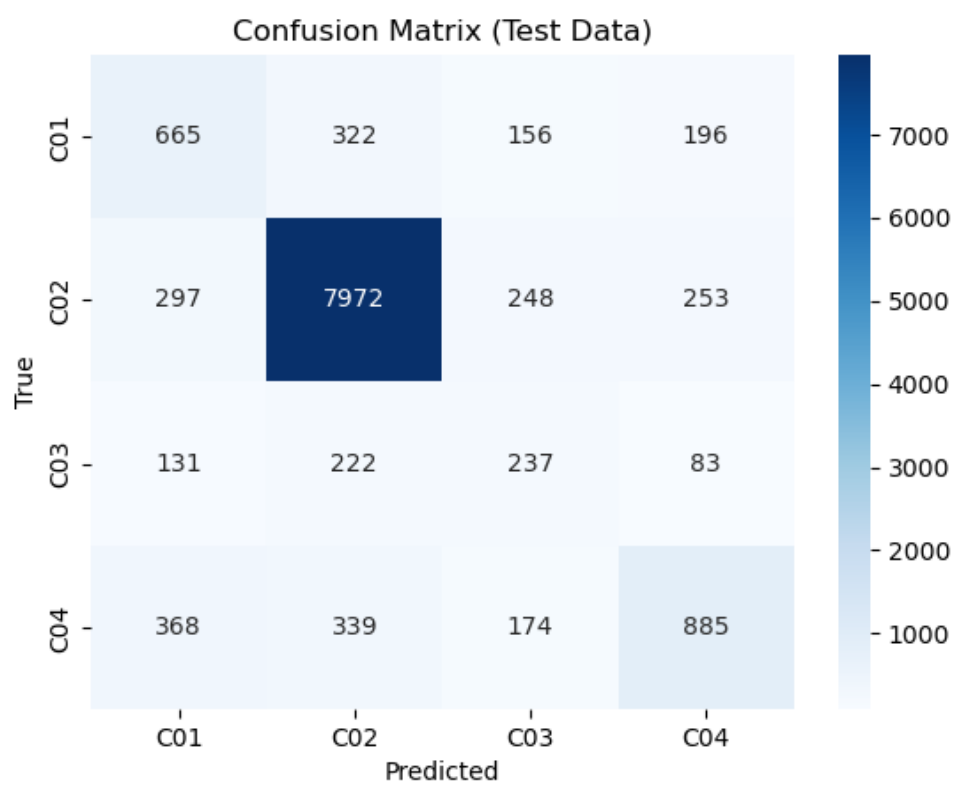Figure E1(b): Confusion matrix for Adaboost model

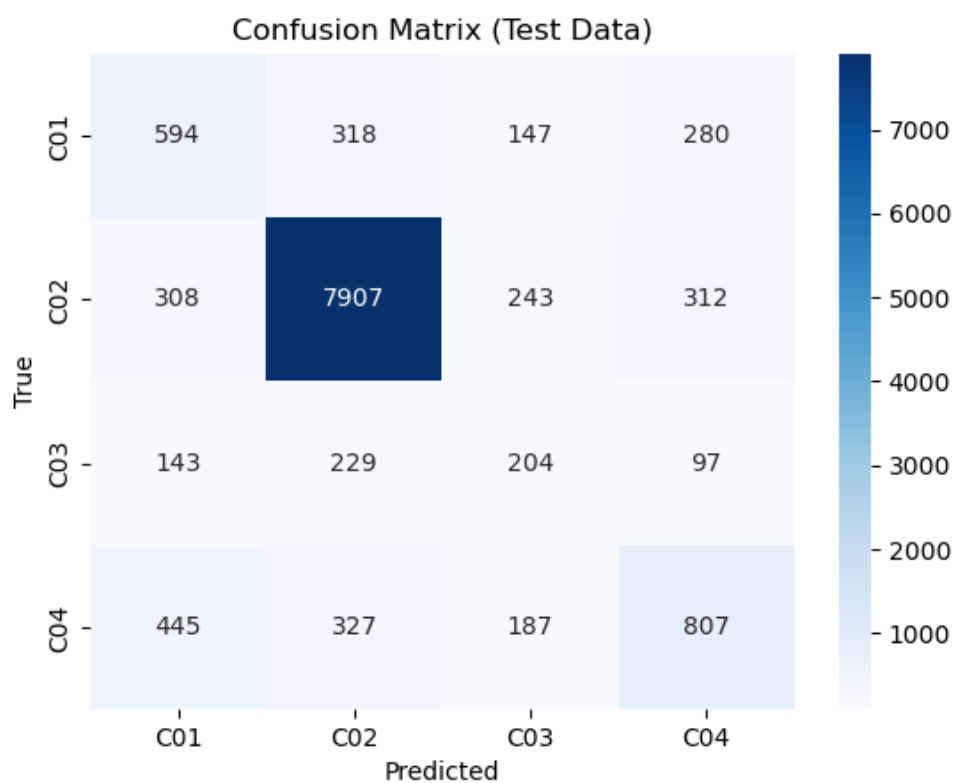Figure E1(c): Confusion matrix for LightGBM model

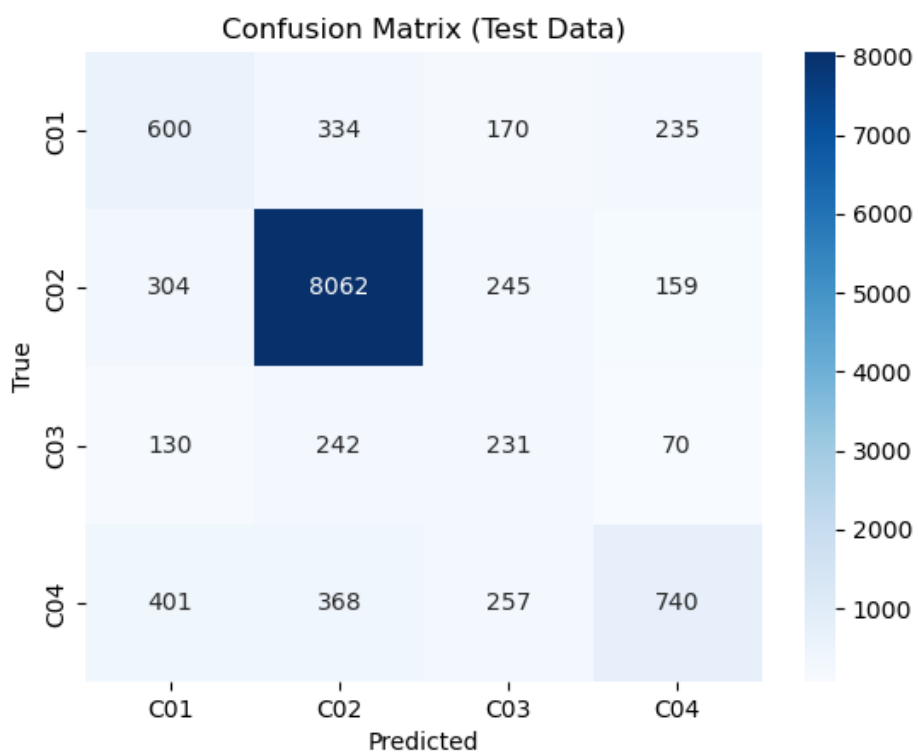Figure E1(d): Confusion matrix for CatBoost model

Figure E1(e): Confusion matrix for Random Forest model

Table E2: Individual class performance metrics for the optimized metrics

| Models	Class	Precision	Recall	F1 score	AUC
XGBoost + TPE	C01	0.59	0.46	0.52	0.87
	C02	0.88	0.95	0.91	0.91
	C03	0.45	0.29	0.35	0.82
	C04	0.68	0.63	0.65	0.90
Adaboost + TPE	C01	0.48	0.42	0.45	0.81
	C02	0.88	0.94	0.91	0.86
	C03	0.38	0.26	0.31	0.73
	C04	0.59	0.55	0.57	0.86
LightGBM + TPE	C01	0.56	0.46	0.50	0.86
	C02	0.88	0.95	0.91	0.91
	C03	0.41	0.30	0.35	0.82
	C04	0.66	0.58	0.62	0.90
CatBoost + TPE	C01	0.56	0.46	0.51	0.87
	C02	0.89	0.94	0.91	0.91
	C03	0.41	0.29	0.34	0.81
	C04	0.65	0.61	0.63	0.89
RF + TPE	C01	0.56	0.49	0.52	0.86
	C02	0.89	0.95	0.92	0.91
	C03	0.43	0.25	0.32	0.81
	C04	0.66	0.61	0.64	0.89
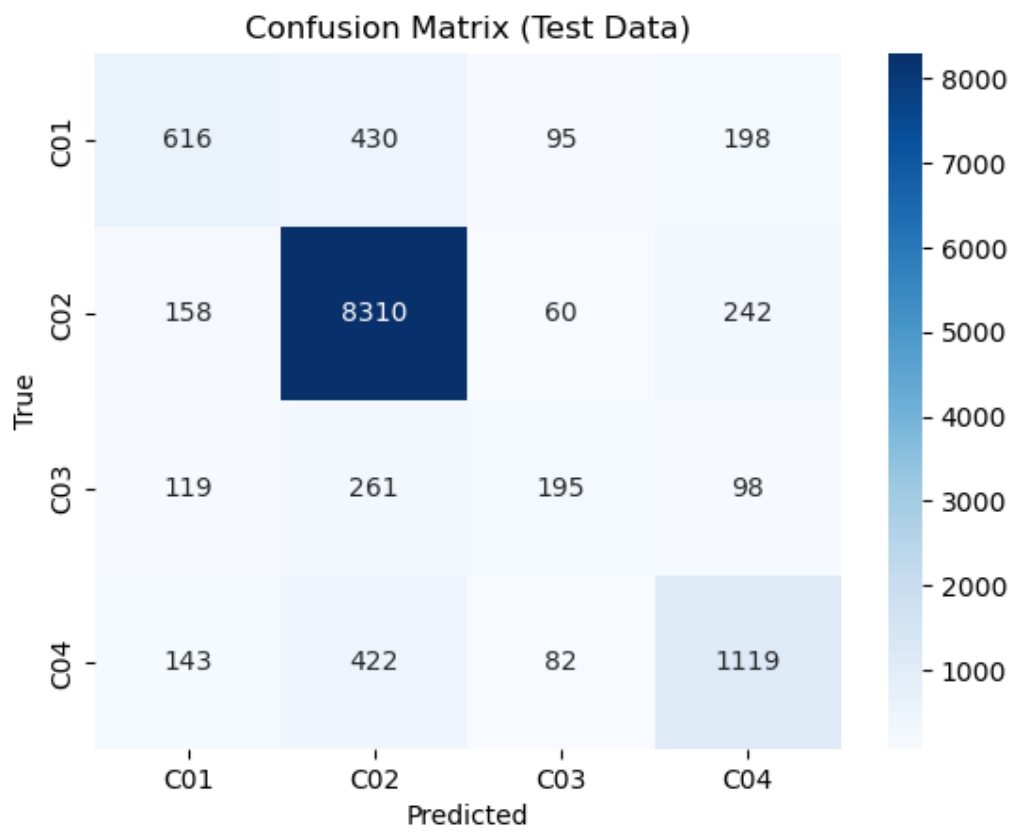
Figure E2(a): Confusion matrix for XGB+TPE model
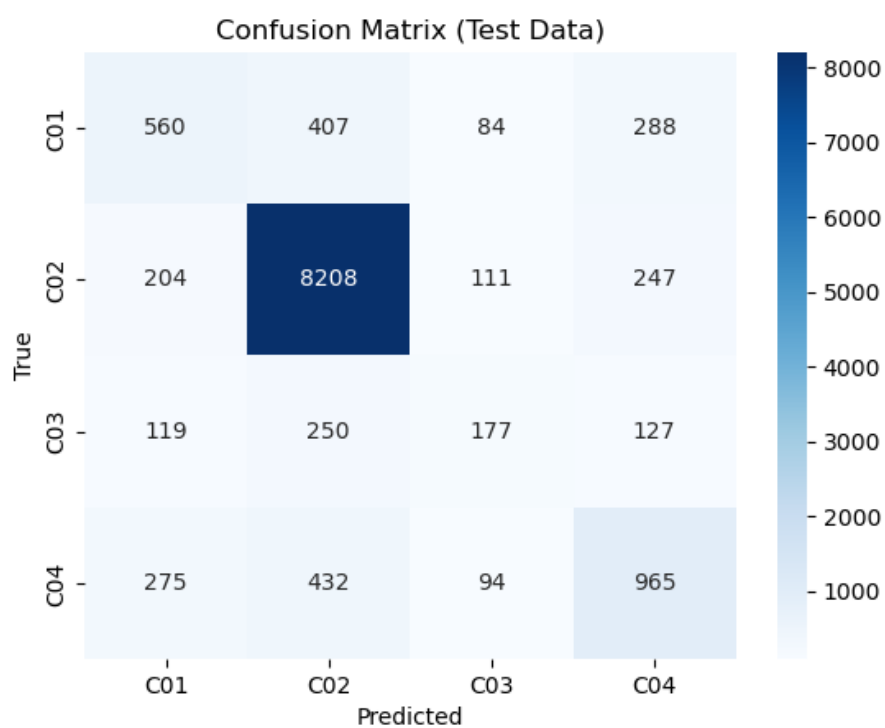
Figure E2(b): Confusion matrix for Adaboost + TPE model

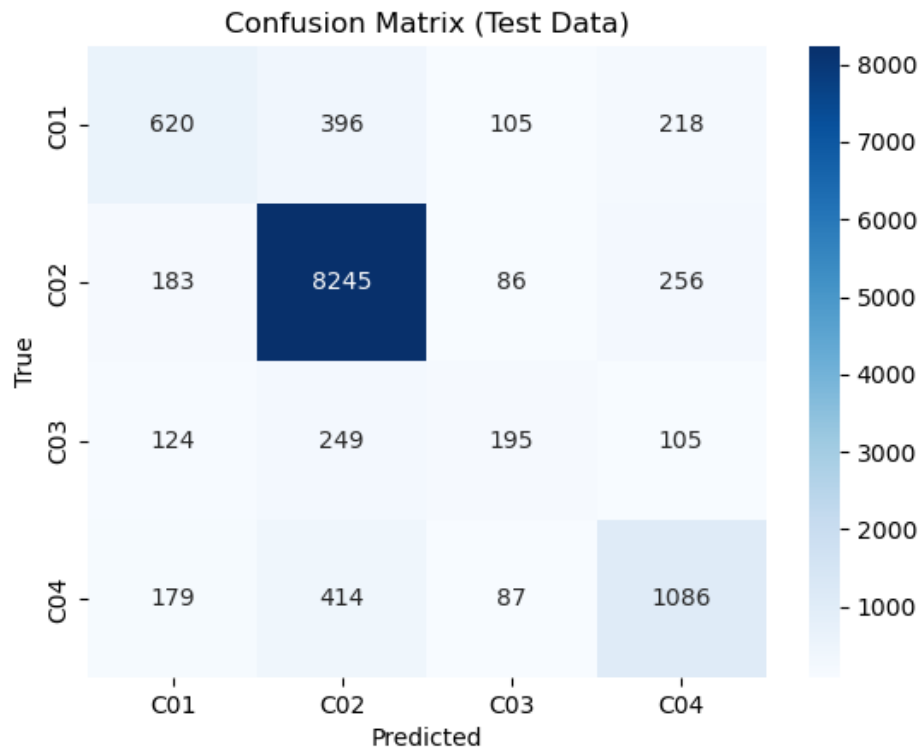Figure E2(c): Confusion matrix for LightGBM + TPE model

Figure E2(d): Confusion matrix for CatBoost + TPE model

Figure E2(e): Confusion matrix for RF + TPE model

REFERENCES

Abdelmageed, S., Tariq, S., Boadu, V., & Zayed, T. (2022). Criteria-based critical review of artificial intelligence applications in water-leak management. *Environmental Reviews*, *April*, 1–18. https://doi.org/10.1139/er-2021-0046

Abdi, A. (2016). *Three types of Machine Learning Algorithms* (Issue November). https://doi.org/10.13140/RG.2.2.26209.10088

Adabre, M. A., Chan, A. P. C., Edwards, D. J., & Adinyira, E. (2021). Assessing critical risk factors (CRFs) to sustainable housing: The perspective of a sub-Saharan African country. *Journal of Building Engineering*, *41*(March), 102385. https://doi.org/10.1016/j.jobe.2021.102385

Affolter, C., Barbezat, M., Piskoty, G., Neuner, O., & Terrasi, G. (2018). Failure of a sag water pipe triggered by aging of the GFRP composite relining. *Engineering Failure Analysis*, *84*(November 2016), 358–370. https://doi.org/10.1016/j.engfailanal.2017.09.017

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2623–2631. https://doi.org/10.1145/3292500.3330701

Al-Adeeb, A. M., & Matti, M. A. (1984). Leaching corrosion of asbestos cement pipes. *International Journal of Cement Composites and Lightweight Concrete*, *6*(4), 233–240. https://doi.org/10.1016/0262-5075(84)90018-6

Al-Ali, A. M., Laurent, J., & Dulot, J. P. (2020). Developing deterioration prediction model for the potable water pipes renewal plan – case of Jubail industrial city, KSA. *Desalination and Water Treatment*, *176*(March), 324–332. https://doi.org/10.5004/dwt.2020.25539

Al-Zahrani, M., Abo-Monasar, A., & Sadiq, R. (2016). Risk-based prioritization of water main failure using fuzzy synthetic evaluation technique. *Journal of Water Supply: Research and Technology - AQUA*, *65*(2), 145–161. https://doi.org/10.2166/aqua.2015.051

Ali, A. H., Kineber, A. F., Elyamany, A., Ibrahim, A. H., & Daoud, A. O. (2023). Modelling the role of modular construction's critical success factors in the overall sustainable success of Egyptian housing projects. *Journal of Building Engineering*, *71*(April), 106467. https://doi.org/10.1016/j.jobe.2023.106467

Almheiri, Z., Meguid, M., & Zayed, T. (2020). Intelligent Approaches for Predicting Failure of Water Mains. *Journal of Pipeline Systems Engineering and Practice*, *11*(4), 04020044. https://doi.org/10.1061/(asce)ps.1949-1204.0000485

Almheiri, Z., Meguid, M., & Zayed, T. (2021). Failure modeling of water distribution pipelines using meta-learning algorithms. *Water Research*, *205*(September), 117680. https://doi.org/10.1016/j.watres.2021.117680

Alshurideh, M., Al Kurdi, B., Salloum, S. A., Arpaci, I., & Al-Emran, M. (2020). Predicting the actual use of m-learning systems: a comparative approach using PLS-SEM and

machine learning algorithms. *Interactive Learning Environments*, *0*(0), 1–15. https://doi.org/10.1080/10494820.2020.1826982

Amiri-Ardakani, Y., & Najafzadeh, M. (2021). Pipe Break Rate Assessment While Considering Physical and Operational Factors: A Methodology based on Global Positioning System and Data-Driven Techniques. *Water Resources Management*, *0123456789*. https://doi.org/10.1007/s11269-021-02911-6

Andreou, S. A., Marks, D. H., & Clark, R. M. (1987). A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory. *Advances in Water Resources*, *10*(1), 2–10. https://doi.org/10.1016/0309-1708(87)90002-9

Andriamamonjy, A., Saelens, D., & Klein, R. (2019). A combined scientometric and conventional literature review to grasp the entire BIM knowledge and its integration with energy simulation. *Journal of Building Engineering*, *22*, 513–527. https://doi.org/10.1016/j.jobe.2018.12.021

Arık, S., & Pfister, T. (2021). TabNet: Attentive Interpretable Tabular Learning. *35th AAAI Conference on Artificial Intelligence, AAAI 2021*, *8A*, 6679–6687. https://doi.org/10.1609/aaai.v35i8.16826

Arriba-Rodriguez, L. De, Villanueva-Balsera, J., Ortega-Fernandez, F., & Rodriguez-Perez, F. (2018). Methods to evaluate corrosion in buried steel structures: A review. *Metals*, *8*(5). https://doi.org/10.3390/met8050334

Aryai, V., Baji, H., & Mahmoodian, M. (2022). Failure assessment of corrosion affected pipeline networks with limited failure data availability. *Process Safety and Environmental Protection Journal Homepage*, *157*, 306–319. https://doi.org/10.1016/j.psep.2021.11.024

Aryai, V., Baji, H., Mahmoodian, M., & Li, C. Q. (2020). Time-dependent finite element reliability assessment of cast-iron water pipes subjected to spatio-temporal correlated corrosion process. *Reliability Engineering and System Safety*, *197*(July 2019), 106802. https://doi.org/10.1016/j.ress.2020.106802

Aryai, V., & Mahmoodian, M. (2017). Spatial-temporal reliability analysis of corroding cast iron water pipes. *Engineering Failure Analysis*, *82*(August), 179–189. https://doi.org/10.1016/j.engfailanal.2017.08.017

Auma, R., Musaazi, I. G., Tumutungire, M. D., & Sempewo, J. I. (2023). Logistic pipe failure prediction models for an urban water distribution network in the developing world: a case study of Kampala water, Uganda. *Water Practice and Technology*, *18*(1), 264–273. https://doi.org/10.2166/wpt.2022.159

AWWA. (2002). *Effect of water age on distribution system water quality*.

Baharun, N., Razi, N. F. M., Masrom, S., Yusri, N. A. M., & Abd Rahman, A. S. (2022). *Auto Modelling for Machine Learning: A Comparison Implementation between RapidMiner and Python*. *12*(05), 1–7. https://doi.org/10.46338/ijetae0522

Barton, N. A., Farewell, T. S., Hallett, S. H., & Acland, T. F. (2019). Improving pipe failure predictions: Factors effecting pipe failure in drinking water networks. *Water Research*, *164*. https://doi.org/10.1016/j.watres.2019.114926

Beig Zali, R., Latifi, M., Javadi, A. A., & Farmani, R. (2024). Semisupervised Clustering Approach for Pipe Failure Prediction with Imbalanced Data Set. *Journal of Water*

Resources Planning and Management, 150(2), 1–15.
https://doi.org/10.1061/jwrmd5.wreng-6263

Bello, I. T., Taiwo, R., Esan, O. C., Adegoke, A. H., Ijaola, A., Li, Z., Zhao, S., Chen, W., Shao, Z., & Ni, M. (2023). AI-enabled Materials Discovery for Advanced Ceramic Electrochemical Cells. *Energy and AI*.

Ben Seghier, M. E. A., Keshtegar, B., Tee, K. F., Zayed, T., Abbassi, R., & Trung, N. T. (2020). Prediction of maximum pitting corrosion depth in oil and gas pipelines. *Engineering Failure Analysis*, 112(March), 104505. https://doi.org/10.1016/j.engfailanal.2020.104505

Ben Seghier, M. E. A., Mustaffa, Z., & Zayed, T. (2022). Reliability assessment of subsea pipelines under the effect of spanning load and corrosion degradation. *Journal of Natural Gas Science and Engineering*, 102(November 2021), 104569. https://doi.org/10.1016/j.jngse.2022.104569

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.

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(4). https://doi.org/10.3390/w10040411

BS 9295. (2010). *Guide to the structural design of buried pipelines*.

Burn, S., Davis, P., & Schiller, T. (2005). Long-term performance prediction for PVC pipes. *AWWA Research Foundation*. https://espace.library.uq.edu.au/view/UQ:193983

Carrión, A., Solano, H., Gamiz, M. L., & Debón, A. (2010). Evaluation of the Reliability of a Water Supply Network from Right-Censored and Left-Truncated Break Data. *Water Resources Management*, 24(12), 2917–2935. https://doi.org/10.1007/s11269-010-9587-y

Chen, C., & Seo, H. (2023). Prediction of rock mass class ahead of TBM excavation face by ML and DL algorithms with Bayesian TPE optimization and SHAP feature analysis. *Acta Geotechnica*, 0123456789, 3825–3848. https://doi.org/10.1007/s11440-022-01779-z

Chen, L., Duan, Z., Li, W., Yang, G., Jia, J., Ma, L., Li, M., Zhong, Y., Sun, H., Jiang, C., & Chen, Y. (2021). Physical Properties and Boundary Influence of Singularity in Fluid Pipelines Based on Vibration Wave's Transmission Characteristics. *Journal of Pipeline Systems Engineering and Practice*, 13(1), 04021066. https://doi.org/10.1061/(asce)ps.1949-1204.0000608

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-Augu, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, T. Y.-J., Vladeanu, G., Yazdekhasti, S., & Daly, C. M. (2022). Performance Evaluation of Pipe Break Machine Learning Models Using Datasets from Multiple Utilities. *Journal of Infrastructure Systems*, 28(2). https://doi.org/10.1061/(asce)is.1943-555x.0000683

Chik, L., Albrecht, D., & Kodikara, J. (2017). Estimation of the Short-Term Probability of Failure in Water Mains. *Journal of Water Resources Planning and Management*, 143(2), 04016075. https://doi.org/10.1061/(asce)wr.1943-5452.0000730

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In *G.*

A. Marcoulides (Ed.) (pp. 295–336). Lawrence Erlbaum Associates Publishers.

Clair, A. M., & Sinha, S. (2014). Development of a Standard Data Structure for Predicting the Remaining Physical Life and Consequence of Failure of Water Pipes. *Journal of Performance of Constructed Facilities*, *28*(1), 191–203. https://doi.org/10.1061/(asce)cf.1943-5509.0000384

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *11*(2), 155–159.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Cooper, N. R., Blakey, G., Sherwin, C., Ta, T., Whiter, J. T., & Woodward, C. A. (2000). The use of GIS to develop a probability-based trunk mains burst risk model. *Urban Water*, *2*(2), 97–103. https://doi.org/10.1016/S1462-0758(00)00047-9

Dao, V. H., Ryu, H. K., & Yoon, K. B. (2021). Leak failure at the TP316L welds of a water pipe caused by microbiologically influenced corrosion. *Engineering Failure Analysis*, *122*(November 2020), 105244. https://doi.org/10.1016/j.engfailanal.2021.105244

Davis, P., De Silva, D., Marlow, D., Moglia, M., Gould, S., & Burn, S. (2008). Failure prediction and optimal scheduling of replacements in asbestos cement water pipes. *Journal of Water Supply: Research and Technology - AQUA*, *57*(4), 239–252. https://doi.org/10.2166/aqua.2008.035

Davis, P., Moglia, M., Burn, S., & Farlie, M. (2004). Estimating failure probability from condition assessment of critical cast iron water mains. *Proceedings 6th National Conference, Australasian Society of Trenchless Technology (ASTT)*, 1–9.

Dawood, T., Elwakil, E., Novoa, H. M., & Gárate Delgado, J. F. (2021). Toward urban sustainability and clean potable water: Prediction of water quality via artificial neural networks. *Journal of Cleaner Production*, *291*. https://doi.org/10.1016/j.jclepro.2020.125266

De-Silva, D., Moglia, P., Davis, P., & Burn, S. (2006). Condition assessment to estimate failure rates in buried metallic pipelines. *Journal of Water Supply: Research and Technology - AQUA*, *55*(7–8), 453–459. https://doi.org/10.2166/aqua.2006.053

De Leon, D., & Macías, O. F. (2005). Effect of spatial correlation on the failure probability of pipelines under corrosion. *International Journal of Pressure Vessels and Piping*, *82*(2), 123–128. https://doi.org/10.1016/j.ijpvp.2004.07.018

Debón, A., Carrión, A., Cabrera, E., & Solano, H. (2010). Comparing risk of failure models in water supply networks using ROC curves. *Reliability Engineering and System Safety*, *95*(1), 43–48. https://doi.org/10.1016/j.ress.2009.07.004

Debrah, C., Chan, A. P. C., & Darko, A. (2022). Artificial intelligence in green building. *Automation in Construction*, *137*(January), 104192. https://doi.org/10.1016/j.autcon.2022.104192

Demissie, G., Tesfamariam, S., & Sadiq, R. (2016). Considering Soil Parameters in Prediction of Remaining Service Life of Metallic Pipes: Bayesian Belief Network Model. *Journal of Pipeline Systems Engineering and Practice*, *7*(2), 04015028. https://doi.org/10.1061/(asce)ps.1949-1204.0000229

Demšar, J., & Zupan, B. (2013). Orange: Data mining fruitful and fun - A historical perspective. *Informatica (Slovenia)*, *37*(1), 55–60.

Doyle, G., Seica, M. V., & Grabinsky, M. W. F. (2003). The role of soil in the external corrosion of cast iron water mains in Toronto, Canada. *Canadian Geotechnical Journal*, *40*(2), 225–236. https://doi.org/10.1139/t02-106

Ductile Iron Pipe Research Association. (2017). *Corrosion Control Polyethylene encasement*.

Economou, T., Bailey, T. C., & Kapelan, Z. (2007). Bayesian modelling of time aggregated water pipe bursts with a zero-inflated , non-homogeneous Poisson process. *Proceedings of the 22nd International Workshop on Statistical Modelling*, 227–232.

Ellison, D., & Spencer, D. (2016). The True Causes of AC Pipe Failures — According to the Data. *Pipelines*, 637–647.

Elsawah, H., Bakry, I., & Moselhi, O. (2016). Decision Support Model for Integrated Risk Assessmentand Prioritization of Intervention Plans of Municipal Infrastructure. *Journal of Pipeline Systems Engi- Neering and Practice*, *7*(4).

Fan, X., Wang, X., Zhang, X., & ASCE Xiong (Bill) Yu, P. E. F. (2022). Machine learning based water pipe failure prediction: The effects of engineering, geology, climate and socio-economic factors. *Reliability Engineering and System Safety*, *219*(July 2021), 108185. https://doi.org/10.1016/j.ress.2021.108185

Fan, X., Wang, X., Zhang, X., Xiong, P. E. F. A., & Yu, B. (2022). Machine learning based water pipe failure prediction : The effects of engineering , geology , climate and socio-economic factors. *Reliability Engineering and System Safety*, *219*(November 2021), 108185. https://doi.org/10.1016/j.ress.2021.108185

Fan, X., Zhang, X., & Yu, X. B. (2023). Uncertainty quantification of a deep learning model for failure rate prediction of water distribution networks. *Reliability Engineering and System Safety*, *236*(January 2022), 109088. https://doi.org/10.1016/j.ress.2023.109088

Fares, H., & Zayed, T. (2010). Hierarchical Fuzzy Expert System for Risk of Failure of Water Mains. *Journal of Pipeline Systems Engineering and Practice*, *1*(1), 53–62. https://doi.org/10.1061/(asce)ps.1949-1204.0000037

Farewell, T. S., Hallett, S. H., & Truckell, I. G. (2012). Soil and climatic causes of water mains infrastructure bursts. *NSRI Research Report. NSRI, Cranfield University, UK.*, *111pp.*

Farh, H. M. H., Ben Seghier, M. E. A., Taiwo, R., & Zayed, T. (2023). Analysis and Ranking of Corrosion Causes for Water Pipelines: A Critical Review. *Npj Clean Water*.

Farshad, M. (2006). Plastic pipe systems. In *Elsevier*. https://doi.org/10.1016/b978-185617496-1/50002-1

Fiedler, S., Vepraskas, M. J., & Richardson, J. L. (2007). Soil Redox Potential: Importance, Field Measurements, and Observations. *Advances in Agronomy*, *94*(06), 1–54. https://doi.org/10.1016/S0065-2113(06)94001-2

Folkman, S. (2018). *Water main break rates in the USA and Canada: A comprehensive study* (Issue March).

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39. https://doi.org/10.2307/3151312

Francisque, A., Rodriguez, M. J., Sadiq, R., & Miranda, L. F. (2009). *Prioritizing monitoring locations in a water distribution network : a fuzzy risk approach and Franc¸ois Proulx*. https://doi.org/10.2166/aqua.2009.011

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139.

Friedl, F., Moderl, M., Rauch, W., Liu, Q., Schrotter, S., & Fuchs-Hanusch, D. (2012). Failure Propagation for Large-Diameter Transmission Water Mains Using Dynamic Failure Risk Index. *World Environmental and Water Resources Congress 2012: Crossing Boundaries*, 2247–2252.

Furxhi, I., Murphy, F., Mullins, M., & Poland, C. A. (2019). Machine learning prediction of nanoparticle in vitro toxicity: A comparative study of classifiers and ensemble-classifiers using the Copeland Index. *Toxicology Letters*, *312*(May), 157–166. https://doi.org/10.1016/j.toxlet.2019.05.016

Gabriel, L. (2011). *Corrugated polyethylene pipe design manual and installation guide*.

Gao, Y. (2017). *Systematic Review for Water Network Failure Models and Cases*. http://scholarworks.uark.edu/etdhttp://scholarworks.uark.edu/etd/2608

Garcia, D. M., Lee, J., Keck, J., Yang, P., & Guzzetta, R. (2019). Utilizing spatiotemporal based business risk exposure to analyze cast iron water main failures in California. *Journal of Water Supply: Research and Technology - AQUA*, *68*(2), 111–120. https://doi.org/10.2166/aqua.2019.120

García, L., Bacon, J., & Campos, C. (2018). *Modelling the fate and transport of faecal bacteria from Modelling the fate and transport of faecal bacteria from sewage overflows : The Dart Estuary case study*. 0–6.

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*(4). https://doi.org/10.3390/W12041153

Gong, J., Lambert, M., Zecchin, A., Simpson, A., Arbon, N., & Kim, Y. Il. (2016). Field study on non-invasive and non-destructive condition assessment for asbestos cement pipelines by time-domain fluid transient analysis. *Structural Health Monitoring*, *15*(1), 113–124. https://doi.org/10.1177/1475921715624505

Grigg, N. S. (2017). Water Supply Pipeline Failures: Investigative Procedures and Data Management. *Journal of Performance of Constructed Facilities*, *31*(6), 06017004. https://doi.org/10.1061/(asce)cf.1943-5509.0001113

Gudergan, S. P., Ringle, C. M., Wende, S., & Will, A. (2008). Confirmatory tetrad analysis in PLS path modeling. *Journal of Business Research*, *61*(12), 1238–1249. https://doi.org/10.1016/j.jbusres.2008.01.012

Gurung, T. R., Stewart, R. A., Beal, C. D., & Sharma, A. K. (2015). Smart meter enabled water end-use demand data: Platform for the enhanced infrastructure planning of contemporary urban water supply networks. *Journal of Cleaner Production*, *87*(1), 642–654. https://doi.org/10.1016/j.jclepro.2014.09.054

Habibian, A. (1994). EFFECT OF TEMPERATURE CHANGES ON WATER MAIN BREAKS. *Journal of Transportation Engineering*, *120*(2), 312–321.

Hair, J. F., Jr., H. G. T. M., Ringle, C. M., & Sarstedt, M. (2014). A primer on partial least squares structural equations modeling (PLS-SEM). Sage Publications. *Journal of Tourism Research*, *6*(2), 211–213.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). Partial Least Squares Structural Equation Modeling: Rigorous Applications, Better Results and Higher Acceptance. *Long Range Planning*, *46*(1–2), 1–12. https://doi.org/10.1016/j.lrp.2013.01.001

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Hair, J., Hollingsworth, C. L., Randolph, A. B., & Chong, A. Y. L. (2017). An updated and expanded assessment of PLS-SEM in information systems research. *Industrial Management and Data Systems*, *117*(3), 442–458. https://doi.org/10.1108/IMDS-04-2016-0130

Hair, J., Sarstedt, M., Hopkins, L., & Kuppelwieser, V. (2014). Partial least squares structural equation modeling (PLS-SEM) An emerging tool in business research. *European Business Review*, *26*(2), 106–121. https://doi.org/10.1108/EBR-10-2013-0128

Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00369-8

Hekmati, N., Rahman, M. M., Gorjian, N., Rameezdeen, R., & Chow, C. W. K. (2020). Relationship between environmental factors and water pipe failure: an open access data study. *SN Applied Sciences*, *2*(11). https://doi.org/10.1007/s42452-020-03581-6

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). *A new criterion for assessing discriminant validity in variance-based structural equation modeling*. 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hou, Y., Lei, D., Li, S., Yang, W., & Li, C. (2016). Experimental Investigation on Corrosion Effect on Mechanical Properties of Buried Metal Pipes. *International Journal of Corrosion*, 1–13.

Hu, Y., & Hubble, D. W. (2007). Factors contributing to the failure of asbestos cement water mains. *Canadian Journal of Civil Engineering*, *34*(5), 608–621. https://doi.org/10.1139/L06-162

Hussein, M., & Zayed, T. (2021). Critical factors for successful implementation of just-in-time concept in modular integrated construction: A systematic review and meta-analysis. *Journal of Cleaner Production*, *284*, 124716. https://doi.org/10.1016/j.jclepro.2020.124716

IAEA. (1993). Applicability of the leak before break concept. In *Iaea-Tecdoc-710*.

Ismaeel, M., & Zayed, T. (2018). Integrated Performance Assessment Model for Water Networks. *Journal of Infrastructure Systems*, *24*(2), 04018005. https://doi.org/10.1061/(asce)is.1943-555x.0000419

Jallouf, S., Schmitt, C., Pluvinage, G., Hadj-Taïeb, E., & Lebienvenu, M. (2011). A probabilistic safety factor for defect assessment of water pipes subjected to water hammer. *Journal of Strain Analysis for Engineering Design*, *46*(1), 14–26. https://doi.org/10.1243/03093247JSA656

Jara-arriagada, C., & Stoianov, I. (2021). Pipe breaks and estimating the impact of pressure control in water supply networks. *Reliability Engineering and System Safety*, *210*(February), 107525. https://doi.org/10.1016/j.ress.2021.107525

Ji, J., Hong Lai, J., Fu, G., Zhang, C., & Kodikara, J. (2020). Probabilistic failure investigation of small diameter cast iron pipelines for water distribution. *Engineering Failure Analysis*, *108*(June), 104239. https://doi.org/10.1016/j.engfailanal.2019.104239

Ji, J., Robert, D. J., Zhang, C., Zhang, D., & Kodikara, J. (2017). Probabilistic physical modelling of corroded cast iron pipes for lifetime prediction. *Structural Safety*, *64*, 62–75. https://doi.org/10.1016/j.strusafe.2016.09.004

Jiang, R., Rathnayaka, S., Shannon, B., Zhao, X. L., Ji, J., & Kodikara, J. (2019). Analysis of failure initiation in corroded cast iron pipes under cyclic loading due to formation of through-wall cracks. *Engineering Failure Analysis*, *103*(September 2018), 238–248. https://doi.org/10.1016/j.engfailanal.2019.04.031

Jun, H. J., Park, J. K., & Bae, C. H. (2017). Deep Leaning Neural Networks for Determining Replacement Timing of Steel Water Transmission Pipes. *Proceedings - 2017 International Conference on Control, Artificial Intelligence, Robotics and Optimization, ICCAIRO 2017*, *2018-Janua*, 219–225. https://doi.org/10.1109/ICCAIRO.2017.49

Jun, H. J., Park, J. K., & Bae, C. H. (2020). Factors Affecting Steel Water-Transmission Pipe Failure and Pipe-Failure Mechanisms. *Journal of Environmental Engineering*, *146*(6), 04020034. https://doi.org/10.1061/(asce)ee.1943-7870.0001692

Kabir, G., Tesfamariam, S., Francisque, A., & Sadiq, R. (2015). Evaluating risk of water mains failure using a Bayesian belief network model. *European Journal of Operational Research*, *240*(1), 220–234. https://doi.org/10.1016/j.ejor.2014.06.033

Kanakoudis, V. K. (2004). A troubleshooting manual for handling operational problems in water pipe networks. *Journal of Water Supply: Research and Technology - AQUA*, *53*(2), 109–124. https://doi.org/10.2166/aqua.2004.0010

Karamouz, M., Yousefi, M., Zahmatkesh, Z., & Nazif, S. (2012). Development of an Algorithm for Vulnerability Zoning of Water Distribution Network. *World Environmental and Water Resources Congress 2012: Crossing Boundaries*, 3011–3020.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *2017-Decem*(Nips), 3147–3155.

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 and Resilient Infrastructure*, *8*(1), 1–23. https://doi.org/10.1080/23789689.2020.1787033

Khodabandelu, A., & Park, J. W. (2021). Agent-based modeling and simulation in construction. *Automation in Construction*, *131*(July), 103882. https://doi.org/10.1016/j.autcon.2021.103882

Kim, T., Kim, K., Hyung, J., & Koo, J. (2021). *Integrated water suspension risk assessment using fault tree analysis and genetic algorithm in water supply systems*. *227*, 104–115. https://doi.org/10.5004/dwt.2021.27358

Kleiner, Y., & Rajani, B. (2012). Comparison of four models to rank failure likelihood of

individual pipes. *Journal of Hydroinformatics*, *14*(3), 659–681. https://doi.org/10.2166/hydro.2011.029

Kline, R. (2016). *Principles and Practice of Structural Equation Modeling* (Fifth). Guilford Press.

Konstantinou, C., & Stoianov, I. (2020). A comparative study of statistical and machine learning methods to infer causes of pipe breaks in water supply networks. *Urban Water Journal*, *17*(6), 534–548. https://doi.org/10.1080/1573062X.2020.1800758

Kotsiantis, S. B. (2014). Erratum: Feature selection for machine learning classification problems: A recent overview (Artificial Intelligence Review (2011)). *Artificial Intelligence Review*, *42*(1), 157. https://doi.org/10.1007/s10462-011-9230-1

Kumar, A., Ali, S., Rizvi, A., Brooks, B., Vanderveld, R. A., Wilson, K. H., Kenney, C., Edelstein, S., Finch, A., Maxwell, A., & Zuckerbraun, J. (2018). Using Machine Learning to Assess the Risk of and Prevent Water Main Breaks. *KDD 2018: 24th ACMSIGKDD International Conference on Knowledge Discovery & Data Mining*, 472–480.

Kutyłowska, M., & Hotloś, H. (2013). Failure analysis of water supply system in the Polish city of Głogów. *Engineering Failure Analysis*, *41*, 23–29. https://doi.org/10.1016/j.engfailanal.2013.07.019

Kutyłowska, M., & Orłowska-Szostak, M. (2016). Comparative analysis of water–pipe network deterioration–case study. *Water Practice and Technology*, *11*(1), 148–156. https://doi.org/10.2166/wpt.2016.018

Lee, S. H. (2011). *Prioritizing Water Pipe Replacement and Rehabilitation By* (Issue December). Texas A&M University.

Lei, J., & Sægrov, S. (1998). Statistical approach for describing failures and lifetimes of water mains. *Water Science and Technology*, *38*(6 pt 5), 209–217. https://doi.org/10.1016/S0273-1223(98)00582-4

Leishear, R. (2019). WATER MAIN BREAKS, WATER HAMMER. CORROSION, AND FATIGUE FAILURES. *ASME Journal of Pressure Vessel Technology Paper*, *11*(2), 444–446. https://doi.org/10.1111/j.1559-3584.1899.tb01231.x

Li, C., Firouzi, A., & Yang, W. (2017). Prediction of Pitting Corrosion – Induced Perforation of Ductile Iron Pipes. *Journal of Engineering Mechanics*, *143*. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001258.

Li, W., Mazumder, R. K., & Li, Y. (2021). Time-Dependent Reliability Analysis of Buried Water Distribution Network: Combined Finite-Element and Probabilistic Approach. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, *7*(4), 04021064. https://doi.org/10.1061/ajrua6.0001178

Li, Z., Zhang, B., Wang, Y., Chen, F., Taib, R., Whiffin, V., & Wang, Y. (2013). *Water pipe condition assessment : a hierarchical beta process approach for sparse incident data*. https://doi.org/10.1007/s10994-013-5386-z

Lin, P., Zhang, B., Wang, Y., Li, Z., Li, B., Wang, Y., & Chen, F. (2015). Data driven water pipe failure prediction: A Bayesian nonparametric approach. *International Conference on Information and Knowledge Management, Proceedings*, *19-23-Oct-*, 193–202. https://doi.org/10.1145/2806416.2806509

Lindhe, A., Rosen, L., Norberg, T., & Bergstedt, O. (2009). Fault tree analysis for integrated and probabilistic risk analysis of drinking water systems. *Water Research*, *43*(6), 1641–1653. https://doi.org/10.1016/j.watres.2008.12.034

Liu, W., Xie, Z., & Song, Z. (2023). *Predicting Water Pipe Failures Using Deep Learning Algorithms*. *29*(2019), 1–13. https://doi.org/10.1061/JITSE4.ISENG-2247

Liyanage, K., & Dhar, A. S. (2018). Stresses in cast iron water mains subjected to non-uniform bedding and localised concentrated forces. *International Journal of Geotechnical Engineering*, *12*(4), 368–376. https://doi.org/10.1080/19386362.2017.1282338

Liyanage, K. T. H., & Dhar, A. S. (2017). Effects of Corrosion Pits on Wall Stresses in Cast-Iron Water Mains. *Journal of Pipeline Systems Engineering and Practice*, *8*(4), 04017023. https://doi.org/10.1061/(asce)ps.1949-1204.0000286

Lundberg, S., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *31st Conference on Neural Information Processing Systems*, *4*, 552–564. https://doi.org/10.1016/j.ophtha.2018.11.016

Luo, S., Chu, V. W., Zhou, J., Chen, F., Wong, R. K., & Huang, W. (2017). A Multivariate Clustering Approach for Infrastructure Failure Predictions. *Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017*, 274–281. https://doi.org/10.1109/BigDataCongress.2017.42

Ma, X., Li, Y., Zhang, W., Li, X., Shi, Z., Yu, J., Wang, J., & Liu, J. (2022). A Real-Time Method to Detect the Leakage Location in Urban Water Distribution Networks. *Journal of Water Resources Planning and Management*, *148*(12). https://doi.org/10.1061/(asce)wr.1943-5452.0001628

Mackey, T., Cashman, A., & Cumberbatch, R. (2014). *Identification of Factors Contributing to the Deterioration and Losses in the Water Distribution System in Barbados* (Issue 68).

Mady, R. (2021). Developing a Survivor Curve for Prestressed Concrete Cylinder Pipe Rabia. *Pipelines*, 46–54.

Mahmoodian, M., & Aryai, V. (2017). Structural failure assessment of buried steel water pipes subject to corrosive environment. *Urban Water Journal*, *14*(10), 1023–1030. https://doi.org/10.1080/1573062X.2017.1325500

Mahmoodian, M., & Li, C. Q. (2016). Structural integrity of corrosion-affected cast iron water pipes using a reliability-based stochastic analysis method. *Structure and Infrastructure Engineering*, *12*(10), 1356–1363. https://doi.org/10.1080/15732479.2015.1117114

Mahmoodian, M., & Li, C. Q. (2018). Reliability-Based Service Life Prediction of Corrosion-Affected Cast Iron Pipes Considering Multifailure Modes. *Journal of Infrastructure Systems*, *24*(2), 04018004. https://doi.org/10.1061/(asce)is.1943-555x.0000417

Mailhot, A., Pelletier, G., & Noel, J.-F. (2000). *Modeling the evolution of the structural state of water pipe networks with brief recorded pipe break histories : Methodology and application*. *36*(10), 3053–3062.

Makar, J. ., Desnoyers, R., & Mcdonald, S. . (2001). Failure modes and mechanisms in gray

cast iron pipe NRCC-44218. *Underground Infrastructure Research: Municipal, Industrial and Environmental Applications*. www.nrc.ca/irc/ircpubs

Makar, J. M. (2000). A preliminary analysis of failures in grey cast iron water pipes. *Engineering Failure Analysis*, *7*(1), 43–53. https://doi.org/10.1016/S1350-6307(99)00005-9

Mazumder, R. K., Salman, A. M., & Li, Y. (2021). Failure risk analysis of pipelines using data-driven machine learning algorithms. *Structural Safety*, *89*(July 2020), 102047. https://doi.org/10.1016/j.strusafe.2020.102047

Mazumder, R. K., Salman, A. M., Li, Y., & Yu, X. (2021). Asset Management Decision Support Model for Water Distribution Systems: Impact of Water Pipe Failure on Road and Water Networks. *Journal of Water Resources Planning and Management*, *147*(5), 04021022. https://doi.org/10.1061/(asce)wr.1943-5452.0001365

Mhadbi, N. (2021). *Python Tutorial: Streamlit*. Datacamp. https://www.datacamp.com/tutorial/streamlit

Mikdam, A., Colin, X., Minard, G., Billon, N., & Maurin, R. (2017). A kinetic model for predicting the oxidative degradation of additive free polyethylene in bleach desinfected water. *Polymer Degradation and Stability*, *146*, 78–94. https://doi.org/10.1016/j.polymdegradstab.2017.09.020

Models, E. E., Wall, P., Loss, T., & Pipes, W. (2024). Explainable Ensemble Models for Predicting Wall Thickness Loss of Water Pipes. *Ain Shams Engineering Journal*, *August 2023*, 1–49. https://doi.org/10.1016/j.asej.2024.102630

Moerman, A., Wols, B. A., & Diemel, R. (2016). The effects of traffic loads on drinking water main failure frequencies in the Netherlands. *Water Practice and Technology*, *11*(3), 524–530. https://doi.org/10.2166/wpt.2016.057

Mohammed Abdelkader, E., Zayed, T., Elshaboury, N., & Taiwo, R. (2024). A hybrid Bayesian optimization-based deep learning model for modeling the condition of saltwater pipes in Hong Kong. In *International Journal of Construction Management*. https://doi.org/10.1080/15623599.2024.2304392

Mohandes, S. R., Kineber, A. F., Abdelkhalek, S., Kaddoura, K., Elsayed, M., Hosseini, M. R., & Zayed, T. (2022). Evaluation of the critical factors causing sewer overflows through modeling of structural equations and system dynamics. *Journal of Cleaner Production*, *375*(August), 134035. https://doi.org/10.1016/j.jclepro.2022.134035

Mora-Rodríguez, J., Delgado-Galván, X., Ramos, H. M., & López-Jiménez, P. A. (2014). An overview of leaks and intrusion for different pipe materials and failures. *Urban Water Journal*, *11*(1), 1–10. https://doi.org/10.1080/1573062X.2012.739630

Moser, A., & Folkman, S. (2008). *Buried pipe design*. McGraw-Hill Professional Publication.

Murari, K. (2015). *Impacts of leadership styles on employee empowerment*. Partridge Publishing.

Murray, J. N., & Moran, P. J. (1989). Influence of moisture on corrosion of pipeline steel in soils using in situ impedance spectroscopy. *Corrosion*, *45*(1), 34–43. https://doi.org/10.5006/1.3577885

Najjaran, H., Sadiq, R., & Rajani, B. (2006a). Fuzzy expert system to assess corrosion of

cast/ductile iron pipes from backfill properties. *Computer-Aided Civil and Infrastructure Engineering*, *21*(1), 67–77. https://doi.org/10.1111/j.1467-8667.2005.00417.x

Najjaran, H., Sadiq, R., & Rajani, B. (2006b). *Fuzzy Expert System to Assess Corrosion of Cast / Ductile Iron Pipes from Backfill Properties*. *21*, 67–77.

Nguyen, H., Abdel-Mottaleb, N., Uddin, S., Zhang, Q., Lu, Q., Zhang, H., & Li, M. (2022). Joint maintenance planning of deteriorating co-located road and water infrastructures with interdependencies. *Reliability Engineering and System Safety*, *226*(June), 108678. https://doi.org/10.1016/j.ress.2022.108678

Nguyen, H. V., & Byeon, H. (2023). Predicting Depression during the COVID-19 Pandemic Using Interpretable TabNet: A Case Study in South Korea. *Mathematics*, *11*(14). https://doi.org/10.3390/math11143145

Noor, E., & Al-Moubaraki, A. (2014). *Influence of Soil Moisture Content on the Corrosion Behavior of X60 Steel in Influence of Soil Moisture Content on the Corrosion Behavior of X60 Steel in Different Soils*. *May 2015*. https://doi.org/10.1007/s13369-014-1135-2

Ogutu, G. A., Okuthe, P. K., & Lall, M. (2017). A review of probabilistic modelling using bayesian network. *Journal of Engineering and Applied Sciences*, *12*(12), 3163–3173.

Olanrewaju, O. I., Kineber, A. F., Chileshe, N., & Edwards, D. J. (2022). Modelling the relationship between Building Information Modelling (BIM) implementation barriers, usage and awareness on building project lifecycle. *Building and Environment*, *207*(PB), 108556. https://doi.org/10.1016/j.buildenv.2021.108556

Olawumi, T. O., & Chan, D. W. M. (2018). A scientometric review of global research on sustainability and sustainable development. *Journal of Cleaner Production*, *183*, 231–250. https://doi.org/10.1016/j.jclepro.2018.02.162

Ombudsman. (2014). *The Ombudsman probes Water Supplies Department mechanism for handling leaks (bursts) of private water pipes*. https://ofomb.ombudsman.hk/abc/en-us/press_releases/detail/46

Ott, L., & Longnecker, M. (2016). *An introduction to statistical methods and data analysis* (Seventh). Cengage Learning.

Padmanabhan, D., Agarwal, H., Renaud, J. E., & Batill, S. M. (2006). A study using Monte Carlo Simulation for failure probability calculation in Reliability-Based Optimization. *Optimization and Engineering*, *7*(3), 297–316. https://doi.org/10.1007/s11081-006-9973-8

Paradkar, A.B. (2012). An evaluation of failure modes for cast iron and ductile iron water pipes.

Pękala, A., & Pietrucha-Urbanik, K. (2018). The influence of the soil environment on the corrosivity of failure infrastructure - Case study of the exemplary water network. *Archives of Civil Engineering*, *64*(1), 133–144. https://doi.org/10.2478/ace-2018-0009

Phan, H. C., Dhar, A. S., Hu, G., & Sadiq, R. (2019). Managing water main breaks in distribution networks—A risk-based decision making. *Reliability Engineering and System Safety*, *191*(June), 106581. https://doi.org/10.1016/j.ress.2019.106581

Phan, H. C., Dhar, A. S., & Sadiq, R. (2018). *Prioritizing Water Mains for Inspection and Maintenance Considering System Reliability and Risk*. *9*(2011), 1–12.

https://doi.org/10.1061/(ASCE)PS.1949-1204.0000324

Pietrucha-Urbanik, K., & Tchórzewska-Cieślak, B. (2017). Failure risk assessment in water network in terms of planning renewals – A case study of the exemplary water supply system. *Water Practice and Technology*, *12*(2), 274–286. https://doi.org/10.2166/wpt.2017.034

Pietrucha-Urbanik, Katarzyna. (2015). Failure analysis and assessment on the exemplary water supply network. *Engineering Failure Analysis*, *57*, 137–142. https://doi.org/10.1016/j.engfailanal.2015.07.036

Pietrucha-Urbanik, Katarzyna, & Tchórzewska-Cieślak, B. (2020). Cost Analysis of Water Pipe Failure. *Advances in Intelligent Systems and Computing*, *987*, 411–424. https://doi.org/10.1007/978-3-030-19501-4_41

Pluye, P., & Hong, Q. N. (2014). Combining the power of stories and the power of numbers: Mixed methods research and mixed studies reviews. *Annual Review of Public Health*, *35*, 29–45. https://doi.org/10.1146/annurev-publhealth-032013-182440

Poojitha, S. N., & Jothiprakash, V. (2022). Application of Fine-Tuned Krill Herd Algorithm in Design of Water Distribution Networks. *Journal of Pipeline Systems Engineering and Practice*, *13*(4), 1–18. https://doi.org/10.1061/(asce)ps.1949-1204.0000684

Pouri, Z., & Heidarimozaffar, M. (2022). Spatial Analysis and Failure Management in Water Distribution Networks Using Fuzzy Inference System. *Water Resources Management*, *36*(6), 1783–1797. https://doi.org/10.1007/s11269-022-03104-5

Pranckutė, R. (2021). Web of science (Wos) and scopus: The titans of bibliographic information in today's academic world. *Publications*, *9*(1). https://doi.org/10.3390/publications9010012

Prieto, M. A., Murado, M. A., Bartlett, J., Magette, W. L., & Curran, T. P. (2015). Mathematical model as a standard procedure to analyze small and large water distribution networks. *Journal of Cleaner Production*, *106*, 541–554. https://doi.org/10.1016/j.jclepro.2014.12.011

Pritchard, O., Hallett, S. H., Farewell, T. S., & Pritchard, O. G. (2013). *Infrastructure Transitions Research Consortium Soil corrosivity in the UK – impacts on critical infrastructure Soil Corrosivity in the UK – Impacts on Critical Infrastructure*.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). Catboost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, *2018-Decem*(Section 4), 6638–6648.

Punurai, W., & Davis, P. (2017). Prediction of asbestos cement water pipe aging and pipe prioritization using Monte Carlo simulation. *Engineering Journal*, *21*(2), 1–13. https://doi.org/10.4186/ej.2017.21.2.1

Qian, G., Niffenegger, M., & Li, S. (2011). Probabilistic analysis of pipelines with corrosion defects by using FITNET FFS procedure. *Corrosion Science*, *53*(3), 855–861. https://doi.org/10.1016/j.corsci.2010.10.014

Qian, G., Niffenegger, M., Zhou, W., & Li, S. (2013). Effect of correlated input parameters on the failure probability of pipelines with corrosion defects by using FITNET FFS procedure. *International Journal of Pressure Vessels and Piping*, *105–106*, 19–27. https://doi.org/10.1016/j.ijpvp.2013.02.004

Raharjo, R. D., Soleh, A. M., & Sartono, B. (2019). An empirical study of the performance of two stage optimal ensemble classification using genetic algorithm. *IOP Conference Series: Earth and Environmental Science PAPER*, *299*. https://doi.org/10.1088/1755-1315/299/1/012024

Rahbaralam, M., Modesto, D., Abdollahi, A., Cucchietti, F. M., & Barcelona, D. (2020). Predictive Analytics for Water Asset Management: Machine Learning and Survival Analysis. *ArXiv*, 1–19.

Rahman, A., Sarkar, A., Yadav, O. P., Achari, G., & Slobodnik, J. (2020). Potential human health risks due to environmental exposure to nano- and microplastics and knowledge gaps: A scoping review. *Science of the Total Environment*, *757*, 143872. https://doi.org/10.1016/j.scitotenv.2020.143872

Rajani, B., & Tesfamariam, S. (2005). Estimating time to failure of ageing cast iron water mains under uncertainties. *NRC Publications Archive*.

Rajeev, P., Kodikara, J., Robert, D., Zeman, P., & Rajani, B. (2014). Factors contributing to large daimeter water pipe failure. *Water Asset Management International*, *10*(3), 9–14. https://www.researchgate.net/publication/271446112_Factors_contributing_to_large_diameter_water_pipe_failure

Rak, J. R., Wartalska, K., & Kaźmierczak, B. (2021). Weather risk assessment for collective water supply and sewerage systems. *Water (Switzerland)*, *13*(14). https://doi.org/10.3390/w13141970

Randal, O., Urbanowicz, R., Andrews, P., Lavender, N., Kidd, L. C., & Moore, J. (2016). Automating Biomedical Data Science Through Tree-Based Pipeline Optimization. *European Conference on the Applications of Evolutionary Computation*, *1*, 123–137. https://doi.org/10.1007/978-3-319-31204-0

Raspati, G. S., Bruaset, S., Bosco, C., Mushom, L., Johannessen, B., & Ugarelli, R. (2022). A Risk-Based Approach in Rehabilitation of Water Distribution Networks. *International Journal of Environmental Research and Public Health*, *19*(3). https://doi.org/10.3390/ijerph19031594

Rathnayaka, S., Shannon, B., Robert, D., & Kodikara, J. (2017). Experimental evaluation of bursting capacity of corroded grey cast iron water pipeline. *Structure and Infrastructure Engineering*, *13*(12), 1553–1562. https://doi.org/10.1080/15732479.2017.1303840

Raziani, S., & Azimbagirad, M. (2022). Deep CNN hyperparameter optimization algorithms for sensor-based human activity recognition. *Neuroscience Informatics*, *2*(3), 100078. https://doi.org/10.1016/j.neuri.2022.100078

Rehan Sadiq, Y. K. and B. R. (2010). *Aggregative risk analysis for water quality failure in distribution networks NRC Publications Archive (NPArC)*. JANUARY, 241–261.

Render. (2023). *Render*. https://render.com/

Renzetti, S., & Dupont, D. (2013). *Buried Treasure: The Economics of Leak Detection and Water Loss Prevention in Ontario*.

Rezaei, H., Ryan, B., & Stoianov, I. (2015). Pipe failure analysis and impact of dynamic hydraulic conditions in water supply networks. *Procedia Engineering*, *119*(1), 253–262. https://doi.org/10.1016/j.proeng.2015.08.883

Rifaai, T. M., Abokifa, A. A., & Sela, L. (2022). Integrated approach for pipe failure prediction and condition scoring in water infrastructure systems. *Reliability Engineering and System Safety*, *220*(December 2021), 108271. https://doi.org/10.1016/j.ress.2021.108271

Robert, D. J., Rajeev, P., Kodikara, J., & Rajani, B. (2016). An equation to predict maximum pipe stress incorporating internal and external loadings on buried pipes. *Canadian Geotechnical Journal*, 5–24.

Robles-Velasco, A., Cortés, P., Muñuzuri, J., & Onieva, L. (2021). Estimation of a logistic regression model by a genetic algorithm to predict pipe failures in sewer networks. *OR Spectrum*, *43*(3), 759–776. https://doi.org/10.1007/s00291-020-00614-9

Robles-velasco, A., Cortés, P., Muñuzuri, J., Onieva, L., Organización, D. De, Empresas, G. De, Etsi, I. I., Sevilla, U. De, & Descubrimientos, C. C. D. L. (2020). Prediction of pipe failures in water supply networks using logistic regression and support vector classification. *Reliability Engineering and System Safety*, *196*(March 2019), 106754. https://doi.org/10.1016/j.ress.2019.106754

Rogers, P. D., & Grigg, N. S. (2007). Failure assessment model to prioritize pipe replacement in water utility asset management. *8th Annual Water Distribution Systems Analysis Symposium 2006*, 19. https://doi.org/10.1061/40941(247)19

Romanoff, M. (1957). *Underground Corrosion, National Bureau of Standards (NBS)*.

Roy, J. K., & Pijush, B. (2014). Water Hammer In Piped Water Distribution System : Investigation In Practical System And Protection Scheme Water Hammer In Piped Water Distribution System : Investigation In Practical System And Protection Scheme. *Proc. of Int. Conf. on Computing, Communication & Manufacturing*, *March 2016*, 189–200.

Rubinstein, R. Y., & Kroese, D. . (2008). *Simulation and the Monte Carlo method* (2nd ed.). Wiley, New York.

Sadiq, R., Rajani, B., & Kleiner, Y. (2004). Probabilistic risk analysis of corrosion associated failures in cast iron water mains. *Reliability Engineering and System Safety*, *86*(1), 1–10. https://doi.org/10.1016/j.ress.2003.12.007

Salehi, S., Jalili-Ghazizadeh, M., Tabesh, M., Valadi, S., & Salamati Nia, S. (2021). A risk component-based model to determine pipes renewal strategies in water distribution networks. *Structure and Infrastructure Engineering*. https://doi.org/10.1080/15732479.2020.1842466

Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F., & Delipetrev, B. (2020). AI Watch - Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence. In *Joint Research Centre (European Commission)*. https://doi.org/10.2760/382730

San, N. O., Nazir, H., & Dönmez, G. (2012). Microbiologically influenced corrosion failure analysis of nickel-copper alloy coatings by Aeromonas salmonicida and Delftia acidovorans bacterium isolated from pipe system. *Engineering Failure Analysis*, *25*, 63–70. https://doi.org/10.1016/j.engfailanal.2012.04.007

Sarbatly, R. H., & Krishnaiah, D. (2007). Free chlorine residual content within the drinking water distribution system. *International Journal of Physical Science*, *2*(8), 196–201.

Scheidegger, A., Scholten, L., Maurer, M., & Reichert, P. (2013). Extension of pipe failure models to consider the absence of data from replaced pipes. *Water Research*, *47*, 3696–3705.

Scholten, L., Scheidegger, A., Reichert, P., Maurer, M., & Lienert, J. (2014). Strategic rehabilitation planning of piped water networks using multi-criteria decision analysis. *Water Research*, *49*, 124–143.

Seica, M. V., Packer, J. A., Grabinsky, M. W. F., & Adams, B. J. (2002). Evaluation of the properties of Toronto iron water mains and surrounding soil. *Canadian Journal of Civil Engineering*, *29*(2), 222–237. https://doi.org/10.1139/l01-090

Shin, H., Kobayashi, J., Koo, J., & Do, M. (2016). Estimating burst probability of water pipelines with a competing hazard model. *Journal of Hydroinformatics*. doi: 10.2166/hydro.2015.016

Shull, D. H. (2021). Sources of Corrosive Bottom Water to Bellingham Bay, Washington State. *Estuaries and Coasts*, *44*(5), 1250–1261. https://doi.org/10.1007/s12237-020-00859-1

Singh, A. (2011). Bayesian analysis for causes of failure at a water utility. *Built Environment Project and Asset Management*, *1*(2), 195–210. https://doi.org/10.1108/20441241111180433

Singh, A., & Adachi, S. (2012). Expectation Analysis of the Probability of Failure for Water Supply Pipes. *Journal of Pipeline Systems Engineering and Practice*, *3*(2), 36–46. https://doi.org/10.1061/(asce)ps.1949-1204.0000094

Snider, B., & McBean, E. A. (2018). Improving time-To-failure predictions for water distribution systems using gradient boosting algorithm. *1st International WDSA / CCWI 2018 Joint Conference*, *July*.

Snider, B., & McBean, E. A. (2021). Combining Machine Learning and Survival Statistics to Predict Remaining Service Life of Watermains. *Journal of Infrastructure Systems*, *27*(3), 1–14. https://doi.org/10.1061/(asce)is.1943-555x.0000629

Surwase, G., Sagar, A., Kademani, B. S., & Bhanumurthy, K. (2011). Co-citation Analysis : An Overview. *BOSLA National Conference*.

Taiwo, R., Abdelfadeel, I., & Zayed, T. (2023). Development of sustainable water infrastructure : A proper understanding of water pipe failure. *Journal of Cleaner Production*, *398*(February 2022), 136653. https://doi.org/10.1016/j.jclepro.2023.136653

Taiwo, R., Ben Seghier, M. E. A., & Zayed, T. (2023a). Predicting Wall Thickness Loss in Water Pipes Using Machine Learning Techniques. *2nd Conference of the European Association on Quality Control of Bridges and Structures - EUROSTRUCT2023*, *6*(5), 1087–1092. https://doi.org/10.1002/cepa.2075

Taiwo, R., Ben Seghier, M. E. A., & Zayed, T. (2023b). Towards sustainable water infrastructure : The state-of-the-art for modeling the failure probability of water pipes. *Water Resources Research*.

Taiwo, R., Shaban, I. A., & Zayed, T. (2023). Development of sustainable water infrastructure: A proper understanding of water pipe failure. *Journal of Cleaner Production*. https://doi.org/https://doi.org/10.1016/j.jclepro.2023

Taiwo, R., Zayed, T., & Ben Seghier, M. E. A. (2024). Integrated intelligent models for predicting water pipe failure probability. *Alexandria Engineering Journal*, *86*, 243–257. https://doi.org/10.1016/j.aej.2023.11.047

Tang, K., Parsons, D. J., & Jude, S. (2019a). Comparison of automatic and guided learning for Bayesian networks to analyse pipe failures in the water distribution system. *Reliability Engineering and System Safety*, *186*, 24–36. https://doi.org/10.1016/j.ress.2019.02.001

Tang, K., Parsons, D. J., & Jude, S. (2019b). Comparison of automatic and guided learning for Bayesian networks to analyse pipe failures in the water distribution system. *Reliability Engineering and System Safety*, *186*(February), 24–36. https://doi.org/10.1016/j.ress.2019.02.001

Tariq, S., Bakhtawar, B., & Zayed, T. (2022). Data-driven application of MEMS-based accelerometers for leak detection in water distribution networks. *Science of the Total Environment*, *809*, 151110. https://doi.org/10.1016/j.scitotenv.2021.151110

Tariq, S., Hu, Z., & Zayed, T. (2021). Micro-electromechanical systems-based technologies for leak detection and localization in water supply networks: A bibliometric and systematic review. *Journal of Cleaner Production*, *289*, 125751. https://doi.org/10.1016/j.jclepro.2020.125751

Tavakoli, R., Sharifara, A., & Najafi, M. (2020). Artificial Neural Networks and Adaptive Neuro-Fuzzy Models to Predict Remaining Useful Life of Water Pipelines Razieh. *World Environmental and Water Resources Congress 2020: ASCE, 2001*, III–IV.

Tawfik, G. M., Dila, K. A. S., Mohamed, M. Y. F., Tam, D. N. H., Kien, N. D., Ahmed, A. M., & Huy, N. T. (2019). A step by step guide for conducting a systematic review and meta-analysis with simulation data. *Tropical Medicine and Health*, *47*(1), 1–9. https://doi.org/10.1186/s41182-019-0165-6

Tchorzewska-Cieslak, B. (2012). Model of failure risk analysis in the water pipe network. *International Journal of Performability Engineering*, *8*(4), 379–388.

Tchórzewska-Cieślak, B., Rak, J. R., & Szpak, D. (2019). Bayesian Inference in the Analysis of the Failure Risk of the Water Supply Network. *Journal of Konbin*, *49*(3), 433–450. https://doi.org/10.2478/jok-2019-0066

Tran, V. Q. C., Le, D. V., Yntema, D. R., & Havinga, P. J. M. (2021). A Review of Inspection Methods for Continuously Monitoring PVC Drinking Water Mains. *IEEE Internet of Things Journal*, *4662*(c), 1–20. https://doi.org/10.1109/JIOT.2021.3077246

Trickey, S. A., Moore, I. D., & Balkaya, M. (2016). Parametric study of frost-induced bending moments in buried cast iron water pipes. *Tunnelling and Underground Space Technology*, *51*, 291–300. https://doi.org/10.1016/j.tust.2015.10.028

Tsai, Y. L., Chang, H. C., Lin, S. N., Chiou, A. H., & Lee, T. L. (2022). Using Convolutional Neural Networks in the Development of a Water Pipe Leakage and Location Identification System. *Applied Sciences (Switzerland)*, *12*(16). https://doi.org/10.3390/app12168034

United Utilities Water Limited. (2019). The pH of drinking water. In *United Utilities Water Limited*. https://www.unitedutilities.com/globalassets/documents/pdf/phfactsheet_acc16.pdf

van Eck, Nees, & Waltman, L. (2009). How to Normalize Cooccurrence Data? An Analysis of SomeWell-Known Similarity Measures. *Journal of the American Society for Information Science and Technology*, *60*(8), 1635–1651.

Van Eck, Ness, & Waltman, L. (2010). Software survey : VOSviewer , a computer program for bibliometric mapping. *Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping*, *84*, 523–538. https://doi.org/10.1007/s11192-009-0146-3

Vanrenterghem-Raven, A. (2007). Risk Factors of Structural Degradation of an Urban. *Journal of Infrastructure Systems*, *13*(1), 55–64. https://doi.org/10.1061/(ASCE)1076-0342(2007)13

Vaulet, T., Al-Memar, M., Fourie, H., Bobdiwala, S., Saso, S., Pipi, M., Stalder, C., Bennett, P., Timmerman, D., Bourne, T., & De Moor, B. (2022). Gradient boosted trees with individual explanations: An alternative to logistic regression for viability prediction in the first trimester of pregnancy. *Computer Methods and Programs in Biomedicine*, *213*. https://doi.org/10.1016/j.cmpb.2021.106520

Vipulanandan, C., Qiao, W., & Hovsepian, H. (2012). WATER PIPELINE FAILURES IN THE ACTIVE ZONE C. Vipulanandan 1 , W. Qiao 2 and H. Hovsepian 3. *Proceedings CIGMAT-2012 Conference*, 1–11.

Vladeanu, G., & Koo, D. (2015). A Comparison Study of Water Pipe Failure Prediction Models Using Weibull Distribution and Binary Logistic Regression. *Pipelines*, 1318–1332.

Wang, D., & Cullimore, D. R. (2010). Bacteriological challenges to asbestos cement water distribution pipelines. *Journal of Environmental Sciences*, *22*(8), 1203–1208. https://doi.org/10.1016/S1001-0742(09)60239-4

Wang, G., Wu, P., Wu, X., Zhang, H., Guo, Q., & Cai, Y. (2020). Mapping global research on sustainability of megaproject management: A scientometric review. *Journal of Cleaner Production*, *259*, 120831. https://doi.org/10.1016/j.jclepro.2020.120831

Wang, W., Wang, Y., Zhang, B., Shi, W., & Li, C. Q. (2021). Failure prediction of buried pipe network with multiple failure modes and spatial randomness of corrosion. *International Journal of Pressure Vessels and Piping*, *191*(July 2019), 104367. https://doi.org/10.1016/j.ijpvp.2021.104367

Wang, X., Yang, Z., Wang, Z., Shi, Q., Xu, B., Zhou, C., & Zhang, L. (2019). The influence of copper on the stress corrosion cracking of 304 stainless steel. *Applied Surface Science*, *478*, 492–498. https://doi.org/10.1016/j.apsusc.2019.01.291

Ward, B., Selby, A., Gee, S., & Savic, D. (2017). Deterioration modelling of small-diameter water pipes under limited data availability. *Urban Water Journal*, *14*(7), 743–749. https://doi.org/10.1080/1573062X.2016.1254252

Wasim, M., Shoaib, S., Mubarak, N. M., Inamuddin, & Asiri, A. M. (2018). Factors influencing corrosion of metal pipes in soils. *Environmental Chemistry Letters*, *16*(3), 861–879. https://doi.org/10.1007/s10311-018-0731-x

Water Supplies Department. (2021). *Annual report*. https://www.wsd.gov.hk/filemanager/common/annual_report/2019_20/en/index.html

Water Supplies Department HKSAR. (2021). *WSD Annual Report*. https://www.wsd.gov.hk/filemanager/common/annual_report/2019_20/en/index.html

Weeraddana, D., Hapuarachchi, H., Kumarapperuma, L., Khoa, N. L. D., & Cai, C. (2020). Long-Term Water Pipe Condition Assessment: A Semiparametric Model Using Gaussian Process and Survival Analysis. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 12085 LNAI*. Springer International Publishing. https://doi.org/10.1007/978-3-030-47436-2_37

Weeraddana, D., MallawaArachchi, S., Warnakula, T., Li, Z., & Wang, Y. (2021). Long-Term Pipeline Failure Prediction Using Nonparametric Survival Analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12460 LNAI*, 139–156. https://doi.org/10.1007/978-3-030-67667-4_9

Wilkowski, G. (2000). Leak-before-break: what does it really mean? *Journal of Pressure Vessel Technology, Transactions of the ASME*, *122*(3), 267–272. https://doi.org/10.1115/1.556183

Wilson, D., Filion, Y., & Moore, I. (2017). State-of-the-art review of water pipe failure prediction models and applicability to large-diameter mains. *Urban Water Journal*, *14*(2), 173–184. https://doi.org/10.1080/1573062X.2015.1080848

Wilson, Daniel, Filion, Y. R., & Moore, I. D. (2015). Identifying factors that influence the factor of safety and probability of failure of large-diameter , cast iron water mains with a mechanistic , stochastic model : A case study in the City of Hamilton. *Procedia Engineering*, *119*, 130–138. https://doi.org/10.1016/j.proeng.2015.08.863

Wilson, Daniel, Moore, I., & Filion, Y. (2017). Using sensitivity analysis to identify the critical factors that lower the factor of safety of large-diameter cast iron mains. *Urban Water Journal*, *14*(7), 685–693. https://doi.org/10.1080/1573062X.2016.1236137

Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W., & Tscheikner-Gratl, F. (2018). Pipe failure modelling for water distribution networks using boosted decision trees. *Structure and Infrastructure Engineering*, *14*(10), 1402–1411. https://doi.org/10.1080/15732479.2018.1443145

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement*, *73*(6), 913–934. https://doi.org/10.1177/0013164413495237

Wu, R., Painumkal, J. T., Volk, J. M., Liu, S., Louis, S. J., Tyler, S., Dascalu, S. M., & Harris, F. C. (2017). Parameter estimation of nonlinear nitrate prediction model using genetic algorithm. *2017 IEEE Congress on Evolutionary Computation, CEC 2017 - Proceedings*, *3*, 1893–1899. https://doi.org/10.1109/CEC.2017.7969532

Yamijala, S., Guikema, S. D., & Brumbelow, K. (2009). Statistical models for the analysis of water distribution system pipe break data. *Reliability Engineering and System Safety*, *94*(2), 282–293. https://doi.org/10.1016/j.ress.2008.03.011

Yang, Li, & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, *415*, 295–316. https://doi.org/10.1016/j.neucom.2020.07.061

Yang, Liqian, Chen, G., Rytter, N. G. M., Zhao, J., & Yang, D. (2019). A genetic algorithm-based grey-box model for ship fuel consumption prediction towards sustainable

shipping. *Annals of Operations Research*. https://doi.org/10.1007/s10479-019-03183-5

Yussif, A. M., Sadeghi, H., & Zayed, T. (2023). Application of Machine Learning for Leak Localization in Water Supply Networks. *Buildings*, *13*(4), 1–21. https://doi.org/10.3390/buildings13040849

Zamenian, H., Faust, K. M., Mannering, F. L., Abraham, D. M., & Iseley, T. (2017). Empirical Assessment of Unobserved Heterogeneity and Polyvinyl Chloride Pipe Failures in Water Distribution Systems. *Journal of Performance of Constructed Facilities*, *31*(5), 04017073. https://doi.org/10.1061/(asce)cf.1943-5509.0001067

Zayed, T. M., & Halpin, D. W. (2004). Quantitative Assessment for Piles Productivity Factors. *Journal of Construction Engineering and Management*, *130*(3), 405–414. https://doi.org/10.1061/(asce)0733-9364(2004)130:3(405)

Zhang, P., Su, L., Qin, G., Kong, X., & Peng, Y. (2019). Failure probability of corroded pipeline considering the correlation of random variables. *Engineering Failure Analysis*, *99*(February), 34–45. https://doi.org/10.1016/j.engfailanal.2019.02.002

Zheng, J., Zhang, J., Xu, J., Liu, C., & Xu, L. (2020). Experiment on frost heave failure mechanism of PPR water pipe. *Engineering Failure Analysis*, *117*(July), 104831. https://doi.org/10.1016/j.engfailanal.2020.104831

Zraick, F., Cazin, C., Bedry, S., Andres, M., & Rabaud, B. (2019). Online monitoring and control of drinking water corrosion potential at a full-scale plant. *Water Practice and Technology*, *14*(4), 772–782. https://doi.org/10.2166/wpt.2019.062

Zywiec, J., Piegdoń, I., & Tchórzewska-Cieślak, B. (2019). Failure analysis of the water supply network in the aspect of climate changes on the example of the central and eastern europe region. *Sustainability (Switzerland)*, *11*(24). https://doi.org/10.3390/su11246886