ANN MODEL FOR SEWER NETWORK

A classification-based machine learning approach for predicting sewer pipe condition

THAO TIFFANY HA

School of Business, Society and Engineering
Course: Degree Project in Environmental Engineering
Course code: MTK343
Credits: 30 hp
Program: Environmental Engineering for Sustainable Development

Supervisor: Heidi Ivan
External supervisors: Emmanuel Okwori, Victor Viñas
Examinor: Monica Odlare
Customer: Behroz Haidarian, Kretslopp och Vatten
Date: 2025-05-28
Email: tha23001@student.mdu.se

ABSTRACT

This degree project explores the application of Artificial Neural Networks (ANNs) for predictive maintenance in urban sewer infrastructure, with a case study focused on Gothenburg, Sweden. The aim was to develop and evaluate classification-based ANN models capable of predicting five distinct types of sewer pipe failures: infiltration (INL), cracks (SPR), rupture (RBR), surface damage (YTS), and deformation (DEF). Using inspection data from Kretslopp och Vatten, the study involved extensive data preprocessing, feature engineering, and model training. Key challenges addressed included severe class imbalance, limited inspection coverage (14% of the network), and data quality issues.

Five separate binary classification models were trained using a multilayer perceptron architecture, with hyperparameters optimized via Bayesian optimization. Evaluation metrics included accuracy, precision, recall, F1-score, and ROC AUC. The YTS model achieved the best performance (F1-score: 0.50, recall: 0.76), while the DEF model failed to detect any positive cases due to extreme class imbalance. Feature importance analysis revealed that pipe age, material, and soil transition were consistently influential predictors. Risk group stratification further demonstrated the models' ability to support maintenance prioritization by identifying high-risk pipe segments.

The findings highlight the potential of ANN models to enhance proactive maintenance planning in sewer networks, particularly for common failure types. However, the study also underscores the need for more representative and balanced datasets, especially for rare failure events, and the importance of integrating environmental and geotechnical data. Future work should explore multi-output classification, alternative modeling techniques, and continuous model updating to improve generalizability and practical implementation.

Keywords: artificial neural networks; condition assessment; sewer network; failure prediction; machine learning; predictive maintenance; urban infrastructure; asset management.

PREFACE

This degree project was conducted in collaboration with Kretslopp och Vatten, the municipal water and wastewater utility in Gothenburg, during the spring semester of 2025. It marks the conclusion of my MSc in Environmental Engineering at Mälardalen University.

Throughout this challenging and rewarding journey, I have gained invaluable experience, from navigating the complexities of data collection to managing the demands of academic and professional life. Each stage of the project sharpened my critical thinking and deepened my understanding of sustainable infrastructure.

I would like to express my sincere gratitude to Glen Nivert and Behroz Haidarian at Kretslopp och Vatten for their generous support and for granting access to the data that made this research possible. Special thanks to Emmanuel Okwori at RISE for his guidance and encouragement - your input meant a great deal to me and to the development of my ANN models.

I am especially thankful to my colleague and supervisor Víctor Viñas Cos at AFRY. Your unwavering support, late nights at the office, and consistent motivation kept me going. I also wish to thank my colleagues at AFRY for their flexibility, patience, and encouragement throughout this period.

My deepest thanks go to my academic supervisor, Heidi Ivan, whose calm presence and clear advice helped me slow down and focus on what truly matters.

Balancing full-time work with full-time studies over the past two years has been immensely demanding - filled with sleepless nights, looming deadlines, and constant time management challenges. It has not been easy, but I'm proud to say: I made it. I would like to thank my friends for their encouragement throughout this journey. A special thank you to Uchit Sangroula for providing valuable input and for taking the time to review my report.

Finally, from the bottom of my heart, I want to thank my mother and my two siblings, whose emotional support, even from afar, has been a pillar of strength. And to Rasmus (min namnam) - thank you for being my anchor, for your endless care, and for bringing calm and balance when I needed it most.

Gothenburg, May 2025

Thao Tiffany Ha

CONTENT

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Abbreviation	Description
ANN	Artificial Neural Network
CCTV	Closed-Circuit Television
I & I	Infiltration & Inflow
SWWA	Swedish Water & Wastewater Association
WWTP	Wastewater Treatment Plant

DEFINITIONS

Description of the following definition are retrieved from Belcic and Stryker (2024).

Definition	Description
Hyperparameters	Configuration variables to manage the training process of a machine learning model.
Hyperparameter tuning	The practice of identifying and selecting the optimal hyperparameters for use in training a machine learning model.
Learning rate	Sets the speed at which a model adjusts its parameters in each iteration. A high learning rate means that a model will adjust more quickly, but at the risk of unstable performance and data drift. A low learning rate is time-consuming and makes it more likely a model's minimum loss.
Batch size	Sets the amount of samples the model will compute before updating gits parameters. It has a significant effect on both compute efficiency and accuracy of the training process.
Number of hidden layers	Determines a neural network's depth, which affects its complexity and learning ability.
Number of nodes/ neurons per layer	Sets the width of the model. The more neurons per layer, the greater the breadth of the model and the better able it is to depict complex relationships between data points.
Momentum	The degree to which models update parameters in the same direction as previous iterations, rather than reversing course.
Epochs	A hyperparameter that sets the amount of times that a model is exposed to its entire training dataset during the training process. Greater exposure can lead to improved performance but runs the risk of overfitting.
Activation function	Introduces nonlinearity into a model, allowing it to handle more complex datasets. Nonlinear models can generalize and adapt to a greater variety of data.

1 INTRODUCTION

This degree project aims to contribute to the knowledge of Artificial Intelligence (AI) models for condition assessment developed for sewer pipelines in Sweden. This chapter gives a framework of the project by presenting the background, purpose, research questions and delimitation.

1.1 Background

Sewer networks are critical infrastructure systems that facilitate wastewater transportation from residential, commercial and industrial areas to treatment facilities (Grigg, 2010, p. Chapter 2). Their proper functioning is essential to safeguard public health, protect natural water bodies, and maintain urban resilience. In recent decades, growing urbanization, aging infrastructure, and climate-related stresses have intensified the need for proactive sewer maintenance and sustainable asset management strategies (Caradot et al., 2021, p. Introduction).

In Sweden, municipalities manage extensive sewer networks, with approximately 78,500 kilometres of sewage and 39,700 kilometres of stormwater across the country for more than 9 million inhabitants (SCB, 2022; SWWA, 2023). Maintaining this infrastructure is both financially and operationally demanding. The Swedish Water and Wastewater Association reports that annual investment for pipeline networks, both water and wastewater, covering reinvestments, demographic expansion, increased demands and climate adaption, reach approximately SEK 10-20 billion (SWWA, 2023, p. 32). This investment underscores the critical importance of maintaining and upgrading the water and wastewater infrastructure.

Failures to maintain and upgrade these systems can result in severe consequences: untreated overflows, groundwater contamination, property damage, and elevated greenhouse gas emissions from unmanaged organic loads (Tscheikner-Gratl et al., 2019; Wear et al., 2021). Addressing these challenges aligns directly with the United Nations Sustainable Development Goals, particularly:

- SDG 6: Clean Water and Sanitation – by ensuring the integrity of wastewater treatment and transport systems (United Nations, 2025a).
- SDG 9: Industry, Innovation and Infrastructure – by modernizing urban infrastructure through digital and intelligent solutions (United Nations, 2025b).
- SDG 11: Sustainable Cities and Communities – via more resilient and reliable urban infrastructure (United Nations, 2025c).
- SDG 13: Climate Action – by strengthening infrastructure adaptation and reducing environmentally damaging overflows (United Nations, 2025d).

Since the 1980s, the condition of sewer pipes has been assessed by using Closed-Circuit Television (CCTV) inspections as the industry standard (Caradot et al., 2021, p. Introduction). This method involves deploying cameras into sewer pipelines to visually inspect and document defects, such as cracks, root intrusions, and blockages. While CCTV inspections provide valuable insights, they are time-consuming, labour-intensive, and subject to human interpretations errors (Cheng & Wang, 2018, p. Abstract). Consequently, municipalities face difficulties in scaling assessments to match the size and complexity of their networks.

To meet these challenges, attention has shifted toward AI-based asset management solutions. In particular, Artificial Neural Networks (ANNs) offer robust predictive capabilities for estimating future sewer conditions by integrating diverse datasets such as pipe age, material, inspection history, surrounding soil type, and hydraulic performance (Mohammadagha et al., 2025; Mohammadi et al., 2019, p. Introduction). These models support the transition from reactive to predictive maintenance, enabling data-driven decision-making and optimized budget allocation for rehabilitation.

Moreover, leveraging AI in this domain contributes directly to global sustainability agendas. According to Ziemba et al. (2024, p. 523), the application of AI in predictive maintenance enhances infrastructure longevity, minimizes unscheduled failures and supports SDG 6 targets through improved reliability of water and sanitation systems. Mehmood et al. (2020, p. 4) emphasize the value of AI in identifying sewer anomalies, optimizing intervention schedules, and aligning urban planning with the resilience goals of Agenda 2030.

In Sweden, the ongoing digitization of the sector, combined with the increasing need for proactive asset management, indicates a foreseeable increase in the need for intelligent decision support tools in the coming years (Sörensen et al., 2024, p. Abstract). As digital platforms and smart infrastructure evolve, there is a growing demand for intelligent condition assessment models that can complement existing monitoring practices and enable more strategic, sustainable management of wastewater systems (Nguyen & Seidu, 2022, p. Abstract).

1.2 Purpose/Aim

This degree project aims to develop an Artificial Neural Networks (ANNs)-based sewer condition assessment model to predict the future condition of pipes, based on data provided by the water and wastewater utility Kretslopp och Vatten (Gothenburg), thereby enhancing decision-making processes in sewer network management. By integrating AI into sewer asset management, municipalities can reduce operational costs, minimize environmental risks and improve the long-term sustainability of wastewater infrastructure.

1.3 Research questions

The following research questions (RQs) will be investigated in this work:

RQ1: What are the main challenges and limitations when implementing ANN model for predictive maintenance in sewer network, and how can these be addressed through improved data collection or modeling techniques?

RQ2: How does ANN model performance vary across different sewer pipe failure types, and what role do input features and data imbalance play in this variation?

By addressing these questions, this research seeks to improve failure detection accuracy, optimize maintenance schedules and enhance the resilience of urban sewer systems. Attention will be given to Sweden's sewer maintenance strategies and the role of smart monitoring technologies in ensuring efficient wastewater management.

1.4 Delimitation

This degree project is delimited to the application of ANN for the predictive condition assessment of sewer pipeline infrastructure systems, including sanitary, stormwater and combined sewer systems. The research focuses on modelling sewer deterioration based on static attributes such as pipe age, material, diameter, soil data, and historical CCTV inspection data. The analysis excludes real-time sensor data, flow dynamics, and hydraulic modelling parameters, while being constrained to the available data from Gothenburg municipal water and wastewater utility Kretslopp och Vatten, including 80 007 pipes with 14% having inspection records. Furthermore, the scope was limited to binary classification rather than multi-output classification, as this decision was made in consideration of project time constraints and to mitigate risks associated with data quality issues and the added complexity of multi-output modeling.

While other machine learning approaches are acknowledged in previous studies, the research deliberately constrains its methodological focus to ANNs to enable exploration of architectural complexity variations. Model complexity is explored primarily through architectural modifications, regularization techniques and training configurations, rather than through advanced neural network architectures like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and Generative Adversarial Network (GANs).

The study utilizes inspection data rather than longitudinal deterioration tracking, with limits the ability to model deterioration rates directly. Additionally, the findings may not generalize beyond similar urban infrastructures with comparable pipe materials, age and other geographical distributions, noting that the dataset is dominated by concrete pipes, potentially biasing model performance across different material types.

2 LITERATURE STUDY

Modern cities are dynamic systems in which human use, infrastructure, and the natural environment are in constant interaction - there is nowhere this is more evident than in the urban water cycle. As Wei et al. (2018, p. 350) outline, urban water systems operate through a dual process of natural and engineered flows, including precipitation, infiltration, runoff, and evaporation, and human-activated processes such as water supply, consumption, drainage, and wastewater treatment, see Figure 2-1. Sewer systems play a pivotal role in this cycle, serving as the conduits that link residential dwellings, industrial facilities, and public areas to treatment plants and ultimately to receiving water bodies. In Sweden, where municipal control and environmental administration are highly intertwined, sewer systems have developed over a century. However, a significant portion of the network is currently facing challenges related to age, capacity, and climate change. Conventional inspection-based maintenance is both time-consuming and typically reactive, whereas predictive capabilities are gaining popularity. The utilization of Artificial Intelligence (AI), particularly Artificial Neural Networks (ANN), has the potential to enhance the evaluation of conditions and the decision-making process by facilitating the detection of failure risks in advance. This literature review outlines the development, state of the art, and future advances in sewer condition evaluation, including the role of AI-based modeling in water sustainability in urban environments.

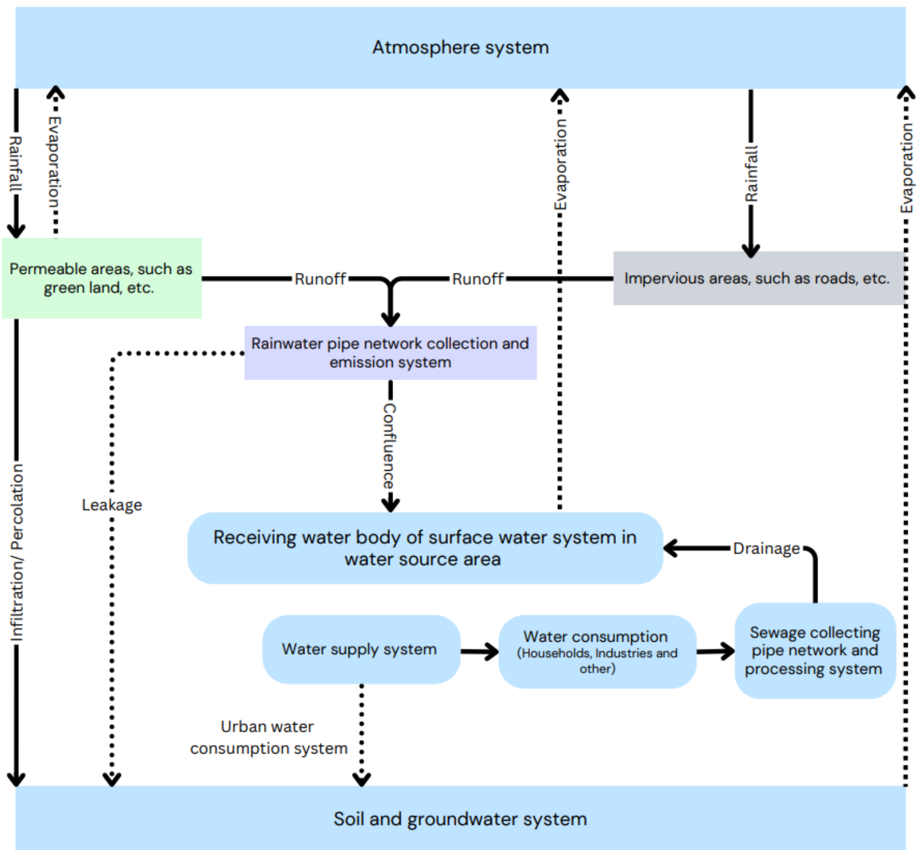


Figure 2-1 Urban water cycle structure. Retrieved from Wei et al. (2018, p. 350).

2.1 Sewer pipe failure

As Sweden's sewer infrastructure continues to age, municipalities are facing increasing challenges from pipe failures.

Neglecting asset management and condition management can increase the risk of pipe failure, as sewer systems depend on regular inspection and maintenance to ensure their long-term performance and reliability (Mohammadi et al., 2019, p. Chapter 2). Sewer pipe failure refers to an event in which the sewer network does not function as intended. Sewer pipe failures can be generalized into structural, operational and hydraulic capacity failures, according to previous studies (Table 2-1) (Mohammadi et al., 2019; Najafi, 2005). Section 2.2.1 and 2.2.2 will discuss these three main types of failures and their consequences.

Table 2-1 Summary of main types of failures and consequences. Adapted from Mohammadi et al. (2019, p. Chapter 2).

Failure type	Origin	Impact
Structural failure	Cracks, internal, external corrosion, pipe deflection, misaligned joints, and breaks	Social, economic and environmental impacts
Operational failure	Debris, infiltration, root intrusion, sediment accumulation, obstruction and grease build-up	
Hydraulic capacity failure	Occurs when flow is higher than pipe capacity, Infiltration/inflow (I/I)	

2.1.1 *Main types of failures*

2.1.1.1. Structural failure

Structural failure in wastewater pipes occurs when the pipe's physical structure degrades to a point where it can no longer perform its intended function (Mohammadi et al., 2019, p. 16). Failure can arise from various factors, including material degradation, soil interaction, and environmental conditions. The most common type of defects for structural failure are cracks, corrosion (both internal and external), deflection, joint displacements, and breaks. Davies et al. (2001, p. 89) has summarized the factors believed to influence the structural deterioration of sewer pipes, see Table 2-2:

Table 2-2 Summary of the most common factors identified that influence the structural failure of pipes. Adapted from Davies et al. (2001, p. 89); Rajani and Kleiner (2001).

Construction factors	Local external factors	Other factors
Bedding material and type	Ground movement	Sewage characteristics
Connections	Groundwater level	Use of inappropriate maintenance methods
Installation method	Infiltration/ Exfiltration	
Joint type and material	Traffic characteristics	Asset age
Sewer size/ depth/ material	Frost	Sediment level

Standard of workmanship Pipe length	Root interference Surface use, loading, type Soil/ backfill type	Surcharge Chemical processes (corrosion and sulfuric acid)

Table 2-3 presents the failure modes for various types of pipe material, as according to EPA (2009, p. Structural failure), different pipe material has different degree of failure.

Table 2-3 Failure models for various types of pipe material. Adapted from EPA (2009, p. Structural failure); Malek Mohammadi (2019, p. 37)

Pipe material	Failure models
Ferrous pipe (Ductile iron, Cast iron, Steel)	Internal or external corrosion are the primary failure mode for metal pipes
Concrete pipe (RCP, PCCP)	Corrosion is often a main factor in the structural failure of concrete pipes when the concrete break up at the result of corroded reinforcing steel inside the pipe
Ceramic-based pipe (Brick, Vitrified Clay pipe)	Collapse caused by weakened mortar is one of the main reasons for brick pipes failure Loss of surrounding soil into the pipe is the other important mode of failure for ceramic based pipes
Plastic pipe (Polyvinyl Chloride (PVC), High-density Polyethylene (HDPE))	Environmental stress cracking is the primary mode for plastic pipe failure

Understanding structural failure in wastewater pipes requires recognizing its progressive nature. The deterioration occurs in stages, influenced by various factors including corrosion, inadequate joint sealing, soil settlement, and external loading pressures. These cumulative factors gradually weaken the pipe structure until actual failure or breakage occurs. Study by Rajani and Kleiner (2001) identifies three distinct categories of pipe breakage: circumferential breakage (occurring around the pipe's circumference), longitudinal breaks (running along the pipe's length), and bell split or joint failure (failure at connection points), see Figure 2-2. Each type represents different stress patterns and failure mechanisms within the wastewater pipe system.

Pipe failure typically develops incrementally rather than suddenly. Small cracks expand into larger fractures, with deterioration accelerating when combined with infiltration, disturbed bedding, and ground movements (Mahamud, 2023, p. 21; Rajani & Kleiner, 2001). Structural failures often trigger operational issues including blockages, infiltration/exfiltration, and overflows. These consequences can lead to service interruptions, environmental contamination when sewage enters waterways, and health hazards. Sewer backups into buildings frequently result in costly repairs and significant health concerns for affected properties.

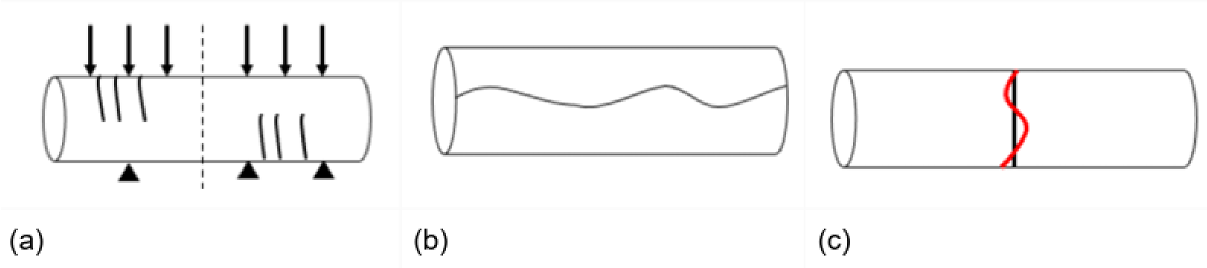


Figure 2-2 a) Circumferential breakage, b) Longitudinal crack and c) Bell split/ joint failure (Mahamud, 2023, p. 21; Rajani & Kleiner, 2001).

2.1.1.2. Operational failure

Operational failures represent the predominant type of dysfunction in wastewater collection networks (Malek Mohammadi, 2019, p. 37). These failures typically arise from physical cause that can be addressed through standard maintenance protocols without compromising the structural integrity of the pipe infrastructure. Common operational failures include debris, infiltration, root intrusion, sediment accumulation, obstruction and grease build-up.

Root intrusion, sediment deposition, fats, oils, grease or non-disposables entering the sewer system remain as significant contributors to blockages (Alda-Vidal et al., 2020; Malm, Horstmark, Jansson, et al., 2011). Blockages can significantly disrupt flow, leading to sewer backups and basement flooding, particularly during periods of heavy rainfall. Combined sewer systems are especially susceptible to these failures due to their dual handling of sewage and stormwater. Blockages may lead to ultimate failure and consequently environmental pollution, property damage and health risks (Okwori, 2021, p. 1).

Another operational issue is sewer overflow, specifically sanitary sewer overflows (SSOs) and combined sewer overflows (CSOs). A SSO is defined as a discharge of untreated wastewater from a sanitary sewer system, the cause of which can vary from inadequate sewer design to insufficient operation and maintenance (Enfinger & Stevens, 2007, pp. 2, 8). On the other hand, a CSO refers to the discharge of untreated wastewater from a combined sewer system, which frequently occurs during periods of heavy rainfall when the combined flow rate of wastewater and stormwater exceeds the designed capacity of the system. While overflows can temporarily relieve backpressure, they also pose serious environmental and public health risks by discharging untreated wastewater into natural water bodies. The severity of impacts depends on factors like wastewater volume, recipient body characteristics, and weather conditions, with dry-weather overflows causing particularly concentrated pollution (Malm, Horstmark, Jansson, et al., 2011, p. 80).

Infiltration and inflow (I/I) further contribute to operational failures (Malm, Horstmark, Jansson, et al., 2011, p. 38). Infiltration refers to groundwater entering sewer pipes through cracks or defective joints, while inflow results from inappropriate connections like storm drains and roof leaders. Elevated I/I levels can overwhelm treatment plants, trigger SSOs/CSOs, and increase operational costs.

2.1.1.3. Hydraulic capacity failure

Hydraulic capacity in sewer networks is defined as a continuous process that results in a reduction of flow capacity due to a decrease in the cross-sectional area of the pipes and an increase in hydraulic resistance (Rodríguez et al., 2012, p. Section 2.1). It has been documented that the primary mechanism for hydraulic deterioration is sediment deposition and accumulation (Okwori, 2021, p. 8).

Hydraulic capacity failure may also be the result of infiltration/ inflow (I/I), where the groundwater and stormwater enter the sewer system through connections, manholes, cracks, and defects (Malek Mohammadi, 2019, p. 38). The risk of hydrophilic capacity failure is increased by other factors, including pipe deformation and inadequate slope along the pipe. The absence of adequate slope along the pipe may be attributed to various factors, including the loss of pipe bedding or inadequate construction and design. Hydraulic capacity failures often signal underlying structural problems such as cracks, leaks, or pipe breakage.

2.1.2 Consequences of sewer pipe failure

The failure of sewer pipes poses significant challenges to urban infrastructure systems, with consequences that span environmental degradation, public health risks, economic burdens and social disruptions. As many sewer networks across the globe, and in Sweden, age under increasing pressure from urbanization and climate-related stresses, the consequences of pipe failure have become a central concern in asset management planning.

2.1.2.1. Environmental and public health impacts

Ones of the most direct impacts of sewer pipe failure is the occurrence of sanitary sewer overflows (SSOs), which lead to the release of untreated wastewater into the environment. This contamination can result in the spread of waterborne diseases and pollution of soil and water bodies, ultimately endangering public health and aquatic ecosystems (F. Alqahtani, 2023, p. Introduction). SSOs are particularly concerned in densely populated areas, where proximity to water bodies and limited drainage capacity increase the risks.

In Swedish contexts, inflow and infiltration from damaged sewer lines into treatment plants have also been linked to capacity overloads and the discharge of untreated water, particularly during heavy rainfall events (Malm, Horstmark, Jansson, et al., 2011, p. 39). These overload events can discharge untreated sewage into local waterways, leading to ecological damage and public complaints.

2.1.2.2. Economic and infrastructure costs

From a financial perspective, the costs associated with sewer pipe failure extend beyond direct repair expenses (Elmasry et al., 2017). A large proportion of wastewater in Sweden, up to 49% is attributed to infiltration and inflow water, which increases treatment costs, energy consumption and the need for expanded infrastructure (Clementson, 2020). These inefficiencies place a financial burden on municipalities, especially when untreated water causes flooding and property damage.

The annual cost of addressing sewer system deterioration in Sweden is substantial. The total replacement value of the national wastewater pipeline network is estimated at approximately SEK 680 billion, with an annual reinvestment need of about SEK 6.8 billion to maintain current service levels (Najar & Persson, 2023, p. 672). Moreover, maintenance costs are not evenly distributed. For example, medium-sized municipalities experience approximately 30% more sewer blockages per kilometer per year than others, increasing the demand for maintenance resources (Okwori et al., 2020, p. 46). Specific cases, such as Malmö's Vanåsgatan, have recorded extremely high per-meter maintenance costs due to recurring root intrusions (Rolf & Stål, 1994).

2.1.2.3. Social impacts

Operational failures in sewer systems, such as overflows, blockages, and structural collapses, can lead to severe social and functional disruptions in urban environments. These failures are often increased by aging infrastructure, increasing urban density, and climate change-related stressors, including intense rainfall and rising groundwater levels. Key operational issues include infiltration and inflow (I/I), structural defects, and hydraulic overloading, all of which can reduce system capacity and increase the risk of failure.

Blockages and stoppages within sewer pipes frequently result in backups and overflows, which can cause localized flooding. These events disrupt daily urban life by damaging private property, limiting mobility, and interrupting access to critical services (Kargar & Joksimovic, 2024; Okwori et al., 2020). In particular, sewer overflows in populated areas pose risks to public health due to potential exposure to untreated sewage, while also contributing to environmental contamination through runoff into nearby soil and water bodies (Saravanan & Vipulanandan, 2014, p. Abstract).

Furthermore, continuous leaks from deteriorating sewer pipes can lead to soil erosion around pipe joints, which increases the risk of ground subsidence and the formation of sinkholes. These effects often extend beyond the underground infrastructure, causing damage to surface structures such as roads, pavements, and building foundations (Saravanan & Vipulanandan, 2014; Vipulanandan & Liu, 2005). The cumulative impact of such events is significant, often requiring substantial municipal response and repair efforts (Zamanian et al., 2020, p. 51).

2.2 Factors influencing sewer deterioration

The aging of sewer infrastructure is motivated by numerous interrelated factors that may encompass physical design, environmental exposure, and operational conditions. The interaction between these factors may be important in forecasting pipe life, setting inspection and rehabilitation priorities, and informing data-driven asset management practices.

Previous studies, especially the early work of Al-Barqawi and Zayed (2006) categorize the reasons for deterioration into three broad groups: physical, environmental, and operational factors. The factors are outlined in Table 2-4, defining the most important variables associated with structural degradation, hydraulic failure, and service disruption.

Table 2-4 Factors affecting sewer pipe deterioration. Adapted from Al-Barqawi and Zayed (2006).

Physical Factors	Environmental Factors	Operational Factors
Connections	Backfill type	Blockages
End invert elevation	Bedding material	Burst history
Installation method	Ground movement	Debris
Joint type	Groundwater level	Flow velocity
Pipe length	pH	Hydraulic condition
Pipe shape	Road type	Infiltration/exfiltration
Pipe slope	Root interference	Previous maintenance
Sewer age	Soil corrosivity	Sediment level
Sewer depth	Soil fracture potential	Sewer function
Sewer pipe material	Soil moisture	Surcharge
Sewer size	Sulfate soil	
Start invert elevation	Surface type	

2.2.1 Physical, environmental and operational factors

2.2.1.1. Physical factors

The deterioration of sewer pipes is significantly influenced by physical attributes such as age, material, diameter, length, and installation depth. Among these, age is one of the most consistent predictors of failure, with older pipes showing increased deterioration due to prolonged exposure to operational and environmental stressors. This is often illustrated by the "bathtub curve" (Figure 2-3)Figure 2-3 Theoretical bathtub curve of buried pipe (Singh & Adachi, 2013). representing high failure rates during early and late life stages (Davies et al., 2001; Singh & Adachi, 2013). Material type also plays a critical role. Concrete pipes, for instance, although widely used for their structural capacity, are highly susceptible to biogenic sulfuric acid corrosion in acidic or anaerobic conditions, particularly in combined or pressurized systems (Anwar et al., 2022, p. 545; Taheri et al., 2020, p. 116245). In contrast, plastic pipes such as PVC (Polyvinyl Chloride) and HDPE (High-Density Polyethylene)) are corrosion-resistant but can deform under high external loads and suffer from joint displacement or leakage when improperly installed or backfilled (Singh & Adachi, 2013).

Diameter and length influence structural resilience and hydraulic performance. Larger diameter pipes tend to have lower failure rates due to greater hydraulic capacity and wall thickness, though they may be subject to higher loads during installation and operation (Laakso et al., 2018, p. 901; Tran et al., 2009). Longer pipes, while reducing the number of maintenance access points, may be more prone to joint defects and sediment accumulation (Ana et al., 2009, p. 303). Pipe depth is also relevant: shallow pipes are more susceptible to root intrusion and surface loads, whereas deeper pipes face increased hydrostatic pressure and infiltration risks (Zafar Khan et al., 2010). In practice, these physical variables often interact with environmental and operational factors, underscoring the need for integrated condition assessments in sewer asset management.

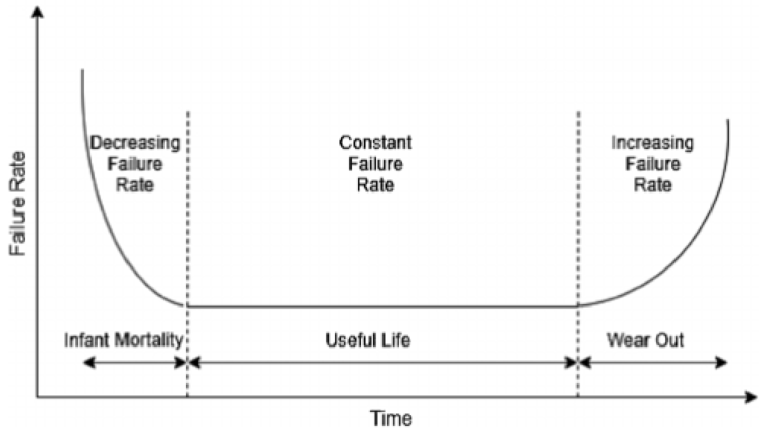


Figure 2-3 Theoretical bathtub curve of buried pipe (Singh & Adachi, 2013).

2.2.1.2. Environmental factors

Environmental conditions play a critical role in the structural integrity and long-term performance of sewer pipes. Key influences include soil type, moisture variability, groundwater levels, and bedding conditions[1], each of which interacts with the pipe's physical and operational context. Soil type is one of the most decisive environmental variables, particularly in relation to corrosion potential. Corrosive soils, such as alluvial soils, accelerate pipe material degradation due to their chemical composition, moisture retention, and microbial activity. In comparison, podzolic soils, which are formed through rock weathering, tend to have lower corrosivity (Denison & Ewing, 1935; Smith, 1968, p. 221; Wang et al., 2016, p. 357)

In Sweden, especially around Gothenburg, the clay-rich soils create specific challenges because they tend to hold moisture. This characteristic makes them particularly risky for pipe stability, notably during periods of freezing and thawing or during wet seasons. When these soils become saturated, they exert sideways pressure on underground pipes, increasing the likelihood of pipes becoming deformed or even collapsing due to changing loads. Research has shown that repeated loading and significant soil movement can greatly weaken clay soils, which impacts the safety and stability of buildings and slopes (Åhnberg et al., 2013, p. Abstract). Additionally, the swelling and shrinking of soils with seasonal moisture variations can cause cracks, joint separation, and eventual failure, particularly in soils prone to expansion. Studies have also highlighted that this seasonal swelling and shrinking is a critical factor behind pipe failures, with the highest rates occurring during dry summer periods (Weerasinghe et al., 2015). If pipes aren't sufficiently supported by suitable bedding

[1] Bedding condition refers to the quality and stability of the material that supports a sewer pipe from below (and sometime around). Improper bedding can lead to pipe misalignment, cracking, sagging or collapse overtime.

materials, these soil behaviors become even more problematic, increasing the risk of pipes sagging or collapsing (Davies et al., 2001).

Groundwater levels also pose a significant challenge. In clay-rich soils, changing groundwater conditions can weaken the soil's strength, reducing the lateral support needed to keep buried pipes stable (Malek Mohammadi et al., 2019, p. 49). This makes the pipes more vulnerable to deformation, water infiltration, or even floating upward. These issues become especially problematic for deeper installations or in locations where drainage is insufficient. Regularly monitoring groundwater levels and installing proper drainage systems are essential steps to prevent such problems and protect sewer infrastructure.

2.2.1.3. Operational factors

Operational factors, including sewer type, internal flow conditions, hydraulic performance, and maintenance practices, play a crucial role in the deterioration of sewer pipelines. Combined systems, which carry both sanitary wastewater and stormwater in a single pipe, have been observed to deteriorate faster due to shallower installation depths and greater flow variability during rainfall events (O'Reilly et al., 1989, p. Results). However, some studies suggest that improved construction practices in combined sewers can mitigate these effects, resulting in slower deterioration compared to separate systems (Davies et al., 2001; Malek Mohammadi, 2019). The type of wastewater conveyed also influences deterioration rates, with sanitary sewers causing more significant degradation than storm or combined systems due to higher pollutant loads.

Internal flow velocity is another operational consideration. Adequate velocity is necessary to prevent sediment deposition and blockage formation within pipes. Although a self-cleansing velocity is important for daily operation, research has indicated that flow velocity alone is not a dominant factor in predicting long-term structural deterioration (Koo & Ariaratnam, 2006, p. Chapter 7). Conversely, the hydraulic condition of a sewer, whether rated as good, fair, or poor, has been shown to have a strong relationship with physical pipe deterioration. Studies by Tran et al. (2006, p. 176) found that pipes in poor hydraulic condition[2] are significantly more prone to structural failures, although other findings suggest this relationship may not always be consistent (Micevski et al., 2002, pp. 19,21).

Maintenance activities also have a profound impact on the operational lifespan of sewer systems (Malek Mohammadi, 2019, p. 53). Proper maintenance is essential to prevent blockages and sustain hydraulic performance, but inappropriate methods can inadvertently accelerate deterioration. Techniques such as high-pressure water jetting or aggressive flushing have been linked to the development of additional defects in pipe walls (Malek Mohammadi, 2019, p. 54). Therefore, selecting appropriate cleaning practices, based on the pipe material and condition, is critical for preserving the structural and operational integrity of wastewater networks.

---

[2] Hydraulic condition refers to how effectively a sewer pipe conveys wastewater, considering factors like flow capacity, velocity, slop and blockage risk.

2.2.2 Condition Assessment Practices

The reinvestment cost of the Swedish wastewater pipe network is estimated at SEK 800 million, according to the VASS operation and maintenance report (SWWA, 2024, p. 14) . Given the limited lifespan of sewer infrastructure, a continuous and strategic investment is required to maintain functionality and service quality. To ensure that these substantial investments are made efficiently, municipalities must be able to assess pipe condition accurately and determine which assets should be prioritized for rehabilitation or replacement (Mahamud, 2023, p. 18). Without sufficient renewal, the cost of maintaining and repairing the network will increase significantly in the future. This highlights the critical importance of robust and objective condition assessment practices as the foundation for sustainable asset management and long-term infrastructure planning

2.2.2.1. Current practice of sewer inspection

Traditional sewer condition assessment methods primarily rely on Closed-Circuit Television (CCTV) inspections, where a camera is deployed through the pipeline to record internal conditions. This footage is later reviewed by trained inspectors to identify and grade structural defects such as cracks, joint displacements, blockages, and corrosion (Denha, 2023). In Sweden, CCTV has been the dominant technique for many years and is considered a fundamental tool in utility management. The method is valued for its direct visual documentation and its compatibility with standardized condition rating systems, such as those used in the P122 methodology developed by Swedish Water and Wastewater Association.

However, CCTV inspections are labor-intensive and depend heavily on the operator's expertise, leading to potential inconsistencies in defect identification and grading (Mohammadagha et al., 2025, p. 2). Additionally, while they provide internal visuals, they don't offer insights into external factors affecting the pipes, such as soil conditions or external corrosion.

While traditional condition assessment methods such as CCTV inspections and visual grading remain foundational in sewer infrastructure management, they are increasingly supplemented or replaced by model-based approaches (Mohammadagha et al., 2025, pp. 2,3). The limitations of traditional techniques, particularly their cost, subjectivity, and reactive nature, have driven interest in predictive and data-driven solutions. These modern methods leverage statistical modeling, machine learning, and sensor data to provide more proactive, consistent, and cost-effective assessments of pipe condition. Model-based approaches aim to shift the paradigm from reactive maintenance to predictive planning, improving long-term investment outcomes and operational reliability.

2.2.2.2. Model-based approaches in condition assessment

Model-based condition assessment involves the use of data-driven models to estimate the current or future condition of sewer assets without requiring full physical inspection (Mohammadi et al., 2019, p. 4). These models use historical data, CCTV inspection data, environmental variables, and asset attributes (e.g., age, material, diameter, and depth) to predict deterioration and support maintenance decision-making. A general classification of existing sewer deterioration models reveals two predominant categories: statistical models and artificial intelligence models, see Figure 2-4.

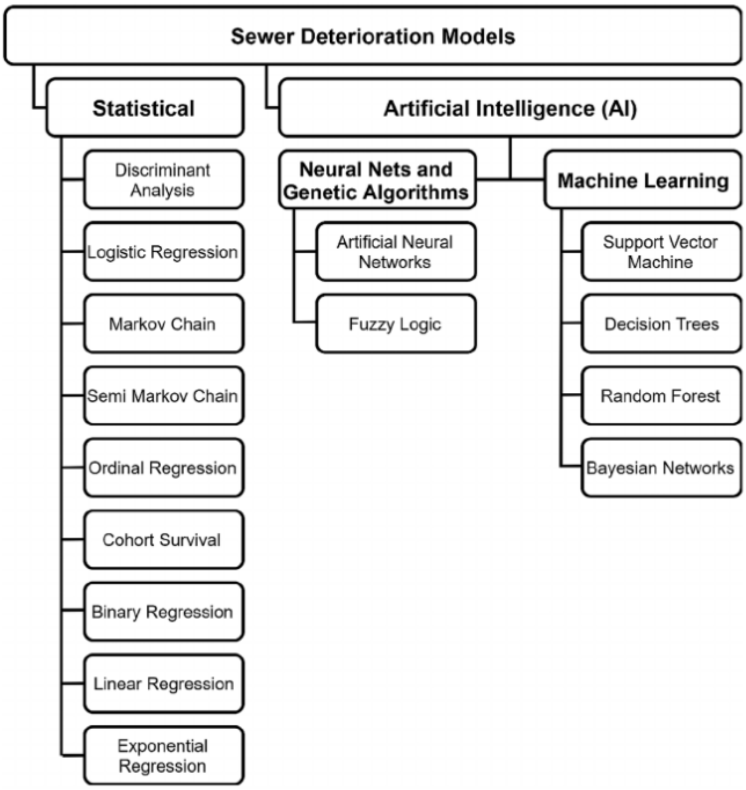


Figure 2-4 Classification of sewer deterioration models. Retrieved from Mohammadi et al. (2019, p. 4).

A variety of techniques are employed in model-based assessments. Statistical models use the probabilistic nature of historical data to characterize the model output as a random variable (Mohammadi et al., 2019, p. 5). A variety of statistical models have been used in previous studies to predict sewer pipe condition, including linear regression, exponential regression, logistic regression, Markov chain, ordinal regression, and cohort survival models (Chughtai & Zayed, 2007; Jeong et al., 2005; Zafar Khan et al., 2010, p. 170). They rely on large datasets for calibration and are commonly applied to forecast condition changes over time (Ana et al., 2009). More recently, AI-based models, including artificial neural networks (ANNs), decision trees, and random forests, have gained popularity due to their ability to capture complex, non-linear relationships between input variables and pipe condition (Mohammadagha et al., 2025; Nguyen & Seidu, 2022).

In Sweden, efforts are underway to integrate these methods into asset management practices. Studies have shown that predictive models can significantly improve the efficiency of inspection planning by identifying high-risk pipes before failures occur, thereby reducing unnecessary inspections and optimizing the use of limited resources (Mahamud, 2023). For example, condition prediction models using AI, trained on historical CCTV and GIS data have been used to support reinvestment planning and reduce subjectivity in condition grading for drinking water network (Sörensen et al., 2024).

These models also enable scenario analysis, allowing municipalities to test the impact of different rehabilitation strategies, renewal rates, or budget levels on long-term network performance. Combined with GIS platforms, model outputs can be spatially visualized, helping engineers and planners prioritize interventions based on both technical and geographical risk (Ghavami et al., 2020, p. 275). While model-based approaches require reliable input data and initial calibration, they offer a powerful complement to traditional inspection methods. When properly implemented, they enable a more proactive, objective, and cost-effective approach to sewer condition management.

2.2.3 Emerging and state-of-the-art methods

To address the shortcomings of traditional and model-based sewer assessments, many municipalities are turning to new technologies that offer more proactive and data-driven approaches to infrastructure management. One area gaining significant attention is the use of artificial intelligence and machine learning. Techniques such as artificial neural networks and decision trees can analyze historical inspection data to predict the condition of sewer pipes, helping utilities prioritize inspections and reduce reliance on subjective assessments (Mahamud, 2023; Nguyen & Seidu, 2022). These tools are also being used to automatically detect defects in CCTV footage, which not only speeds up the evaluation process but also improves consistency.

Another promising development is the use of autonomous inspection robots. Equipped with high-resolution cameras, lasers, or ultrasonic sensors, these robots can navigate through pipes to collect detailed structural information, particularly in areas that are difficult or unsafe for manual inspections (Ahrary et al., 2007, p. 23). The integration of Internet of Things (IoT) technologies is also growing, with sensors installed in sewer systems to continuously monitor flow, gas concentrations, and pipe vibrations (Gerlin et al., 2023, pp. 1-2). This real-time data makes it possible to detect problems early and enables a shift toward condition-based maintenance strategies.

Additionally, the concept of digital twins, dynamic, digital replicas of physical sewer networks is beginning to take hold. These systems combine real-time sensor data, hydraulic modeling, and historical inspection records to simulate network performance, evaluate maintenance scenarios, and support long-term planning (Wang et al., 2024, pp. 21-25).

2.3 AI-based condition assessment methods

Advancements in artificial intelligence in the recent past have opened up new opportunities in infrastructure monitoring, asset management, and failure forecasting in civil engineering. In sewer infrastructure, AI-based systems provide quicker, more precise, and more objective analysis, and hence they serve as a useful addition to conventional techniques.

2.3.1 Overview of AI in civil infrastructure

Artificial intelligence (AI) is increasingly being integrated into the management of civil infrastructure, offering new capabilities for analyzing complex data, identifying patterns, and supporting decision-making processes (Khan, 2025, p. 1021). In general terms, AI refers to systems or algorithms that can perform tasks traditionally associated with human intelligence, such as recognizing patterns, making predictions, or classifying data. Within this broader field, machine learning (ML) and deep learning (DL) represent two of the most prominent subfields. Machine learning enables systems to improve performance based on experience or data, while deep learning, an advanced form of ML, relies on neural networks with multiple layers to model intricate relationships in large data sets (Figure 2-5).

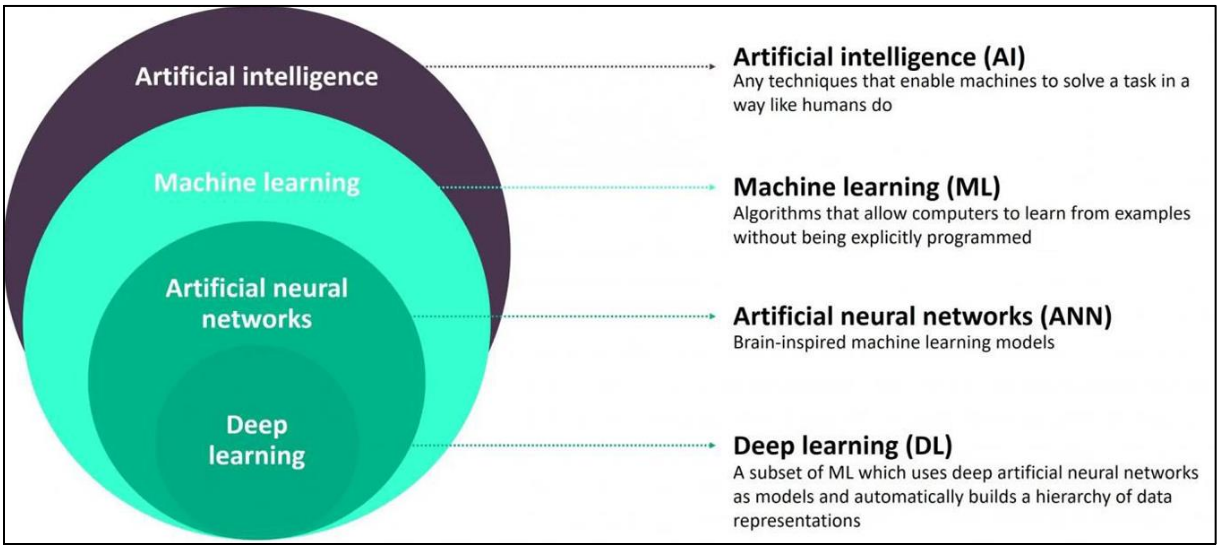


Figure 2-5 Relationship between Artificial intelligence, Machine learning and Deep learning (Ojha, 2024, p. 1021).

In civil engineering, AI has found applications in areas such as structural health monitoring, predictive maintenance, and resource optimization (Vinayak, 2024, p. Results). These methods are particularly relevant in the water sector, where extensive underground networks and aging assets pose challenges for effective monitoring and long-term planning (Mohammadi et al., 2019, p. 4). In the context of sewer systems, AI has emerged as a powerful tool for enhancing condition assessment and maintenance strategies (Salihu et al., 2023, pp. 2-3). By analyzing large volumes of inspection data, such as CCTV footage, operational logs, and pipe characteristics, AI models can support more accurate and objective evaluations of asset condition.

One key application is in condition classification, where AI is used to automatically identify defects and assign condition grades to sewer pipes (Li et al., 2023, p. Abstract). This reduces reliance on manual interpretation and helps standardize the assessment process. AI is also employed to predict the likelihood of pipe failures based on variables such as age, diameter, material, and environmental conditions (Salihu et al., 2023, pp. 2-3). These predictive models assist utilities in identifying high-risk assets before failures occur, enabling more proactive and cost-effective maintenance planning. Beyond prediction, AI is increasingly used to support investment decisions by analyzing inspection data alongside financial and geographical information. This allows municipalities to prioritize rehabilitation efforts based on both technical condition and long-term planning needs (Mohammadi et al., 2019).

Among various AI techniques, Artificial Neural Networks (ANN) have gained particular attention due to their ability to model complex, nonlinear relationships inherent in sewer network data. Their flexibility and high predictive performance make ANNs suitable for forecasting pipe conditions and anticipating failures, supporting enhanced decision-making for asset management. The following section 2.3.2 explains how ANNs work, and previous studies that have effectively employed ANN models compared with other AI models, to assess and predict the condition of sewer systems.

2.3.2 Artificial Neural Networks in predicting sewer condition

2.3.2.1. Artificial Neural Networks

The concept of AI originated from the pioneering work of Warren McCulloch and Walter Pitts in 1943, who developed the first artificial model inspired by biological neurons. Their innovative approach drew upon knowledge of neuronal physiology, propositional logic, and Turing's theory of computation, forming a foundation for numerous subsequent computational models inspired by the human brain (Russell & Norvig, 2010, p. Introduction).

AI is broadly defined as “the study of mental faculties through computational models” (Charniak, 1985, p. Chapter 1). Artificial Neural Networks (ANNs), a subset of AI methods, replicate biological neural systems through multiple layers, each composed of computational units knows as neurons (Figure 2-6 and Figure 2-7)(Walczak & Cerpa, 2003, p. I.A.). These networks learn by identifying patterns in historical data and generalize learned relationships to predict outputs for new inputs. Typically, ANN structures consist of three primary layers: input, hidden, and output layers (Jiang et al., 2016, p. 53). Various training algorithms have been developed for training ANNs that result in different model types. The primary object of employing these algorithms is to identify the optimal relationship between input and output parameters (Huang & Le, 2021, p. Section 2.1.3). Therefore, it is crucial to select a suitable learning algorithm for effective ANN training. Typically, the neural network modeling process includes four main steps: 1) preprocessing the input datasets, 2) building a network model, 3) determining the neural network architecture, and 4) model optimization (Figure 2-8).

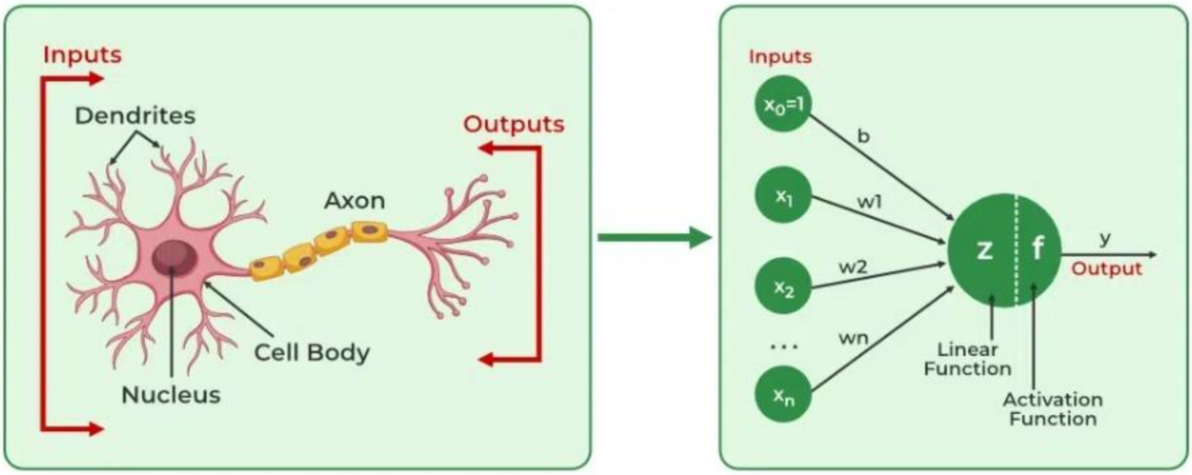


Figure 2-6 The analogy between a biological neuron and an artificial neuron, showing how inputs are received and processed to produce outputs in both systems (Prabhu, 2024).

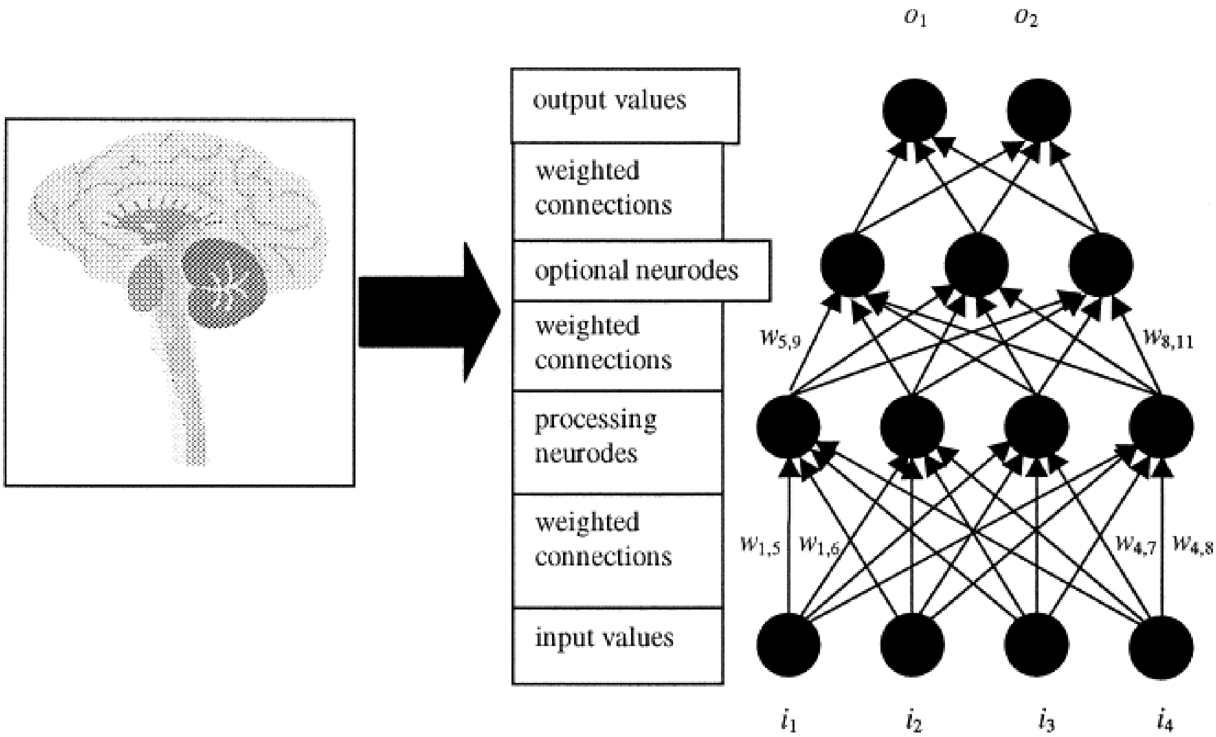


Figure 2-7 Sample artificial neural network architecture (not all weights are shown) (Walczak & Cerpa, 2003, p. I.A.).

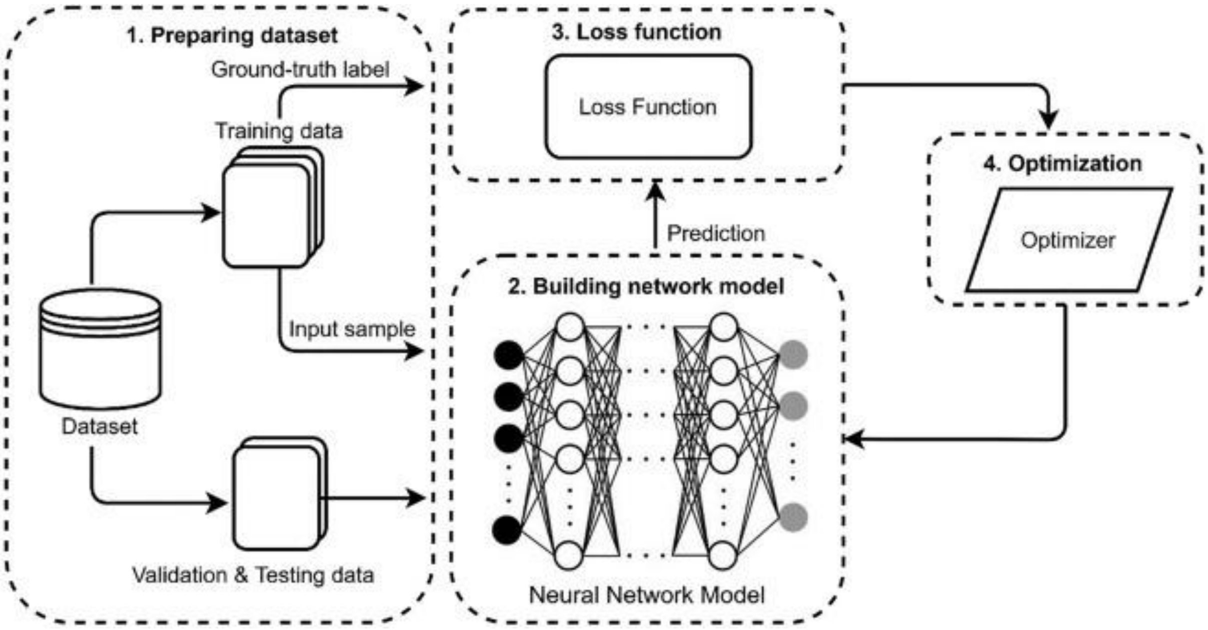


Figure 2-8 Schematic diagram of the neural network training procedure (Huang & Le, 2021, p. Section 2.1.3)

ANNs have been used to predict the deterioration of sewer pipelines. A comparative review of prediction accuracy of AI condition assessment models in sewer management is given in Table 2-5. The table summarizes various research studies comparing predictive models used to assess sewer pipe conditions. Among these, Artificial Neural Networks (ANNs) frequently show superior accuracy compared to other modeling methods. For example, Z. Khan et al. (2010) observed that Back Propagation Neural Networks (BPNN) yielded more reliable predictions than Probabilistic Neural Networks (PNN). Similarly, Sousa et al. (2014) and Jiang et al. (2016) found ANN models more effective than traditional methods such as Support Vector Machines, Logistic Regression, and Multiple Linear Regression. Notably, Alsaqqar et al. (2017) and Mohammadagha et al. (2025) reported high prediction accuracies of 87.3% and 90.6%, respectively, further validating ANN effectiveness. Alternative modeling approaches, including rule-based fuzzy logic studied by Li et al. (2019) and rule-based simulation by Hawari et al. (2017), achieved relatively good but slightly lower accuracy. Interestingly, Hahn et al. (2002) showed that expert systems tended to produce conservative predictions, likely influenced by limited and low-quality input data.

While the reviewed studies highlight the strong performance of ANNs when trained on comprehensive and robust datasets, real-world sewer datasets often present challenges such as missing values, limited failure labels, and class imbalance (Rokstad & Ugarelli, 2015, p. Abstract). In the present study, although the dataset is large and well-maintained by the water utility KoV, it still reflects common issues found in sewer infrastructure data, including rare failure observations. These limitations were considered during model development, especially when selecting preprocessing methods and performance metrics.

Table 2-5 A summary of prediction accuracy of AI condition assessment models for sewer network. Retrieved from Hawari et al. (2020, pp. 6,16); Malek Mohammadi (2019, pp. 71-73).

Author(s)	Model type	Prediction Accuracies

Z. Khan et al. (2010)	Artificial Neural Networks (ANN)	Back Propagation Neural Networks (BPNN) provided better predictions than Probabilistic Neural Network (PNN).
Sousa et al. (2014)		ANN model provided better predictions than Support Vector Machines and Logistic Regression models.
Jiang et al. (2016)		ANN model provided better predictions Multiple Linear Regression.
Alsaqqar et al. (2017)		ANN model provided an overall prediction with an accuracy of 87.3%.
Najafi and Kulandaivel (2005)		Back Propagation Neural Networks (BPNN) is found to be feasible to develop condition prediction model for pipes, although the model accuracy is highly dependent on larger and more inclusive sample size.
Tran (2007)		Performance of neural network calibrated with Markov chain is better than neural network calibrated with backpropagation method.
Mohammadagha et al. (2025)		ANN model provided robust performance metrics with an accuracy of 90.6%.
(Sousa et al., 2014)		ANN model provided better performance than Support Vector Machines with 73-81% correct predictions.
Li et al. (2019)	Rule based fuzzy logic	Adaptive Neuro Fuzzy Inference System (ANFIS) provided better predictions than Multi Linear Regression.
Hawari et al. (2017)	Rule based simulation	Rule based simulation model provided overall prediction with an accuracy of 82%.
Hahn et al. (2002)	Expert systems	Expert system provided 55% similarity in predictions, 37% higher conservative predictions and 8% less conservative predictions. This was justified as low condition data was used in model development.

2.4 Urban wastewater sewage system in Sweden

Sweden's wastewater system reflects more than a century of urban development, public health reform, and environmental protection. From the mid-19th century, when sanitation was primitive and epidemics common, to the present-day advanced but aging wastewater systems, the Swedish reaction has evolved in response to technical, political, and ecological pressures (SWWA, 2007, p. 8). One of the features of this development has been the evolution from combined, as was the norm in early 20th-century cities, to separated and more sustainable systems with advanced levels of treatment technology (SWWA, 2019, p. 16). Nonetheless, despite the system's expansion and modernization, a considerable number of pipes installed during periods of intense construction, particularly the Million Programme housing wave of the 1960s and 1970s, are approaching or have surpassed their designed lifespan (SWWA, 2007, p. 8). This chapter offers a general introduction to Sweden's wastewater system, its history, volume, material, and present state, to contextualize the necessity for data-driven and predictive maintenance techniques such as the ANN-based model that this work aims to develop.

2.4.1 A historical overview of wastewater treatment in Sweden

The development of Sweden's wastewater infrastructure started in response to severe public health issues during the 19th century (SWWA, 2007, p. 8). Cities were afflicted by inadequate sanitation, with sewage frequently discharged directly into surrounding water bodies, causing epidemics of diseases such as cholera (Larsson, 2015, p. Introduction). By the late 1860s and early 1900s, organized efforts to improve urban hygiene were in place. Underground sewer pipes had been laid in 80 cities, marking the first organized extension of urban wastewater infrastructure, see Figure 2-9 (SWWA, 2019, pp. 16-18).

During the 1930s, Sweden began constructing mechanical wastewater treatment plants, introducing the first level of pollutant removal (SWWA, 2019, p. 16). Combined sewer systems (transporting both stormwater and sewage) dominated at this point, especially in dense urban areas. The 1950s and 1960s saw the introduction of biological treatment, while chemical treatment technologies became common in the 1970s. Around the same time, many municipalities began implementing separate sewer systems, particularly in new suburbs (SWWA, 2019, p. 17).

One of the major milestones was the Million Programme (1965–1975), a state-driven housing programme that resulted in the building of large housing areas and related sewerage systems (SWWA, 2007, p. 8). More than 100,000 km of municipal pipes are estimated to have been constructed, much of which is still in operation today (EPA, 2022, p. 5). In the 1990s, Sweden shifted its focus toward nitrogen and phosphorus removal to achieve stricter environmental requirements, taking a leading role in nutrient removal technologies (EPA, 2022, p. 6).

By the 2000s, the country had almost universal sewer coverage. However, a new challenge emerged: aging infrastructure. According to a report by Svenskt Vatten (Swedish Water & Wastewater Association), the annual renewal rate of wastewater pipes is only 0.4%, a rate considered insufficient for maintaining current quality in the long term of the wastewater system (Malm, Horstmark, Jansson, et al., 2011, p. 11). As the infrastructure is aging, there is a growing recognition of the need for strategic maintenance and rehabilitation planning.

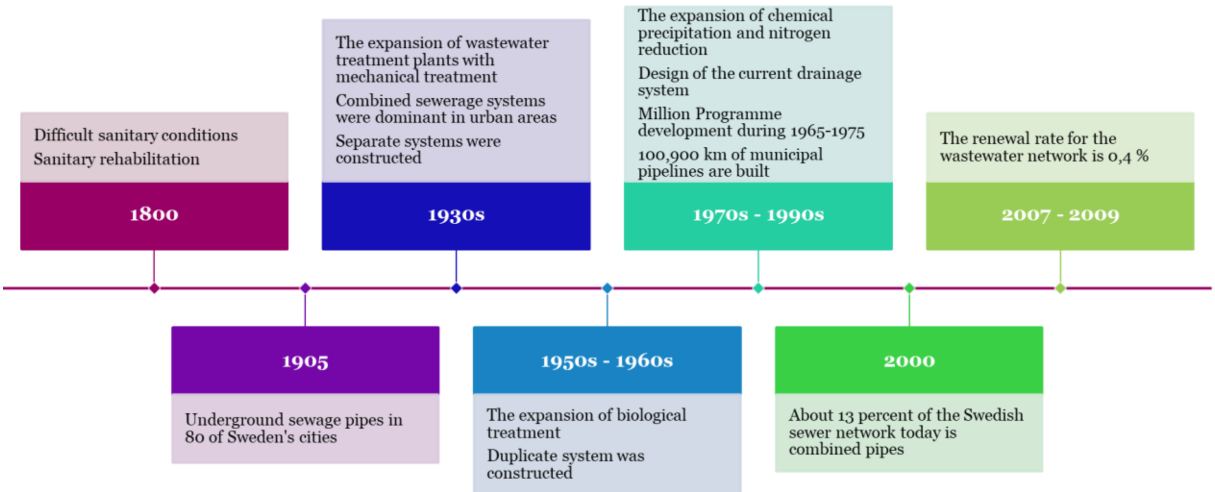


Figure 2-9 Wastewater treatment timeline in Sweden. Adapted from Swedish Water and Wastewater P110 (SWWA, 2019).

2.4.2 An overview of modern wastewater system

Sweden’s sewer infrastructure today is a mature, decentralized system, and its operation is undertaken mostly by local authorities. It is a combination of the older combined sewer systems that manage sewage and stormwater in the same pipe together, and the separate systems with separate networks for rainwater and wastewater. This structural divide reflects the historical development of Swedish cities, where older city areas (e.g., Gothenburg or Malmö’s city centers) have combined systems, while the majority of the city have separate systems (Haghighatafshar et al., 2018, p. 62).

Sweden’s sewage pipes and stormwater pipes amount to approximately 76,740 kilometres and 42,000 kilometres respectively, as of 2023, serving a population of more than nine million (SWWA, 2023, p. 32). The design, maintenance, inspection and renovation of each respective pipe network are typically the responsibility of each municipality. Centralized management of water is facilitated by Swedish Water & Wastewater Association, which issues technical guidelines, promotes standardization and encourages sustainable development (Blomkvist et al., 2023, p. 4).

2.4.2.1. Component of sewer system

Wastewater can be of the following origins (SWWA, 2007, p. 10):

- Wastewater: contaminated water from homes, hospitals, schools, hotels, restaurants, offices, shops etc. (domestic wastewater) and in industries, laboratories, laundries, car care facilities, mechanical workshops, landfills (industrial wastewater)
- Drainage water: groundwater and soil water that is discharged into a pipe, ditch or drainage screen to dewater land areas, e.g. from building foundations.
- Stormwater: surface runoff in the form of rainwater and meltwater from yards, plots, streets, roads, parks and squares, and roofed surfaces. infiltration of groundwater, lake water or sea water.

The components of a typical Swedish sewer system include:

- Gravity pipes, which constitute most of the system and rely on elevation slops to carry flow
- Pumping stations, which are utilized where terrain prevents gravity flow
- Pressure pipe, which are sealed and pressurized by pumps to push wastewater through the system
- Manholes and inspection chambers, which allow access for cleaning, maintenance, and CCTV inspection
- Stormwater inlets and culverts, designed to handle runoff from urban surfaces
- Wastewater treatment plants (WWTPs), which serve as terminal nodes where collected wastewater is processed before discharge into recipient (SWWA, 2007).

Figure 2-10 (Lundberg, 2021, p. 5) illustrates a schematic view of a conventional wastewater pipe network.

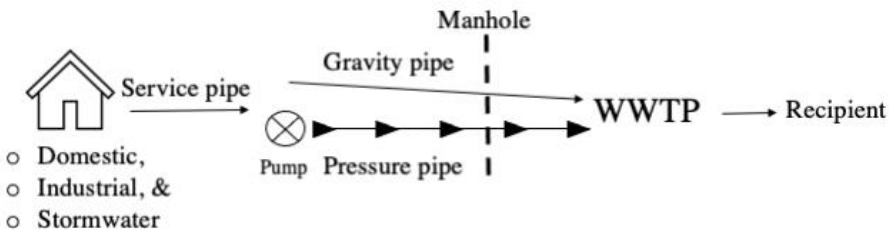


Figure 2-10 Schematic view of a conventional wastewater pipe network including origin, service pipe, type of pipe and pumps, manhole, WWTP and recipient. Retrieved from Lundberg (2021).

Table 2-6 provides a comparative overview of combined, duplicate and separated sewer systems, all of which are present within Sweden's municipal networks due to the historical development of its urban areas.

Table 2-6 Overview of different sewer systems (SWWA, 2007, p. Sewer systems).

System type	Sewage	Stormwater	Drainage	Comments
Combined system	Wastewater, stormwater and drainage water are diverted in the same pipe.			A pipe in the system. Skimmer (Bräddavlopp) is a necessary system function in combined systems.
Duplicate system	Diverted into a separate pipe, possibly together with drainage water.	Diverted into a separate pipe, possibly together with drainage water.	Primarily discharged together with stormwater. In special cases, drainage water can be discharged together with wastewater.	At least two pipes in the system. Drainage water can be diverted in different ways even within the same area.
Separate system	Diverted into a separate pipe, possibly together with drainage water.	Diverted into ditch or LOD [3] system, possibly together with drainage water.	Discharged either together with wastewater or together with stormwater in a ditch or separate pipe.	A pipe and a ditch system, possibly including LOD in the system. Drainage water can be diverted in different ways even within the same area.

2.4.2.2. Management strategies

Network management involves routine maintenance and long-term planning for replacement, typically led by data held in geographic information system (GIS) or asset management systems. However, as Okwori et al. (2024, p. 2) highlight from a national survey, there are considerable challenges for Swedish municipalities to integrate pipe condition data into strategic decision-making. Issues of non-standardized data formats,

---

[3] LOD (Lokalt omhändertagande av dagvatten) system: Local stormwater management. Stormwater is managed on the own property, instead of being discharged into stormwater systems or sewers.

limited use of predictive analytics, and fragmented digital systems are among those that hinder the efficient monitoring and prioritization of aging infrastructure. Moreover, climatic changes in Sweden, such as more precipitation and seasonal snowmelt, put further pressure on urban drainage systems. Research in cities such as Helsingborg (Semadeni-Davies et al., 2008, p. Background) and Malmö (Haghighatafshar et al., 2018, p. 62) demonstrates that climate resilience and adaptability of systems are becoming critical considerations in sewer planning and design. With aging mid-20th century pipes nearing the need of their design life, municipalities are compelled to become more proactive (SWWA, 2021a, p. Introduction). Whereas conventional renewal practices are reactive and expensive, new practices like data-driven approaches provide new opportunities to target high-risk pipes and schedule maintenance before failures happen (Hawari et al., 2020, p. Introduction).

2.4.3 Pipe materials

The materials used in Sweden's sewer pipe networks reflect over a century of infrastructure development, with each era shaped by innovation in construction techniques, environmental policy and material design. Material choice is critical in determining the structural integrity, lifespan, and likelihood of failure of the system. Historically, Swedish municipalities have selected materials depending on local availability, economy and durability in local soil and climate conditions, see Figure 2-11.

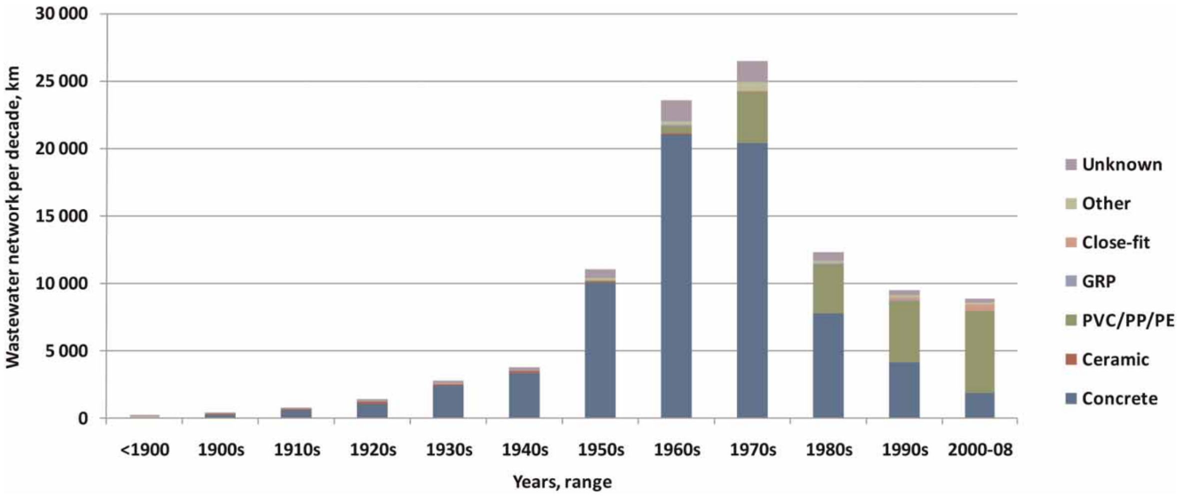


Figure 2-11 Material and age distribution for the Swedish wastewater network, in total 100,900 km (Malm et al., 2013).

Distributions of wastewater pipe networks in 2008 and 2016 are compared in Figure 2-12. Previously, concrete pipes accounted for 66,6% of all pipes, followed by plastic pipes 23,7%, and structure pipes (Malm, Horstmark, Larsson, et al., 2011, p. 12). In modern days with newer technologies, plastic pipes have been dominating new installations, see Figure 2-12 (left). Plastic pipes include PVC (Polyvinyl Chloride), PP (Polypropylene), and PE (Polyethylene). Together with other materials, plastic pipes add up to the 416 km newly installed pipes in Sweden 2016, contributing to over 100,000 km long of Swedish total wastewater pipes in 2017 (Lundberg, 2021, p. 7).

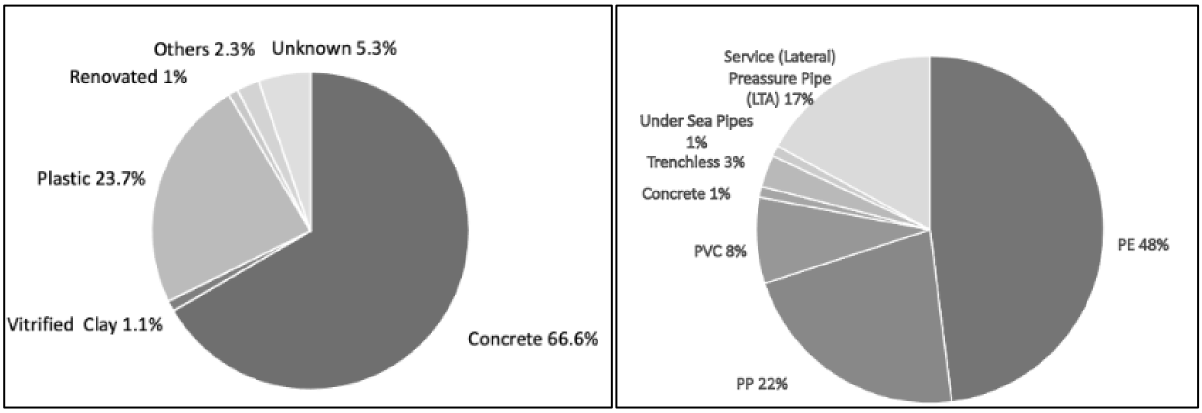


Figure 2-12 Distribution of the wastewater pipe network in 2008 (left), adapted from (Malm, Horstmark, Larsson, et al., 2011, p. 12) and newly installed wastewater pipes in Sweden 2016 (right), adapted from SWWA (2018). Figures are retrieved from Lundberg (2021, pp. 7,8).

Concrete pipe has been the dominant material for wastewater and stormwater infrastructure and has represented a large proportion of total pipe length – especially in infrastructure constructed prior to the 1980s (Malm et al., 2013). Early concrete pipes (pre-1950) also tended to be of poorer quality and more prone to cracking and infiltration, while pipes laid after the 1970s typically offer improved structural integrity due to regulatory changes and better jointing techniques (Malm, Horstmark, Larsson, et al., 2011, pp. 12,13,15). Cast iron (gråjärn), which was common in pressure and house connection lines, has largely been replaced by ductile iron (segjärn) and plastic materials. Although cast iron provided good compressive strength, it tends to corrode and is prone to brittle failure under poor soil conditions.

Starting in the 1970s, there was a significant shift toward plastic materials – primarily PVC and later PE and PP (Malm, Horstmark, Larsson, et al., 2011, p. 26). These materials gained favor due to their light weight, chemical resistance, and easy to install. PVC came to be extensively used for gravity sewer systems, whereas PE, being flexible and long-lasting, is generally used for pressure pipes and stormwater lines. In recent decades, plastic materials have increasingly replaced concrete and iron in new developments. However, long-term performance data for newer plastics are still limited, and their actual lifespan remains somewhat questionable.

Lifetime estimates for pipe materials are predicted on installation time, environmental factors, and construction quality. According to Malm et al. (2013, p. 227), pre-1950 concrete pipes tend to reach a median service life of 60-100 years, while newer concrete pipes have the capacity to last 110-140 years or more. Similarly, older cast iron pipes may reach 80-110 years, while newer corrosion-protected ductile iron pipes may exceed 140 years. Plastic pipes – especially those laid after 1980, are expected to have median lifespans of 100-150 years under normal conditions. Nevertheless, poor installations, corrosive soil conditions, or hydraulic pressure can drastically reduce the operational lifespan. The Swedish sewer system illustrates a complex material history of materials, with older generations of iron pipes and concrete existing alongside new plastic systems. Knowledge of these material characteristics is vital for effective infrastructure planning, risk analysis, and modelling based approaches. As local governments strive to enhance their rehabilitation programs, full material data will continue to play a central role in failure prediction and investment planning.

2.4.4 Sewer pipes' condition

The structural and operational condition of sewer pipes are critical factors in managing the long-term performance and reliability of urban wastewater systems. In Sweden, much of the sewer network was constructed during the mid-20th century, particularly in connection with the Million Programme (SWWA, 2007, p. 8). As a result, many pipes are now approaching or exceeding their expected service lives. Deterioration due to material aging, soil conditions, corrosion and hydraulic stress is increasingly common, especially in concrete and cast iron pipelines installed prior to the 1980s (Malm, Horstmark, Larsson, et al., 2011; Malm et al., 2013).

To monitor and assess the state of this aging infrastructure, Swedish municipalities rely heavily on systematic condition assessments, such as visual inspection techniques, particularly CCTV surveys. The findings from these inspections are evaluated using a standardized scoring method known as the "kortbetyg" system, currently defined in Swedish Water and Wastewater Association P122 (SWWA, 2021b). This method assigns weighted scores to defects identified during CCTV inspections, based on their severity (graded 1-4) and type (e.g., deformation, fracture, root intrusion), see example in Table 2-7. The result is a calculated score for each pipe section, split into two categories: Structural pipe defects (rörfel) and Operational flow-related defects (driftfel). These scores are then mapped to a grading scale from 1 (very good) to 5 (very poor), guiding maintenance priorities and renewal planning. However, consistent national statistics on condition distribution are not yet publicly compiled, and condition grading varies in frequency and completeness across municipalities.

Table 2-7 Example of grading system according to P122 guideline (SWWA, 2021b).

Description	Code	Character	Grading	Weight coefficient
Structural Failure				
Deformation	DEF		1	6
			2	18
			3	54
			4	100
Cracks	SPR	LONGITUDINAL	1	3
			2	6
			3	20
		CIRCULAR	1	2
			2	4
			3	16
		COMPLEX	1	4
			2	8
			3	24
Pipe Break	RBR		2	54
			3	75
			4	100

Surface Damage	YTS		1	0,1
			2	6
			3	54
			4	100
Operational Failure				
Infiltration	INL		1	0,01
			2	3
			3	24
			4	60

Despite a vast network of sewer infrastructure across Sweden (SWWA, 2023, p. 32), renewal rates remain low. Over the past five years, the renewal rate for wastewater pipes has ranged between 0.46% and 0.52% per year, and for stormwater, between 0.28% and 0.33%. These figures fall significantly below the 0.6%-0.7% renewal rate often cited as necessary to maintain sustainable system performance (Malm & Svensson, 2011, p. Summary).

Municipal wastewater systems are also increasingly challenged by inflow and infiltration (I/I), with Swedish networks reporting an average I/I volume of 40.4 m³ per kilometer of pipe per day (SWWA, 2023, p. 32). As described in section 2.1.1.2, I/I is generally considered an unplanned but partly controllable factor in sewer system performance.

Aging infrastructure, particularly deteriorated pipe joints and structural defects such as cracks, are commonly identified as key pathways for stormwater and groundwater infiltration into the sewer system (Zeydalinejad et al., 2024, p. Theoretical background). As a result, high I/I levels are often considered an indirect indicator of the overall structural vulnerability of the network. When combined with limited inspection frequencies and deferred rehabilitation efforts, these operational stresses increase the risk of system failures.

2.5 Summary of research gaps and insights

Previous studies clearly highlight growing concerns about the condition of Sweden's aging sewer infrastructure. Much of the pipe network was built in the mid-1900s, and many of these pipes are now nearing or have exceeded their expected lifespan. At the same time, climate change, increased urbanization, and low renewal rates are putting additional pressure on the system. Traditional maintenance approaches, which are mainly based on visual inspections and manual grading, have been important tools for municipalities. However, these methods are often expensive, time-consuming, and reactive. With limited budgets and resources, many municipalities struggle to keep up with the need for inspections, repairs and long-term planning.

Over the past decade, various prediction models have been developed to support condition assessment and maintenance planning. Among these, ANNs have shown strong potential. This AI technique can handle complex datasets and identify patterns that are difficult to detect. Several studies have shown that ANN-based models can predict sewer pipe conditions with higher accuracy compared to more traditional approaches. Despite this, the use of AI tools in Swedish water utilities is still limited. Most municipalities rely on conventional methods and often lack the technical frameworks needed to make full use of their data. This creates a clear gap between what is possible and what is currently being done.

The goal of this thesis is to explore how ANN models can be applied in a Swedish context, using data from Kretslopp och Vatten in Gothenburg. The aim is to test whether ANN-based predictions can help identify pipe failures and support better decision-making around inspection and renewal planning. This work will also look at how these models can be integrated into existing digital systems and what limitations or challenges might arise in practice.

3 METHODOLOGY

This section outlines the methodology for assessing sewer pipeline conditions using Artificial Neural Networks (ANN). The approach integrates data preprocessing, feature selection, model development and evaluation to predict pipeline deterioration and failure. The methodology is informed by the literature and adapted to the specifics of the dataset provided by the municipality.

3.1 Proposed framework for sewer condition assessment

This study has a four-phased methodology (Figure 3-1). The Literature Review phase encompasses the critical review of sewer infrastructure in Sweden, sewer deterioration causes and condition assessment models. Data Collection and Model Development phase involves data collection, data preprocessing, feature selection, model training and performance evaluation for developing an ANN predictive model for sewer condition assessment. In the Result and Discussion section, the study interprets findings, evaluates the performance of the ANN model and discusses its predictive capability. Finally, the Conclusion and Recommendations section summarizes key findings and makes recommendations for future research areas.

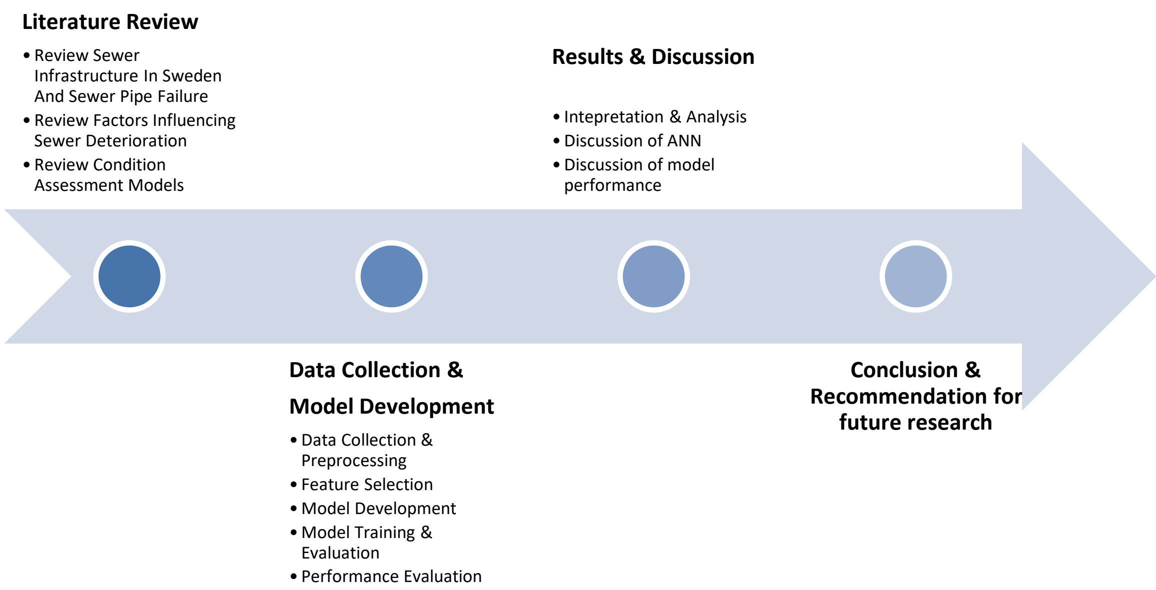


Figure 3-1 The framework for study methodology.

The following flowchart (Figure 3-2) presents the methodology for development of ANN failure prediction model. The process diagram standard Business Process Model and Annotation is utilized Allweyer (2016). Each step is described with more detail in Chapter 4 Current study.

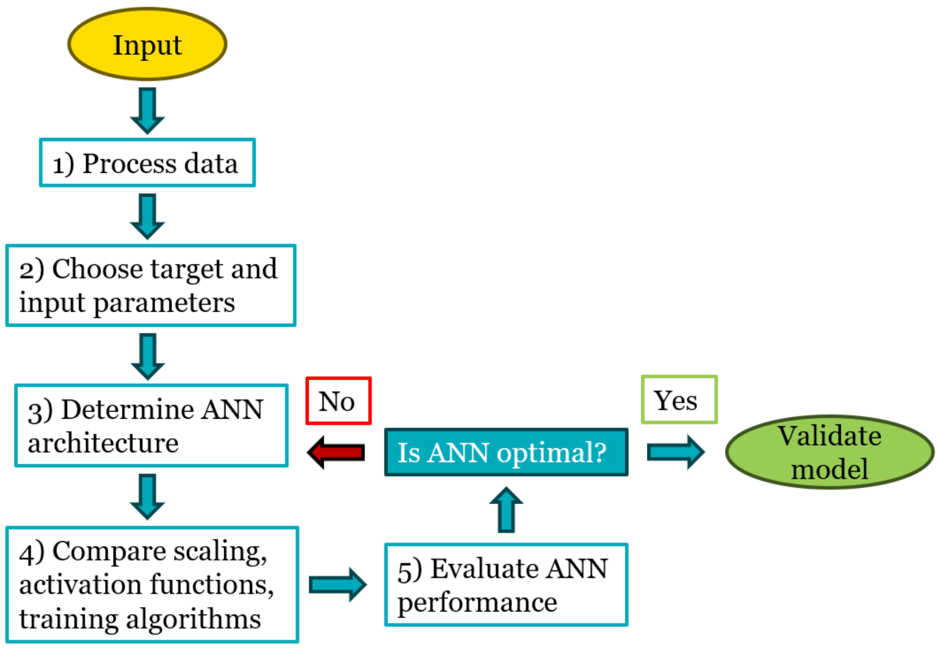


Figure 3-2 Methodology for development of ANN failure prediction model. Adapted from Allweyer (2016); Kerwin et al. (2023).

3.2 Data requirements and processing approach

The success of predictive models for sewer pipe condition assessment relies on the quality and structure of the input data. In many municipalities, particularly those with legacy systems, sewer databases are incomplete, inconsistent, or only partially digitized. These issues often stem from historical gaps in infrastructure documentation, the absence of standardized data collection procedures, and system expansions that occurred before digital asset management practices were common (Malek Mohammadi et al., 2019; Mohammadi et al., 2019, p. Literature review). To ensure model integrity, the first step involves exploratory data analysis and preprocessing.

Exploratory Data Analysis (EDA) serves to develop a comprehensive understanding of available data. An important goal with EDA is to understand the numerical distributions in the data, such as physical attributes and condition ratings across the network (Clementson & Charlesworth, 2025, p. 47). EDA provides the foundation for making informed decisions about model structure and preprocessing requirements. A numerical distribution shows how values in a dataset are distributed or arranged. It describes the values' interval and how often every value or interval occurs. Numerical distribution contributes to the understanding of data's form and spreading, which could show patterns to contribute to a more functional model in the end.

Following EDA, the following preprocessing stages are applied:

- Cleaning data: Handle missing values, remove duplicates, and correct or remove error data points.
- Feature engineering: Create relevant features, remove irrelevant ones, and transform features to appropriate formats (normalization, encoding categorical variables, etc.)

After the preprocessing steps, an evaluation of distribution of the target classes is executed. If significant imbalance is observed, methods for handling this imbalance will be applied.

3.2.1 Data transformation

To ensure that the ANN prediction model is insensitive to units, all input variables are normalized (Isik et al., 2012, p. 873). Normalization transforms each data input to scale it into a range that the network can effectively process. This transformation not only speeds up the training time by aligning the scale of each feature but is also particularly beneficial for modelling applications where inputs vary widely in scale. In this study, two normalization methods are employed for numerical data: (i) statistical normalization and (ii) min-max normalization.

(i) Statistical normalization (also called standard scaling) uses the mean and the standard deviation for each variable across a set of training data to normalize each input variable vector. This transformation produces a dataset such that each variable has a zero mean and a unit variance, see equation (1):

$$y_i = \frac{x_i - \mu}{\sigma} \quad (1)$$

Where:

- x_i = the original raw data,
- y_i = the transformed data,
- μ = the mean,
- σ = the standard deviation of the i^{th} variable of the training data set.

(ii) Min-max normalization is a process that rescales the inputs or output(s) from one range of values to a new range of values. Generally, the variables are rescaled to lie within a range of 0 to 1 or from -1 to 1. The rescaling is executed by:

$$y_i = (x_{target,max} - x_{target,min}) * \frac{x_i - x_{min}}{x_{max} - x_{min}} + x_{target,min} \quad (2)$$

Where:

- $x_{target,max}$, $x_{target,min}$ = the maximum and the minimum target values at the normalized set, respectively,
- x_{min} , x_{max} = the minimum and the maximum values of the raw data set,
- y_i = the transformed data,
- x_i = the raw data

For categorical data such as material, pipe type and soil data, they are encoded to transform non-numerical categories into numerical values that can be effectively processed by the ANN model. While One-Hot encoding is widely used for nominal data and has been applied in sewer condition modeling (e.g., Mohammadagha et al. (2025, p. 7)), it becomes inefficient for categorical variables with many unique values, leading to high dimensionality and increased risk of overfitting (Pargent et al., 2022, p. Results). In contrast, target encoding replaces each category with a smoothed average of the target variable, resulting in a more compact representation that directly captures the relationship between category and outcome. This approach is especially advantageous for high-cardinality features, as it reduces noise, computational cost, and the risk of overfitting, and has been shown in both foundational and recent research to improve predictive performance in machine learning models.

3.2.2 Identify extreme values

In EDA, methods are used to analyze outliers in datasets. One of the popular methods is the Z-score, a measure that indicates how many standard deviations a data point is from the mean in a dataset (Clementson & Charlesworth, 2025, p. 49). The formula for the Z-score is:

$$Z = \frac{X - \mu}{\sigma} \quad (3)$$

Where:

- X = the value of the data point,
- μ = the mean,
- σ = the standard deviation

Typically, data points with a Z-score greater than 3 or less than -3, are considered outliers, as they are far from most of the data. The choice of threshold value depends on the application and can vary between applications. Z-scores are primarily used to identify outliers based on the number of standard deviations from the mean, which works best for normally distributed data.

In addition, the Interquartile Range (IQR) is a measure of the spread in the middle of the dataset and is defined as the difference between the third quartile (Q3) and the first quartile (Q1):

$$IQR = Q3 - Q1 \quad (4)$$

To identify upper and lower bounds for outliers, the following are commonly used:

- Lower bound: Q1 – 1.5 * IQR
- Upper bound: Q3 + 1.5 * IQR

Data points outside these bounds are considered outliers. IQR is effective for identifying outliers in data that is not normally distributed and is generally less sensitive to skewed distributions compared to Z-score. However, IQR-based methods can sometimes miss outliers if the data contains many extreme values or is highly irregular.

3.3 Selection of target and input parameters

The effectiveness of an ANN in predicting sewer pipe conditions is closely tied to how the model's target output is defined. This target depends on the objective and time horizon of the model. For long-term planning, outputs may involve predicting the likelihood of deterioration or failure over several years, while short-term applications might focus on binary outcomes, such as whether a pipe will fail within a specific period. Additionally, the spatial resolution of the model, whether predictions are made at the pipe, segment, or broader network level will affect how input variables are structured and interpreted.

The selection of input variables is informed by a combination of previous research, the scope of the model, and the availability of municipal data. Common predictors identified in the literature include pipe age, diameter, material type, slope, burial depth, surrounding soil type, groundwater conditions, historical inspection or CCTV condition grades, previous failures, maintenance records, infiltration data, and proximity to roads, vegetation, or high traffic zones (Mohammadi et al., 2019, p. 3; Nguyen & Seidu, 2022; Noshahri et al., 2021; Tscheikner-Gratl et al., 2019). These variables are chosen for their documented relevance to sewer pipe deterioration processes and their availability in municipal datasets. For this study, physical attributes, soil attributes and failure types were used for training the model. A detailed description of the chosen input and output parameters for this study are presented in Chapter 4 Current study.

Prior to training, an alternative data split of three subsets was employed to develop and assess the models in this study: training (70%) used to train the model; validation (10%) used to track model parameters and avoid overfitting; and testing (20%) used for checking the model's performance on new data. The following figure recommends this workflow, where "Tweak model" means adjusting anything about the model, such as from changing the learning rate, to adjusting features or to designing a completely new model from scratch.

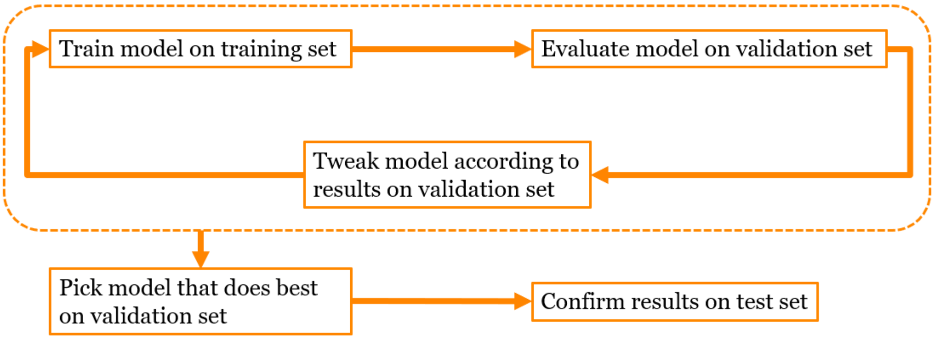


Figure 3-3 Workflow for development and testing. Adapted from Developers Google (2025).

This approach ensures the ANN is trained effectively, while also mitigating overfitting and providing a reliable performance evaluation. Even though this workflow could be optimal, the more the same data is used to make decisions about hyperparameter setting or other model improvements, the less confidence that the model will make good predictions on new data. It is suggested to collect more data to refresh the test set and validation set (Developers Google, 2025).

The careful definition of model outputs and the selection of relevant input variables are foundational to the development of robust predictive models for sewer pipe condition assessment. However, before model training can execute, it is essential to further examine the relationships between these input features and the target variable, as well as address any potential issues related to class imbalance in the dataset.

3.3.1 Correlation analysis

Correlation analysis examines the association between variables to determine whether they move together and the strength of their relationship. The results are expressed as a correlation coefficient, ranging from -1 to +1 (Agahi & Kim, 2021, p. 2)

- +1 = perfect positive correlation (variables increase together)
- 0 = no correlation (variables are independent)
- -1 = perfect negative correlation (one variable increases as the other decrease)

To assess how each input feature relates to the likelihood of pipe failure, Monte Carlo correlation analysis was applied on the training data prior to model development. This method was selected because it can capture uncertainty and handle non-ideal data conditions more effectively than traditional approaches. Conventional correlation measures like Pearson's assume linearity and normally distributed variables (Agahi & Kim, 2021, p. 2), while Spearman's rank correlation requires a monotonic relationship between variables (Curran, 2014, p. 2). Both can produce biased results when data is skewed, noisy or non-linear (Bishara & Hittner, 2015, p. Abstract). In contrast, Monte Carlo approach generates many random subsamples (using bootstrapping[4]), calculates the Pearson correlation for each

[4] Bootstrapping is a statistical resampling method used to estimate the variability (standard errors or confidence intervals) of a statistic – such as the mean, median, or correlation – by generating many simulated samples from the original dataset.

one, and builds a distribution of correlation values (Phuenaree & Sanorsap, 2017, p. 624). This allows the modeler to estimate not just the average correlation but also how much it might vary, which captures confidence intervals and helps to reduce the impact of outliers, sampling noise or class imbalance. The result is a more reliable and generalizable picture of which features matter most.

3.3.2 Class imbalance handling

Having comprehensive and accurate data is important for informed decision-making and effective implementation of data-driven methods. However, sewer network datasets frequently encounter challenges such as data scarcity and class imbalance, which can negatively affect the development of precise failure prediction models (Latifi et al., 2024, p. Chapter 2). In machine learning (ML) applications for predicting pipe failures, class imbalance describes a scenario where the dataset has an uneven distribution among different classes. Specifically, this occurs when the number of pipes that have experienced failures is significantly lower than those without failures, which negatively impacts classifier accuracy. Typically, classes are divided into two groups: the majority class, containing the larger number of instances, and the minority class, with fewer instances. The ratios for majority and minority classes in the dataset can be calculated using the following equations:

$$Majority\ class\ ratio = \frac{Number\ of\ majority\ samples}{Total\ number\ of\ samples} \qquad (5)$$

$$Minority\ class\ ratio = \frac{Number\ of\ minority\ samples}{Total\ number\ of\ samples} \qquad (6)$$

Where *Number of majority/minority samples* refer to the count of instances belonging to majority/minority classes, respectively.

Class imbalance can negatively impact machine learning models, causing them to disproportionately predict the majority class. Various methods, such as oversampling and undersampling, are recommended to address this imbalance. Oversampling increases minority class samples to achieve better balance, enhancing model accuracy by providing sufficient minority instances. Unlike undersampling, oversampling retains all majority class data, preserving valuable information. However, oversampling may lead to overfitting, particularly if synthetic or duplicated samples are used, resulting in poor performance on unseen data. Additionally, synthetic data generation risks introducing noise or misrepresenting the true distribution. Therefore, thorough experimentation and evaluation are essential when selecting methods to handle class imbalance (He & Garcia, 2009, p. Chapter 4).

Random Over-Sampling (ROS), Synthetic Minority Over-sampling Technique (SMOTE), and Adaptive Synthetic Sampling (ADASYN) are common oversampling approaches. ROS addresses class imbalance by randomly duplicating minority class instances, whereas SMOTE generates synthetic samples to enrich the minority class (Chawla et al., 2002, p. 325). Specifically, SMOTE selects a minority class instance and identifies its k nearest neighbors (k being user-defined) within the minority group. It then synthesizes new samples by interpolating between these instances in feature space, repeating this process until the desired oversampling level is achieved (Fernandez et al., 2018, p. 865).

Class weighting is a method used to address class imbalance in ML by giving different levels of importance (weights) to samples based on their class. Typically, a ML model is trained by minimizing an objective function, such as the cross-entropy loss function, calculated across the entire dataset.:

$$Min \sum_{i=1}^{N} Loss\ (y_i, \hat{y_i}), \qquad (7)$$

Where, $Loss\ (y_i, \hat{y_i})$ calculates error between the observation y_i and prediction $\hat{y_i}$; and N is the number of samples. This technique first calculates the class imbalance ratio:

$$Imbalance\ ratio = \frac{n_{majority}}{n_{minority}}, \qquad (8)$$

In which $n_{majority}$, $n_{minority}$ are the number of instances in majority and minority class. By assigning majority class weight and minority class weight based on the imbalance ratio as:

$$\begin{cases} Class\ weight_{majority} = 1 \\ Class\ weight_{minority} = Imbalance\ ratio \end{cases}, \qquad (9)$$

Class weighting assigns class weights in the loss function as:

$$Total\ Loss = Class\ weight_{majority} \sum Loss_{majority} + Class\ weight_{minority} \sum Loss_{minority} \ (10)$$

By assigning greater weights to the minority class, the model is guided to prioritize reducing errors related to that class during training (Burez & Van den Poel, 2009, p. 4249). Robles-Velasco et al. (2021, p. Section 3.2 Implementation) integrated ROS and Random Under-Sampling (RUS) with ANN to address class imbalance in pipe failure prediction models and assess their effectiveness. Their study revealed that under-sampling improved the accuracy of true positive predictions – that is, cases where the model correctly identifies a pipe that actually has a failure - whereas over-sampling enabled the model to predict failures and non-failures with comparable accuracy. Latifi et al. (2024, p. Conclusion) applied all the three primary approaches: under-sampling, over-sampling and class weighting to address the imbalanceness in pipe dataset, by adjusting the representation of minority and majority classes. The paper revealed that better results were achieved by combining different sampling ratios and applying class weights, with under-sampling showing a stronger influence on model performance than over-sampling. As the dataset could vary from study to study, this degree project employed experiments on these three approaches to evaluate which method has a more pronounced impact on predictive performance. Class weight was chosen in the end as the model was able to learn from the full dataset while still addressing the imbalance in class frequencies, which results in a more robust and generalizable classifier.

3.4 Model Implementation

3.4.1 Basic structure of ANN

ANNs are computational models inspired by the structure and functioning of the human brain, consisting of interconnected processing units called neurons. Similar to how the human brain learns, ANNs adapt by adjusting the weights of these neuron connections based on provided examples. Most neural networks employ supervised learning methods, in which they learn from pairs of input data and corresponding desired output during training sessions (Kadhum et al., 2016, p. 30).

In the context of pipe failure prediction, ANNs have been widely utilized for over two decades (Table 3-1), with the Multi-Layer Perceptron (MLP) being the most common structure (Kerwin et al., 2023, p. Section 1.1). An MLP typically includes an input layer, one or more hidden layers and an output layer. Each layer consists of multiple neurons/ perceptrons, that processing inputs by applying weights and transforming them through activation functions to generate outputs. Training an ANN involves iterative adjustment of weights through supervised learning, commonly using the feed-forward back-propagation (BP) algorithm. In BP, the network repeatedly forwards input data through the layers, compares predicted outputs with actual values, and propagates errors backward to update connection weights. Each complete pass of forward and backward propagation is called an epoch. Training continues until the network meets a predefined tolerance or reaches the maximum number of epochs.

The design of an ANN involves careful configuration of its architecture, including the number of neurons in each layer. Selecting the appropriate number of neurons is crucial to achieving reliable prediction results. Too few neurons may restrict the model's learning capacity, while to many can hinder its ability to generalize, leading to overfitting (Atambo et al., 2022, p. 12).

Developing an ANN for pipe failure prediction involves several steps, see Figure 3-2. As mentioned in subsection 3.3, the dataset is split into a training set (70%) to train the ANN, a validation set (10%) to prevent overfitting, and a testing set (20%) to evaluate the ANN's performance. Figure 3-4 illustrates the basic structure of ANN.

Table 3-1 provides an overview of previous studies that applied ANNs to predict various outcomes related to municipal pipe networks, including pipe condition state, failure rate, number of failures, and time to failure. Most studies utilized MLPs with one or two hidden layers and employed the BP algorithm for supervised learning. Common output variables include failure rate, pipe condition scale, and the number of pipe failures with studies drawing on extensive historical datasets. Performance metrics vary, with many studies using the coefficient of determination R2, root mean square error (RMSE) and mean absolute error (MAE) to assess model accuracy. A few studies introduced alternative training methods or architectures, such as radial basis function networks or extreme learning machines. Atambo et al. (2022) applied a more advanced model with three hidden layers and evaluated performance using ROC-AUC (Receiver Operating Characteristic curve – Area Under the Curve), reflecting a shift toward more nuanced classification performance metrics in recent years. The diversity of inputs and configurations highlights the flexibility of ANNs in modeling different aspects of pipe infrastructure performance.

Table 3-1 Literature on the use of ANN for municipal pipe networks. Adapted and adjusted from Kerwin et al. (2023).

Reference	Inputs[5]	Output	ANN development
Sacluti (1999)	11, 12	Area failure density	MLP (1 hidden layer) with BP. Coefficient of determination (R2) used as performance criteria.
Ahn et al. (2005)	4, 11, 12, 13, 14	Network pipe failures	MLP (2 hidden layers) with BP. Mean absolute error (MAE) used as performance criteria.
Najafi and Kulandaivel (2005)	1, 2, 4, 7, 15, 16, 17	Sewer pipe condition scale	MLP (1 hidden layer) with BP. Root mean squared error (RMSE) used as performance criteria.
Al-Barqawi and Zayed (2006)	1, 4, 6, 7, 15, 18, 19, 20,	Water pipe condition scale	MLP (1 hidden layer) with BP. MAE, RMSE, average invalidity percent (AIP), and average validity percent (AVP) used as performance criteria
Achim et al. (2007)	1, 2, 3, 7, 21	Failure rate	MLP (2 hidden layers) with BP. Coefficient of determination (R2) used as performance criteria.
Geem et al. (2007)	1, 4, 6, 7, 10, 15, 21, 22, 23, 30	Pipe condition index	MLP (1 hidden layer) with BP. Coefficient of determination (R2) used as performance criteria.
Tabesh et al. (2009)	1, 2, 7, 10, 15	Failure rate	MLP (1 and 2 hidden layers) with BP. Index of agreement (IOA) and RMSE used as performance criteria.
Amaitik and Amaitik (2010)	7, 10, 15, 24, 25, 26, 27	Number of broken wires in PCCP	MLP (2 hidden layers) with BP. Coefficient of determination (R2) used as performance criteria.
Jafar et al. (2010)	1, 2, 5, 6, 7, 8, 9, 31	Number of failures	MLP (1 hidden layer). Coefficient of determination (R2) and average squared error (ASE) used as performance criteria.
Asnaashari et al. (2013)	1, 2, 4, 5, 6, 7, 29, 30	Failure rate	MLP (1 hidden layer). Coefficient of determination (R2) and average squared error (ASE) used as performance criteria. 50:25:25 data split used.
Kutyłowska (2017)	1, 2, 3, 4, 5, 17,32	Network failure rate per pipe type	Radial Basis Functions ANN (7-4-3) and MLP ANN (7–14-3) compared. 50:25:25 data split used in training phases. Absolute relative error and correlation coefficient used for performance.
Sattar et al. (2019)	1, 2, 3, 5, 6, 28, 29	Failure rate	MLP (one hidden layer) with extreme learning machine training algorithm instead of BP.
Atambo et al. (2022)	1, 2, 4, 6, 15, 16, 19, 33, 35	Sewer pipe condition	MLP (3 hidden layers) with BP. ROC-AUC used as performance criteria.

[5] 1 – Diameter, 2 – Length, 3 – Construction year, 4 – Material, 5 – Number of previous failures, 6 – Soil type, 7 – Age at failure, 8 – Location underground, 9 – Pressure variation, 10 – Pressure head, 11 – Various air temperature measurements, 12 – Various water temperature measurements, 13 – Various soil temperature measurements, 14- Ratio of metered water, 15- Burial depth, 16 – Pipe slope (sewer), 17 – Pipe type, 18 – Road surface, 19 – Pipe failure rate, 20 – Hazen Williams coefficient, 21 – Geographical coordinates, 21 – Electric recharge, 22 – Bedding condition, 23 – Number of road lanes, 24 – Time between failures, 25 – Soil resistivity, 26 – Soil density, 27 – Wire characteristics, 28 – Cathodic protection, 29 – Cement mortar lining, 30 – Presence of pipe surface coating, 31 – Pipe thickness, 32 – Number of service lines, 33- Age, 34- pH, 35- Corrosion Concrete/Steel

(Mohammadagha et al., 2025)	1, 2, 4, 6, 19, 33	Sewer pipe condition	MLP (2 hidden layers) with BP. RMSE, R2, MAE, and relative absolute error (RAE) used as performance criteria.

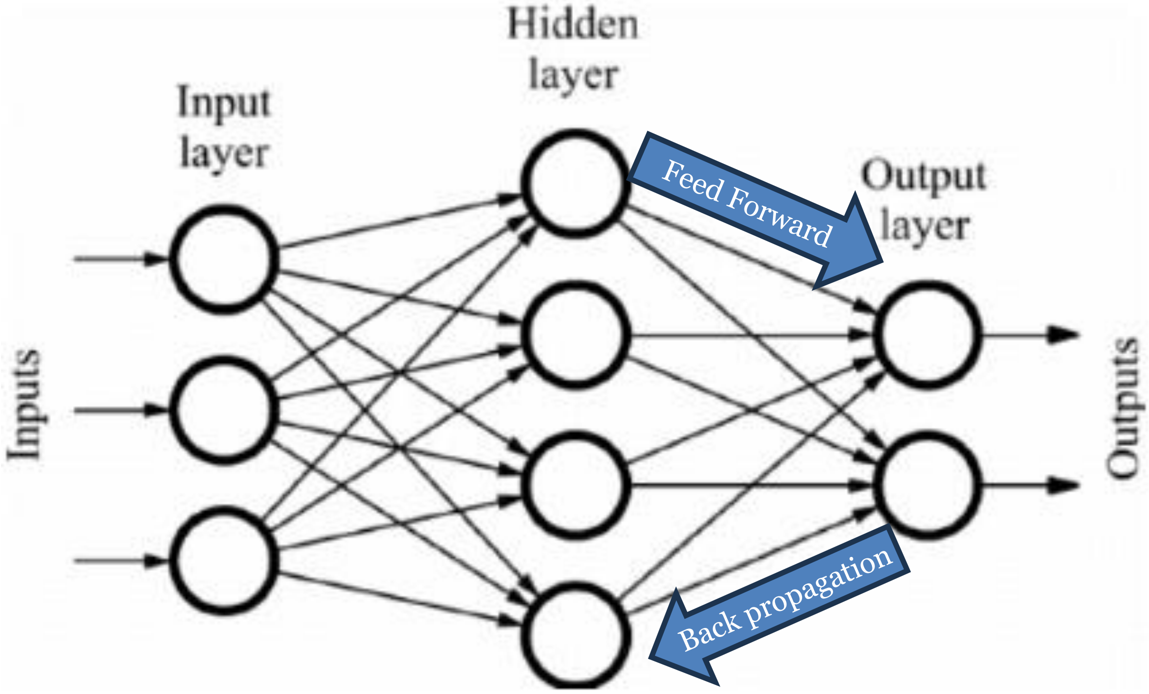


Figure 3-4 Basic structure of Artificial Neural Network with multilayer perceptron with one hidden layer. Adapted from Onorato (2024, p. 10).

3.4.2 Learning process

Neural networks can be trained using a variety of algorithms, depending on the network architecture and the nature of the task. As mentioned in section 3.4.1, one of the most widely adopted approaches is backpropagation – a supervised learning algorithm that iteratively adjusts the weights of the network to minimize prediction errors through gradient descent. In this study, backpropagation was selected due to its effectiveness in training feedforward networks as MLP and its widespread success in similar infrastructure prediction applications (Kerwin et al., 2023, p. Section 1.1; Onorato, 2024, p. 10). As training progresses, the network predicts outcomes based on input data, measures the difference between these predictions and actual values, then modifies its internal weights and biases to reduce this error. Weight optimization typically employs gradient descent techniques, such as Adam optimizer and variations of gradient descent. Early stopping callback was also applied to make sure the model did not train too long and overfit, by stopping when the model stops getting better on validation data.

The fundamental computation in each neuron of the ANN model follows the formula as shown in Eq (11) (Mohammadagha et al., 2025, p. 8):

$$\gamma = f\left(\sum_{i=1}^{n} w_i x_i + b\right) (11)$$

Where:

- γ = the probability of pipe failure (0 to 1),
- f = the activation function,
- w_i = weights assigned to each input feature,
- b = the bias term.

Furthermore, the calculation of the weight updates during backpropagation is performed according to the basic gradient descent rule, as shown in the following equation Eq(12):

$$w_i^{(t+1)} = w_i^{(t)} - \eta \frac{\partial L}{\partial w_i} \qquad (12)$$

Where:

- $w_i^{(t+1)}$ = the updated weight at iteration t+1,
- η = the learning rate controls the step size and how quickly the model learns,
- $\frac{\partial L}{\partial w_i}$ = the gradient of the loss function L (see section 3.4.4), with respect to the weight w_i.

This iterative process is instrumental in ensuring that weights are adjusted to minimize prediction errors. Subsequently, the Adam (Adaptive Moment Estimation) optimization algorithm is employed. This algorithm represents a sophisticated implementation of gradient descent for weight updates during training. The weight update rule follows the principle of gradient descent (presented in Eq(13)), while incorporating adaptive learning rates and momentum:

$$w_i^{(t+1)} = w_i^{(t)} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \qquad (13)$$

Where:

- $w_i^{(t+1)}$ *and* $w_i^{(t)}$ = the weights at time t and t+1, respectively,
- $\hat{m}_t$ = the first moment estimate (mean of gradients),
- $\hat{v}_t$ = the second moment estimate (uncentered variance),
- ϵ = a small positive constant (e.g., 10^{-8}) used to avoid division by zero when computing the final update.

3.4.3 Activation functions

Activation functions are essential mathematical components in neural networks that introduce non-linearity into the learning process (Onorato, 2024, p. 13). They transform the weighted sum of input signals into an output that is passed to the next layer, enabling the network to learn complex data patterns beyond what linear operations can capture. As illustrated in Figure 3-5, each neuron computes a weighted sum of its inputs, adds a bias term, and then applies an activation function to determine its output. This non-linear transformation is what enables deep networks to capture intricate relationships in the data. Without activation functions, the network, regardless of how many layers it has, would behave like a single layer perceptron and be limited to modeling only linear relationships.

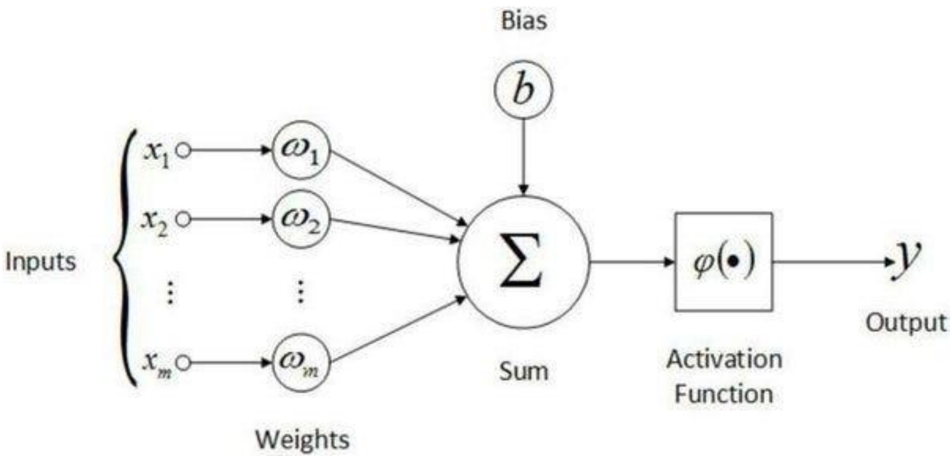


Figure 3-5 Structure of a simple ANN and the role of the activation function. Retrieved from Rallabandi (2023).

The non-linearity provided by activation functions empowers neural networks to model complicated relationships in various data types including time-series, spatial, and categorical inputs (Onorato, 2024, p. 13). Given the range of available activation functions used in neural networks, selecting an appropriate one for each layer is a critical design decision that significantly affects model performance. In this study, LeakyReLU (Leaky Rectified Linear Unit) and ELU (Exponential Linear Unit) activation function were selected for the hidden layers of the ANN after experimentation with other activation functions. Unlike the standard ReLU, which can cause some neurons to become inactive and stop learning (the "dying ReLU" problem), Leaky ReLU allows a small, non-zero gradient when the input is negative (Rallabandi, 2023). This property ensures that all neurons remain trainable throughout the learning process, which is especially beneficial in this type of pipe dataset where the distribution of feature values can vary widely, and some feature may frequently produce negative activations. ELU further addresses the limitations of ReLU by allowing negative outputs that smoothly saturate for large negative inputs, which helps to bring the mean activation closer to zero and can speed up learning and improve model robustness. The output layer used the sigmoid activation function to return a probability value between 0 and 1, which is suitable for binary classification tasks such as predicting pipe failure, as recommended by Mohammadagha et al. (2025, p. 8).

The most common activation functions are shown in Figure 3-6.

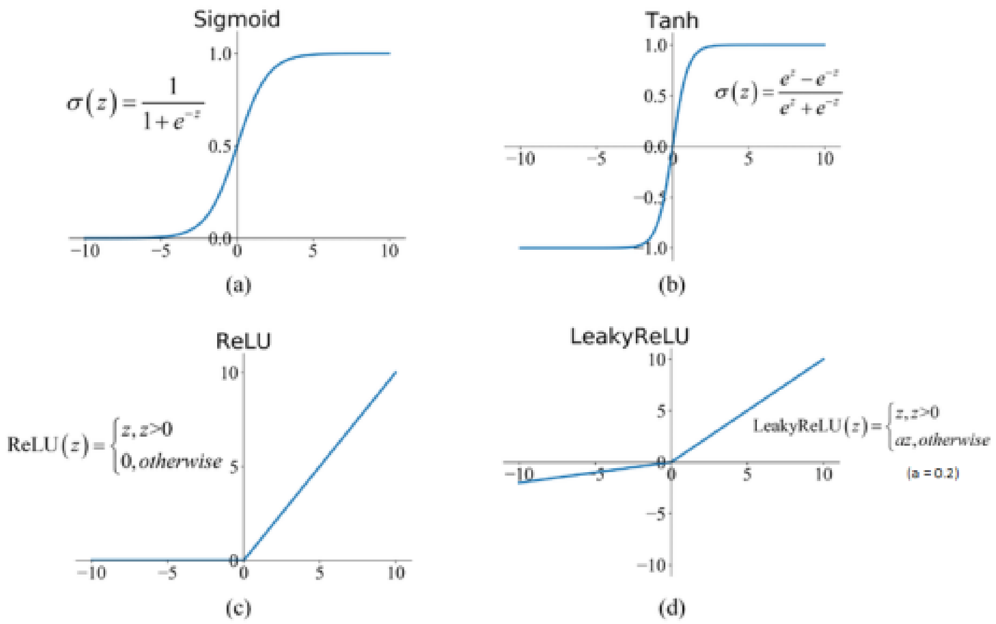


Figure 3-6 The 4 most used activation functions, with LeakyReLU featuring as a hyperparameter. Retrieved from Onorato (2024, p. 13).

3.4.4 Loss function and optimizer

The loss function serves as a critical metric in neural network training as it measures the gap between predicted outputs and actual values at each training step, see Figure 3-7. This measurement guides the network on how to adjust its parameters. Many implementations use the Adam Optimizer to enhance this process by employing backpropagation to calculate precise loss function gradients, then updating weights systematically to improve performance, see Figure 3-8 (Onorato, 2024, p. 15). Adam optimizer was chosen to update weights and biases in the network due to its popularity within pipe condition prediction models from previous studies (Kingma & Ba, 2014; Mohammadagha et al., 2025, p. 8; Sörensen et al., 2024, p. Model development).

The common loss functions for classification tasks are Cross-Entropy Loss (CE), as it measures the performance of a classification model whose output is a probability value between 0 and 1. It is proven to be effective in tasks with imbalanced datasets, compared with other loss functions such as focal loss (Vyas et al., 2023, p. Conclusion). Choosing the right loss function is essential for optimizing ANN performance in classification task. Due to time constraints, this study chose the traditional loss function (CE) instead of exploring other innovative approaches such as hybrid, tangent, and triplet loss functions. It is recommended to explore in further studies to compare and design more effective neural network models for sewer pipe network. More about the chosen loss function is described in 3.4.4.1.

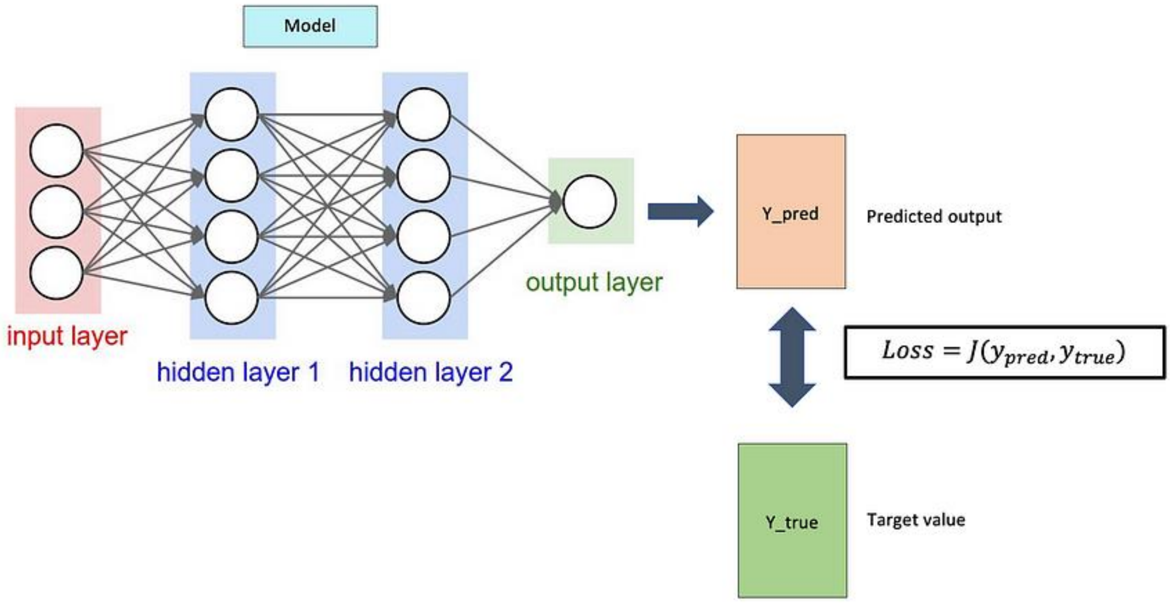


Figure 3-7 A simple neural net to illustrate the purpose of a loss function. Retrieved from Huynh (2023).

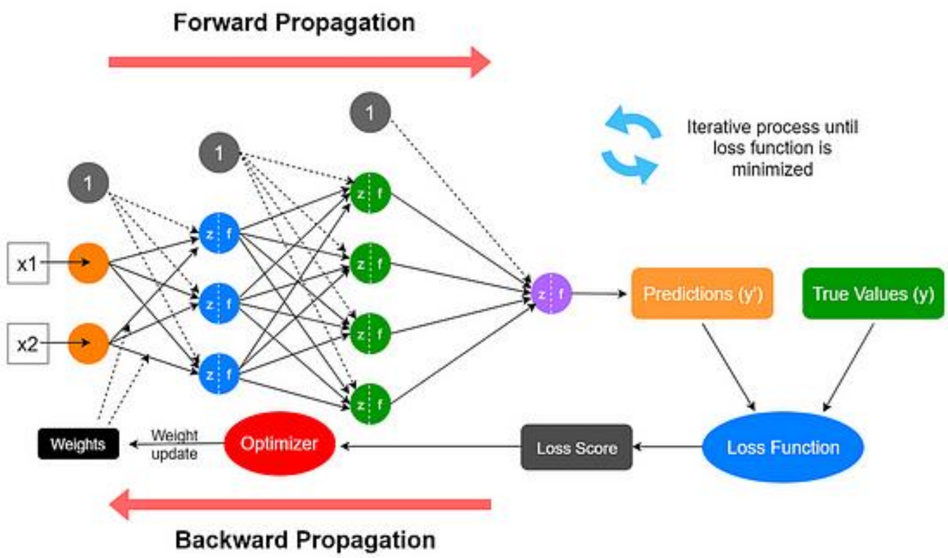


Figure 3-8 Loss function and optimizer during ANN process. Retrieved from Pramoditha and Oak (2024).

3.4.4.1. Loss functions for classification

Binary Cross-Entropy Loss (BCE), also known as Log Loss was chosen as the primary loss function for this study because the task of predicting sewer pipe failure is a binary classification problem, where the model must distinguish between to mutually outcomes: failure or no failure (Onorato, 2024, p. 16). BCE is the standard and most widely accepted loss function for such tasks, as it directly measures the dissimilarity between the predicted probabilities and the actual binary labels (Ruby & Yendapalli, 2020, pp. 5393-5394). Specifically, BCE quantifies the difference between the model's predicted probability (p) for class 1 and the actual binary label (y), which is either 0 or 1, see Eq(14). This metric evaluates the accuracy of binary classification models, typically using a 0.5 threshold to convert predicted probabilities into discrete class assignments (above 0.5 becomes class 1, below becomes class 0).

To address the issue of class imbalance (mentioned in section 3.3.2), where failure events are much rarer than non-failures, class weights were incorporated into the loss calculation. In this weighted binary cross-entropy approach, the loss for each sample is multiplied by a weight corresponding to its class, which is determined by the distribution of the target variable in the training set. This ensures that misclassifications of the minority class (failures) are penalized more heavily, thereby mitigating the effects of class imbalance and improving the model's ability to detect rare failure events (Onorato, 2024, p. 17):

$$BCE(y, p, w) = -w_i(\mathrm{y} \log(p) + w_i(1 - y) \log(1 - p)) \qquad (14)$$

This mechanism works in connection with the sigmoid activation function (mentioned in 3.4.3 and the BCE loss). The output layer of ANN uses a sigmoid activation to produce a probability between 0 and 1 for each sample, representing the model's confidence in predicting pipe failure. The BCE loss then measures the difference between these predicted probabilities and the actual binary labels. By applying class weights within this framework, the model is encouraged to pay greater attention to the minority class, which results in a more balanced and effective classifier. This integrated approach ensures that the model remains sensitive to rare but critical failure events, while still leveraging the full dataset for training.

3.4.5 Hyperparameter tuning using Bayesian Optimization

Hyperparameter tuning is a critical process in optimizing neural network models, as it involves adjusting the parameters that regulate the training process to improve model performance. Neural networks are particularly sensitive to hyperparameters such as the number of neurons, learning rate, and regularization strength and the high dimensionality of the search space makes manual tuning both challenging and inefficient. Common hyperparameter tuning methods are Bayesian Optimization (BO), Grid Search, Random Search, Differential Evolution (DE), Genetic Algorithms (GA), etc. While grid search and random search are straightforward, they can be computationally expensive and may miss optimal regions in high-dimensional spaces due to their exhaustive and random nature. In contrast, Bayesian Optimization is a probabilistic, model-based approach that efficiently explores the hyperparameter space by learning from previous evaluations and focusing on the most promising regions. Previous studies have demonstrated the effectiveness of BO for neural network hyperparameter optimization (Ismail et al., 2024). For example, Taiwo et al. (2025, p. Section 4.2) and Bui and Seidu (2022, p. 124248) used BO for failure prediction models, to effectively optimize batch size, epochs, number of neurons, optimizer, number of filters and learning rate, that eventually maximize validation set performance while avoiding overfitting.

Due to time constraint and the need for efficient optimization, BO was chosen as the primary hyperparameter tuning method in this project. Preliminary experiments with GridSearchCV and RandomizedSearch showed that BO consistently outperformed these methods in terms of both speed and model performance. In practice, the Optuna library was used, which implements advanced Bayesian optimization algorithms such as Tree-structured Parzen Estimator Approach (TPE) and Covariance Matrix Adaptation Evolution Strategy (CMA-ES) to find out the best set of hyperparameters (Bergstra et al., 2011, p. 4). TPE, in particular, models the objective function using two probability density functions – one for good hyperparameter, l(x), and one for bad ones g(x), and samples new hyperparameter sets from regions where good performance is more likely. The goal is to maximize the ratio l(x)/g(x), focusing the search on promising areas of the hyperparameter space:

$$l(x) = P(x|y < y^*)$$

$$g(x) = P(x|y \geq y^*) \qquad (15)$$

where y is the objective function value, and y* is a threshold for good performance.

3.5 Model evaluation metrics

Performance metrics play an important role in evaluating and refining neural network, which offer insights into different dimensions of model performance. The classification workflow consists of three distinct phases: training, validation and testing. During the training phase, the model learns from input patterns (training data) by adjusting its parameters. While the training error indicates how well the model fits the training data, it tends to be optimistically low since the model is evaluated on the same data used for training. The ultimate goal is to predict class labels for unseen data in the testing phase. However, since the true labels for test data are unknown, the testing error cannot be directly estimated. This limitation is the reason why the validation phase is added as it serves two purposes: providing an unbiased evaluation of the model's performance and facilitating the tuning of hyperparameter.

For classification tasks, the most used evaluation tools are (Onorato, 2024, p. 18; Tharwat, 2021, pp. 171-178):

- **Confusion matrix**
  Figure 3-9 illustrates the structure of a 2x2 confusion matrix, a common tool for visualizing the performance of a binary classification model. It categorizes predictions into four outcomes: true positives (TP) for correctly identified positive cases and true negatives (TN) for correctly identified negative cases (representing correct predictions). Incorrect predictions include false negatives (FN), where actual positives are missed (Type II error), and false positives (FP) where actual negatives are

incorrectly flagged as positive (Type I error/ false alarm). This matrix serves as a basis for computing several key classification evaluation metrics discussed below.

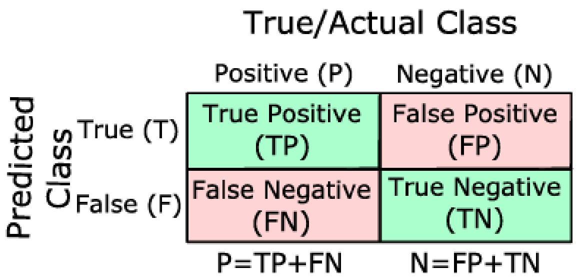


Figure 3-9 Example of 2x2 confusion matrix with two true classes P and N. The output of the predicted class is true or false. Retrieved from Tharwat (2021).

- **Accuracy** is one of the most commonly used measures for classification performance, as it is calculated by dividing the number of correct predictions (TP+TN) by the total number of samples in the dataset:

$$Acc = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \ (16)$$

where P and N indicate the number of positive and negative samples, respectively.

- **Sensitivity and specificity** are crucial for understanding a classifier's performance beyond simple accuracy. Sensitivity (Recall/ True Positive Rate, TPR) measures the model's ability to correctly detect positive instances (detect correctly failed pipes). Specificity (True Negative Rate, TNR) assesses the model's ability to correctly identify negative instances and represent the proportion of actual negatives that are correctly classified (detect correctly non-failed pipes).

$$Sensitivity\ (Recall) = \frac{\text{TP}}{\text{TP+FN}};\ Specificity = \frac{\text{TN}}{\text{TN+FP}} \ (17)$$

- **Precision** represents the accuracy of the model's positive predictions. In the context of this study, it represents the proportion of pipes predicted to fail that actually failed, calculated as the ratio of true positives (correctly identified failures) to the total number of positive predicted samples as follow:

$$Precision = \frac{\text{TP}}{\text{TP+FP}} \ (18)$$

- **F1-score** merges precision and recall into a single balance metric, particularly valuable when working with imbalanced datasets as in this study. F1 ranges from 0, indicating the worst possible classification performance, to 1, representing perfect performance.

$$\text{F1}= \frac{2*Precision*Recall}{Precision+Recall} \ (19)$$

- **Receiver Operating Characteristics (ROC)** curve and its associated **Area Under the Curve (AUC)** capture the relationship between a classifier's sensitivity (recall) and its specificity at different probability thresholds. A summary statistic for this curve is the AUC-ROC, which is independent of the chosen threshold and falls between 0 and 1. Higher AUC-ROC values correspond to better classification performance between classes.

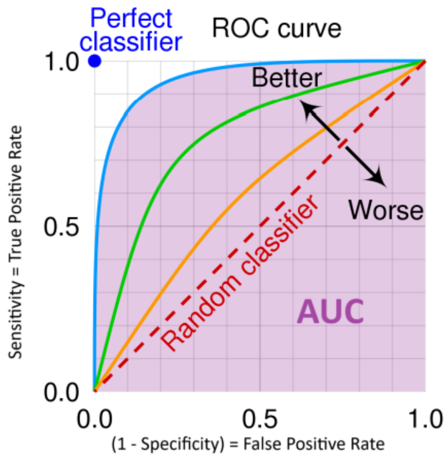


Figure 3-10 ROC curve and AUC (in pink). Retrieved from Onorato (2024, p. 18).

In addition to the above metrics, feature importance was analyzed by aggregating the absolute weights of the first dense layer by feature category. In line with recommendations from the KoV utility, risk group stratification was employed as an intuitive and practical approach to interpret model predictions for asset management. Instead of relying solely on raw probability outputs, the predicted failure probabilities for each pipe segment were divided into discrete risk groups. This stratification enables utilities to categorize pipes into groups such as “very low risk”, “low risk”, “medium risk”, “high risk”, and “very high risk” of failure. For each group, the observed failure frequency and the number of failures per kilometer were calculated and visualized. This approach provides a clear and actionable summary of model results, allowing decision-makers to prioritize inspection, maintenance, or replacement activities based on risk level rather than arbitrary probability thresholds.

3.6 Research ethics considerations

The dataset used in this study was managed exclusively by Kretslopp och Vatten (KoV) throughout the research. To protect sensitive infrastructure information, each pipe segment was identified by a non-descriptive ID, with all geographic and spatial references removed. The author was granted access only to this anonymized version of the data. This setup provided the necessary data for model development while respecting KoV’s information security standards.

Working with infrastructure-related data involves more than concealing location information as it also requires strict access control. The dataset was stored in encrypted environments, and access was governed by agreements established at the start of the project. While the dataset may be categorized as open data within KoV’s internal framework, it has not been considered public in the conventional sense yet. This distinction shaped the data management protocols followed during the research.

All communication of results, both in this thesis and through the ANN model, was guided by a review process designed to prevent the disclosure of sensitive information. From the outset, the ethical framework acknowledged the tension between using infrastructure data for technical insight and protecting it as a critical public asset. This awareness informed key methodological choices during model construction.

4 CURRENT STUDY – DATA AND MODEL

This chapter describes the methodology and data sources used in the current study, providing a comprehensive outline of the research approach and implementation.

4.1 Site description

This case study focuses on Gothenburg's sewer pipe network, managed by Kretslopp och Vatten (KoV), the municipal water and wastewater utility. Gothenburg is Sweden's second-largest city, with more than half a million residents in the city and over one million in the metropolitan area (Figure 4-1) (SCB, 2025). Located on Sweden's west coast, the city experiences a maritime climate with average annual precipitation of 1049 mm, creating significant demands on the wastewater infrastructure (SMHI, 2021; StormTac, 2025, p. 46).

The network comprises 80,007 pipe segments with a total length of 2,387 kilometres. Sanitary sewers represent 43% of the network, stormwater sewers 42%, and combined sewers 15%. Concrete (BTG) is the predominant material (83%), followed by polypropylene (PP) at 7%. The Artificial Neural Network (ANN) models developed in this research are trained using data from a subset of the entire sewer network within Gothenburg to predict pipe failures.

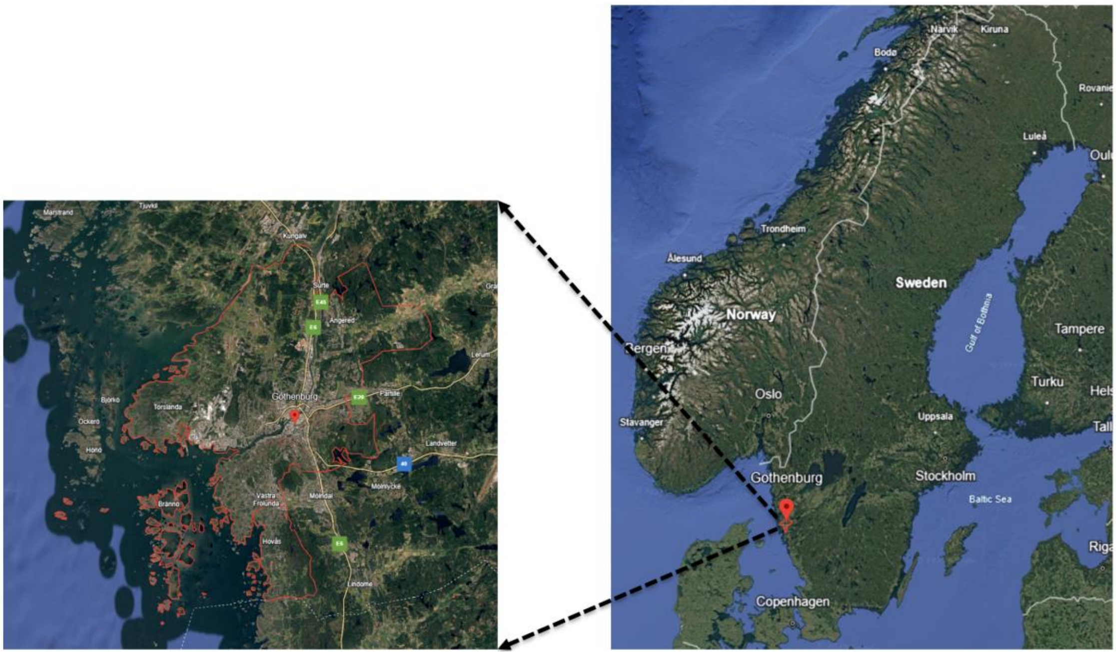


Figure 4-1 Location of the study area, Gothenburg, Sweden. The left panel shows a zoomed-in view of the Gothenburg municipal area, while the right panel indicates its location within Scandinavia.

4.2 Data sources and database characteristics

4.2.1 Overview of available databases

Three primary databases were utilized in this research: a pipe database containing physical factors, a failure database with inspection records, and a Geographic Information System (GIS) database with soil type and environmental factors. Table 4-1 summarizes the parameters used in each database.

The Pipe and Failure databases were merged into a single database (Microsoft Excel file) by Kretslopp och Vatten (2025). The Pipe database contains detailed information about the physical attributes of the pipes, including Pipe ID, Pipe Type, Construction year, Diameter, Material, Length and Renovation status (year and method). For this degree work, the focus is on five recorded failure types that hold data related to the conditions of the pipes: Infiltration (INL), Rupture (RBR), Cracks (SPR), Surface Damage (YTS), and Deformation (DEF).

The GIS database includes geographical data for the pipes, such as Soil data (soil type, soil change, soil change distance), High traffic load, Near building, Tramway, Watercourse, etc. The geographical data (excluding Soil data) contains binary values: 0 (absence) and 1 (presence). For most of these factors, the value 0 significantly outnumbers the value 1, creating an imbalance that could lead to biased modeling results (Haixiang et al., 2017, p. 222). After preliminary trials with the data, environmental factors and renovation status with severe imbalances were excluded from further analysis. For soil data, the feature "soil_change" represents the transition between soil type, while "soil_change_dist" provides information about the distance from a particular pipe to the nearest soil transition. For example, a "soil_change" value of "40,890" in the dataset indicates a transition of glacial clay to bedrock (or vice versa), with the lowest soil type ID always listed first. A "soil_change_dist" of "61.391" means that the shortest distance from the pipe to this soil transition is approximately 61 meters. If the distance value is equal to 0, it means that the pipe directly intersects the soil transition.

Table 4-1 Parameters found in databases and used as input (pipe & soil databases) and output (failure database) in the ANN model(s).

Parameters	Pipe database	Failure database	GIS database
Pipe ID	X		
Pipe Type	X		
Construction year	X		
Age (2025-construction year)	X		
Diameter	X		
Material	X		
Length	X		
Infiltration (INL)		X	
Rupture (RBR)		X	
Cracks (SPR)		X	
Surface Damage (YTS)		X	
Deformation (DEF)		X	
Soil type			X
Soil change			X
Soil change distance			X

4.2.2 Inspection and failure records

Out of a total of 80,007 pipes in the database, 11,179 pipes (14%), with a total length of 389.15 km have been inspected through CCTV, while the remaining 68,828 pipes (86%) have not undergone inspection (Kretslopp och Vatten, 2025). This relatively low inspection rate results from KoV's prioritization strategy, where inspections are typically conducted based on reports of potential issues rather than as routine procedures (Kretslopp och Vatten, 2025; Laakso et al., 2019, p. 3). This is a common approach in wastewater management due to resource constraints. However, this inspection strategy introduces a potential selection bias in the dataset, as inspected pipes are likely overrepresented with failures compared to the overall network. This bias occurs because inspections are primarily triggered by suspected problems rather than random sampling. Consequently, the model trained on this data might overestimate the failure risk for uninspected pipes that share similar characteristics with the inspected ones. Despite this limitation, the ANN model still provides valuable insights by learning from the characteristics of already-inspected pipes and their associated failure types. The model's predictive capability can support utilities like KoV in identifying high-risk pipes that have not yet been inspected, thereby helping to prioritize inspections more strategically. However, it is important to acknowledge that the model's risk estimates should be interpreted with this potential bias in mind, and the results should be validated against actual failure data when possible.

The severity of each failure is graded on a scale from 1-4, following the system defined by Svenskt Vatten P122 (SWWA, 2021b, pp. 2,4). In this system, higher grades indicate more serious conditions, see Figure 4-2 for examples of failure types. A summary of grading ranges and their interpretations is provided in Table 4-2. However, the accuracy and consistency of the recorded failure types remain uncertain and may vary depending on inspection quality.

For model development, the subset of data corresponding to inspected pipes (14% of the total dataset) was used, as inspection data provides direct information about pipe conditions. This limitation reflects the current availability of inspection data rather than a deliberate design choice to reduce model complexity. While this approach ensures usable training data, the model may not fully capture the variability present in the entire pipe network and the model will not be applicable to the parts of the network that are comprised of parameters not included in the training set. Future work could explore integrating additional features from uninspected pipes to improve generalization.

Table 4-2 Failure types and grading system. Adapted from SWWA (2021b).

Failure Type	Grading Range	Explanation
Infiltration (INL)	1-4	Infiltration through pipe joints or cracks.; higher grades indicate more persistent and larger leaks
Rupture (RBR)	2-4	Pipe breakage means that pipe sections are misaligned or missing (grade 2-3). Collapse (grade 4) means that the pipe's cross-section has changed, i.e., the pipe is only supported by the surrounding backfill.
Cracks (SPR)	1-3	Severity and type of cracks (longitudinal, circumferential, complex). One or more cracks are visible on the pipe wall, but the pipe cross-section is unchanged and all parts are still in place.

Surface damage (YTS)	1-4	Degree of surface material deterioration affecting pipe integrity
Deformation (DEF)	1-4	Degree of pipe shape distortion impacting flow and structural stability. Expressed as a percentage change of the pipe cross-section. Used for flexible pipes.

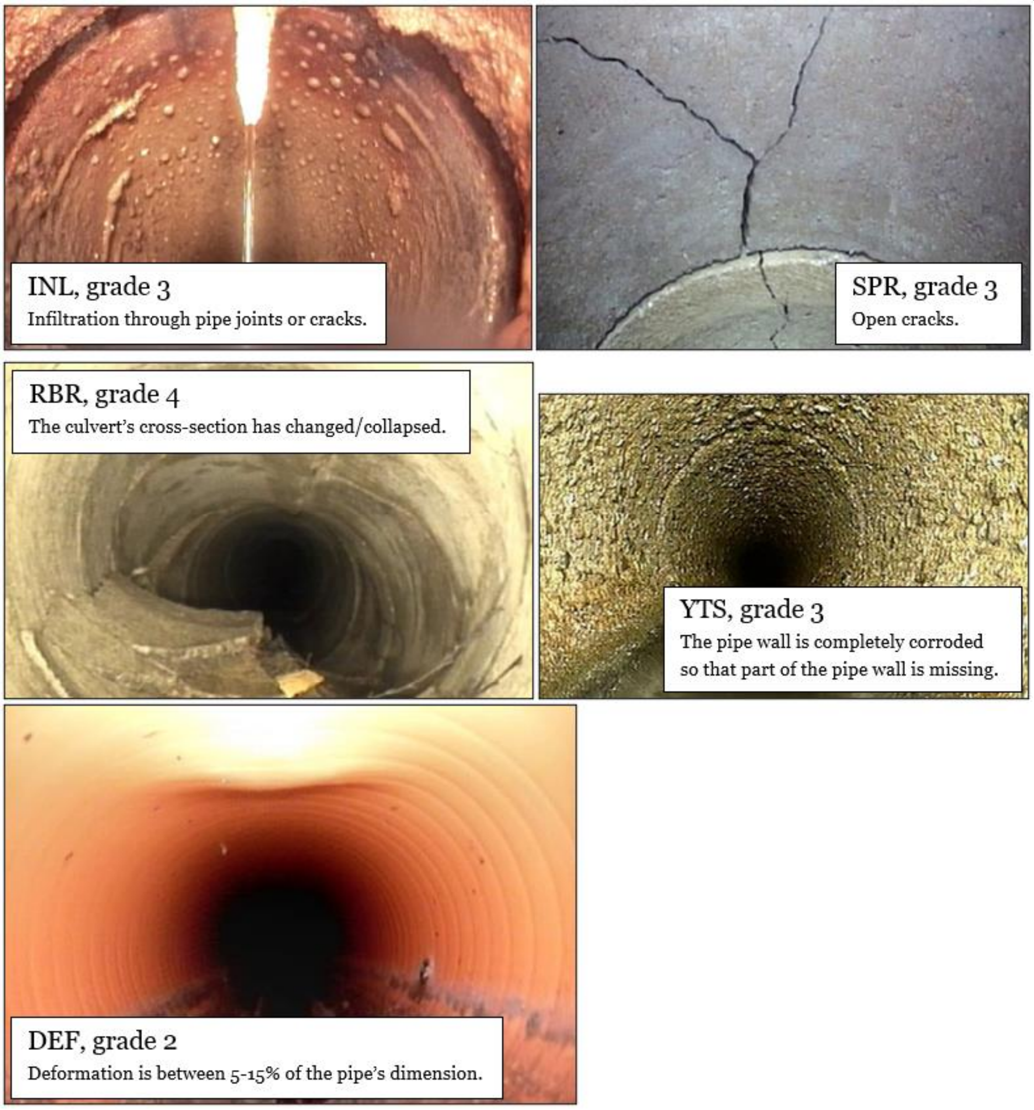


Figure 4-2 Examples of failure types, retrieved from SWWA (2021b).

4.2.3 Data overview for inspected pipes

A total of 11,179 pipes with inspection data were included in the modeling subset, each described by a set of numerical and categorical features (see Table 4-3). The target variables are five binary indicators (see section 4.3) representing different failures types INL, SPR, RBR, YTS and DEF. The following subsections provide an overview of the dataset, including descriptive statistics for the numerical features, the distribution of the most common categories for the categorical features and the correlation between each failure type and the input parameters.

Table 4-3 Example of input data structure for model development. INND = Diameter, LLANG = Length, FTYP = Pipe Type, soil_change = Soil transition and soil_change_dist = Soil Change Distance.

	Numerical variables				Categorical variables			
ID	AGE **(year)**	INND **(mm)**	LLANG **(m)**	soil_change_dist **(m)**	MATERIAL	FTYP	soil	soil_change
9001	3	225	0.13	41.8732	PP	S	40	40,200
19	12	160	0.15	16.0099	PP	S	890	17,890
9000	3	250	0.17	41.7764	PP	S	40	40,200
8635	3	160	0.18	3.7959	PP	D	40	40,890
161	9	110	0.21	10.6035	PP	S	31	31,890

4.2.3.1. Numerical features

The analysis of numerical features revealed important characteristics of the sewer pipe network. The inner diameter (INND) of inspected pipes ranges from 75 mm to 2200 mm, with a mean diameter of 410.66 mm. The pipe lengths (LLANG) vary considerably, ranging from 0.13m to 367.69 m, with an average length of 34.81 m. The age distribution of the pipes shows significant variation, with pipes ranging from newly installed (0 years) to 160 years old, with a mean age of 46.53 years. Additionally, the distance to the nearest soil transition (soil_change_dist) ranges from 0 to 370 meters, with most pipes located within 50 meters of a soil boundary. These distributions highlight the heterogeneity of the network and the importance of accounting for a wide range of physical and environmental factors in predictive modeling.

Summary statistics and distribution of numerical features are presented in Table 4-4 and Figure 4-3:

Table 4-4 Summary statistics of numerical features. Count = Number of non-missing values for the feature; mean = arithmetic mean (average) of the feature; std = standard deviation (spread) of the feature; min = minimum value observed; 25% = 25th percentile (first quartile); 50% = 50th percentile (median); 75% = 75th percentile (third quartile); max = maximum value observed.

Statistics	**AGE** (year)	**INND** (mm)	**LLANG** (m)	**soil_change_dist** (m)
Count (pipes)	10017	11137	11179	11176
mean	46.52	410.65	34.81	48.44
std	29.13	319.45	24.04	56.57
min	0.0	75.0	0.13	0.0
25%	14.0	225.0	15.13	8.18
50%	52.0	250.0	30.5	28.89
75%	63.0	450.0	51.35	69.53
max	160.0	2200.0	367.69	385.44

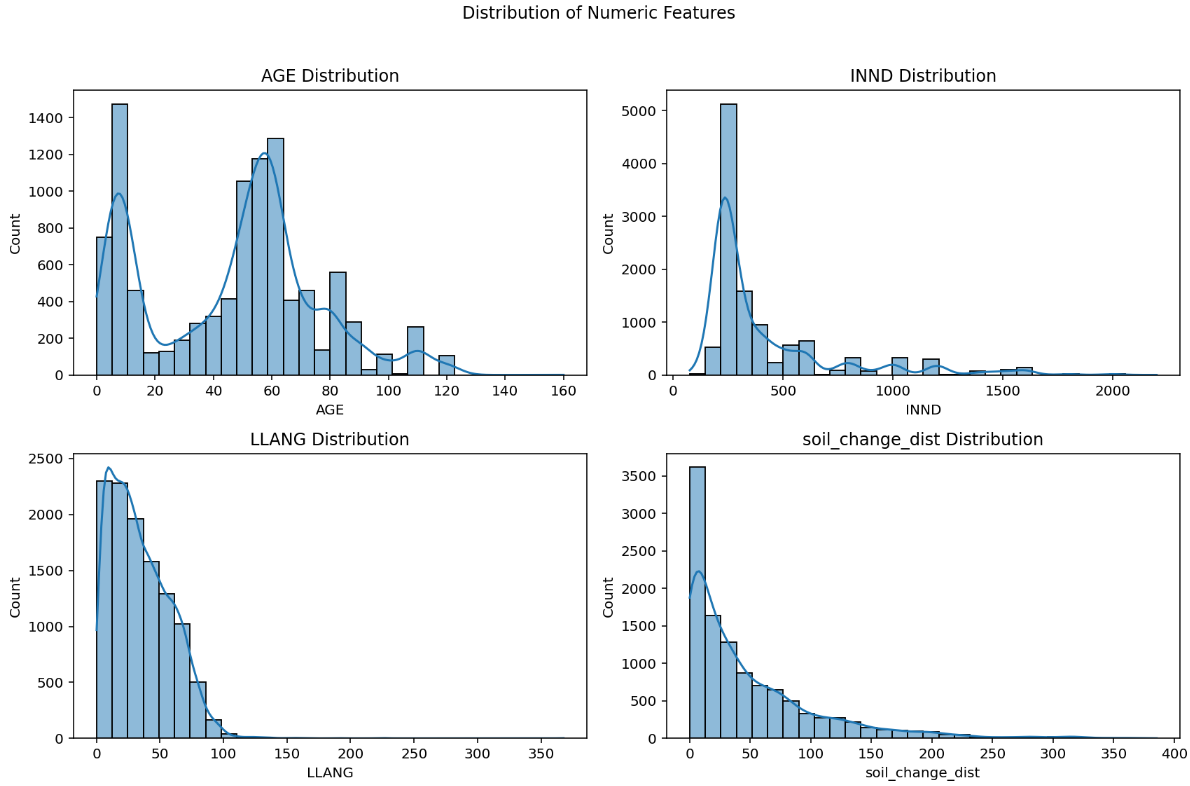


Figure 4-3 Distribution of numerical features, including Age, Diameter, Length and Soil Change Distance for the subset.

4.2.3.2. *Categorical features*

The categorical features provide insights into the structural composition of the network (for inspected pipes) and soil information: Material, Pipe types (FTYP), Soil type and Soil transition (soil_change), see Table 4-3.

Table 4-5 and Table 4-6 present the frequency (counts and percentages) for each categorical feature in the dataset.

Table 4-5 Count and percentages of pipe type and material in the inspected pipes dataset.

Pipe Attribute	Attribute Type	Count	Percentage
Pipe Network	S - Sewer	5668	50.7
	D - Stormwater	3206	28.68
	K - Combined sewage	2305	20.62
Pipe Material	BTG - Concrete	9042	80.88
	PP - Polypropylene	1447	12.94
	GJJ – Cast iron/ Grey iron	136	1.22
	PVC – Polyvinyl chloride	133	1.19
	PP_URIB – Polypropylene, structure reinforced	103	0.92
	PE - Polyethylene	96	0.86
	H – Egg-shaped concrete/ Brick	70	0.63
	SGN – Ductile iron	35	0.31

	LER - Ceramics	29	0.26
	WEHOLITE – Double-walled PE pipe	29	0.26
	GAP – Glass-fiber reinforced polyester	8	0.07
	ST - Steel	4	0.04

The pipe types (FTYP) are predominantly S (sewer)-type pipes, accounting for 50.70% of the network, followed by D (stormwater)-type (28.66%) and K (combined)-type (20.62%) pipes. The material composition shows that concrete (BTG) is the dominant material, comprising 80.88% of all inspected pipes, followed by polypropylene (PP) at 12.94%.

The model's predictive capacity was furthermore enhanced by incorporating geological data from the surrounding environment, with each pipe matched to corresponding soil information through unique ID (that also match with pipe database). This database provided not only the primary soil type in which the majority of each pipe segment is located (soil type) but also detailed information about soil transitions. Specifically, the variables "soil_change" and "soil_change_distance" capture the combination of soil types at the nearest boundary and the shortest distance from the pipe to this transition, respectively, with a value of zero indicating that the pipe intersects the soil boundary.

Given the large number of possible soil transition combinations in the dataset, and to focus the analysis on the most relevant geotechnical influences, this study only considers soil transitions that occur within 50 meters of the pipe. In the data preparation process, all soil transition types (soil_change) associated with a distance greater than 50 meters (soil_change_dist > 50) were set to 0 ("Other" in Table 4-6). This approach ensures that only nearby soil transitions, those most likely to influence the pipe's condition, are considered by the model, while transitions farther away are treated as having no meaningful effect. By applying this 50-meter cutoff, the model emphasizes local soil conditions, reduces noise from distant transitions and simplifies the feature space, therefore reducing the risk of overfitting. This targeted feature engineering step was informed by engineering judgement and preliminary data exploration and could be further refined in future research through sensitivity analysis or expert input.

Table 4-6 summarizes the distribution of the most common soil types and soil transition types (soil_change) among the inspected pipes. The dataset is dominated by a few soil types, with type 17 (postglacial clay) and 40 (glacial clay) together accounting for nearly 60% of all pipes, see subsection 2.2.1.2 for more information about challenges with clay-rich soils. Similarly, the soil transition type 40,890 (transition of glacial clay and bedrock) is the most frequent, representing 16.2% of all 1811 transitions, followed by 31,890 (postglacial sand and bedrock) and 17,890 (postglacial clay and bedrock).

Table 4-6 Count and percentages of soil type (soil) and soil transition type (the first 14 categories) in the inspected pipes dataset.

soil	Count	Percentage	soil_change	Count	Percentage
17 - postglacial clay	3376	30.2	0 (Other)	3858	34.51
40 - glacial clay	3173	28.38	40,890	1811	16.2

			glacial clay ↔ bedrock		
890 - bedrock	1793	16.04	31,890 postglacial sand↔ bedrock	987	8.83
200 - fill	1205	10.78	17,890 postglacial clay ↔ bedrock	945	8.45
31 - postglacial sand	999	8.94	17,40 postglacial clay ↔ glacial clay	468	4.19
95 - sandy moraine	244	2.18	17,200 postglacial clay ↔ fill	433	3.87
28 - postglacial fine sand	184	1.65	200,890 fill ↔ bedrock	425	3.8
33 - washed sediment, gravel	65	0.58	95,890 sandy moraine ↔ bedrock	258	2.31
50 - glaciofluvial sediment	60	0.54	40,200 glacial clay ↔ fill	254	2.27
5 - fen peat	32	0.29	31,40 postglacial sand ↔ glacial clay	237	2.12
9147 - moraine alternating with sorted sediments	24	0.21	17,91 postglacial clay ↔ water	179	1.6
1 - bog peat	9	0.08	28,890 postglacial fine sand ↔ bedrock	169	1.51
9 - overbank sediment, clay silt	6	0.05	17,31 postglacial clay ↔ postglacial sand	136	1.22
16 - gyttja clay	6	0.05	31,200 postglacial sand ↔ fill	119	1.06

4.2.3.3. Failure types and input parameters

To identify which input features are most relevant for each type of pipe failure, a Monte Carlo correlation analysis was performed for all five failure types. This approach estimates the mean correlation and confidence intervals for each feature.

The distribution of pipe failures can be summarized in Table 4-7. For each of these five failure types, the dataset was binarized such that:

- 1 indicates the presence of the specific failure type for a pipe (with grading 1-4),
- 0 indicates no failure.

INL (Infiltration) and YTS (Surface damage) are the most common failure types, with 1670 and 2472 failures, respectively. RBR (Rupture) and SPR (Cracks) are less common, with 534 and 1345 failures, respectively. DEF (Deformation) is extremely rare, with only 36 failures observed in the dataset.

Table 4-7 Binary distribution of pipe failures. 0 = No failure; 1 = Failure.

Binary distribution	INL (Infiltration)	SPR (Cracks)	RBR (Rupture)	YTS (Surface damage)	DEF (Deformation)
0 (No failure)	9509	9834	10645	8707	11143
1 (Failure)	1670	1345	534	2472	36

The relationship between input features and each failure type was investigated by using Monte Carlo correlation analysis, as this approach estimates the mean absolute correlation between each feature and the occurrence of a specific failure. Figure 4-4 illustrates a grouped bar chart presenting the mean absolute correlation between various feature categories and each of the five failure types. The chart indicates that Age is a consistently strong predictor for all failure types, with particularly high correlations observed for YTS and SPR. Material and Length also show notable correlations, especially for INL and YTS, suggesting that older, longer pipes and certain materials are more susceptible to these types of failures. For RBR, both Age and Soil Change are highly correlated, indicating that older pipes and those located in areas with varying soil conditions are more prone to structural failures such as rupture or collapse. In the case of DEF, Soil Change and Soil Type are the most influential categories, implying that environmental and soil-related factors play a significant role in the occurrence of general defects.

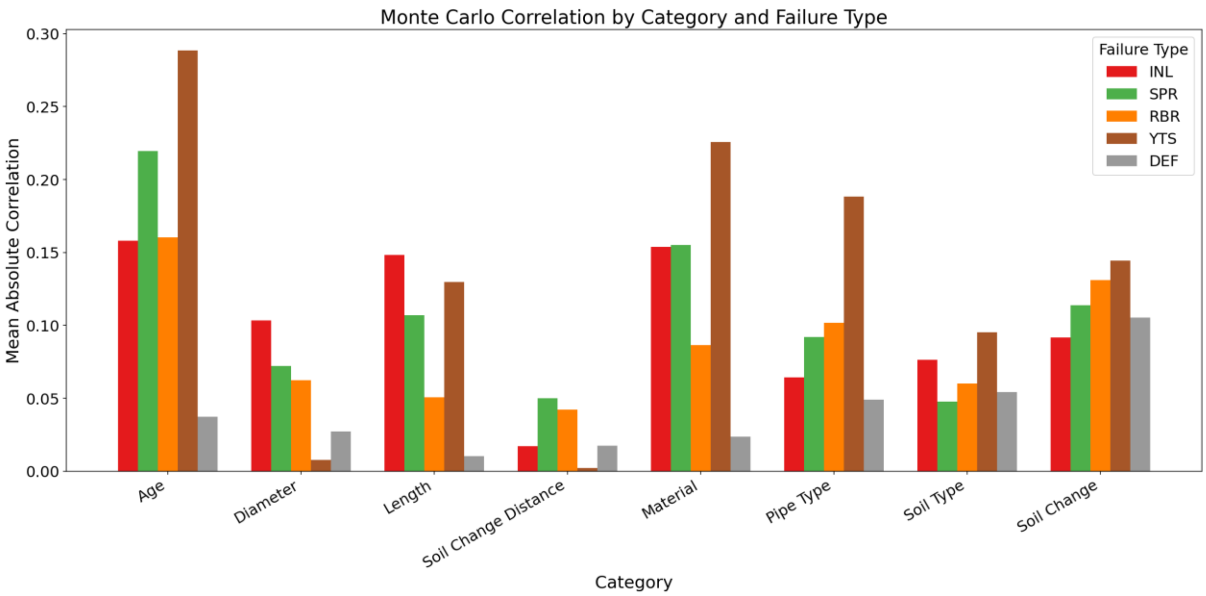


Figure 4-4 Mean absolute Monte Carlo correlation between feature categories and each failure type (INL, SPR, RBR, YTS, DEF) in the dataset. Each color represents a different failure type for direct comparison of the relative importance of each feature across failure types.

4.2.4 Analysis of network characteristics

While the model development uses only the subset of data corresponding to inspected pipes (14%), examining the characteristics of the entire pipe database remains valuable. This analysis provides context for understanding the overall network composition, age distribution and material occurrence, which helps interpret model results and assess their applicability to the uninspected pipes of the network. This view could support future inspection planning by identifying potential gaps in current inspection coverage across different pipe categories.

4.2.4.1. Material distribution and properties

Table 4-8 provides a summary of the pipe database with various materials and their associated characteristics. Material information is available for approximately 98% (78,378 pipes) of the total 80,007 pipes. The table presents key information including material type, number of pipes, construction year range, average age, average length, total length, network share, diameter range, and percentages of pipes with and without failures.

Table 4-8 Pipe database summarizing table for each material. The materials are represented by their abbreviations (in Swedish), with the full names (in English) provided below in Table 4-9 for clarity.

Material	No. of pipes	Build year	Avg. Age (year)	Avg. Pipe length (m)	Total pipe length (km)	Network share (%)	Diameter (mm)	(*) No failure (%)	(**) At least one failure (%)
BTG	66560	1795-2025	52,6	31,9	2122991,1	88,9	100-2400	94,11	5,89
G	1	N/A	-	28,6	28,6	0,0	1600,00	100,00	-
GAP	51	1905-2023	16,3	71,3	3638,0	0,2	190-1720	96,08	3,92
GJJ	917	1905-2001	61,3	17,2	15790,1	0,7	50-1200	92,69	7,31
H	357	1905-2023	82,8	38,1	13592,6	0,6	150-1600	91,60	8,40
K	4	1964-1981	55,3	7,5	30,1	0,0	50-225	100,00	-
LER	73	1905-2019	98,4	28,8	2105,3	0,1	110-450	78,08	21,92
PE	1106	1905-2024	23,6	27,1	29968,6	1,3	20-1600	98,55	1,45
PLÅT	1	1995	30,0	17,8	17,8	0,0	1200,00	100,00	-
PP	5849	1945-2025	8,8	17,3	101051,6	4,2	32-1400	99,01	0,99
PP_URIB	868	1986-2023	14,3	28,3	24588,5	1,0	90-1200	99,65	0,35
PVC	1925	1905-2024	36,7	23,3	44842,5	1,9	63-800	98,86	1,14
RFS	11	1997-2024	18,0	14,7	162,0	0,0	204-1200	100,00	-
SGN	513	1905-2024	44,9	16,2	8310,1	0,3	100-1200	98,83	1,17
ST	84	1935-2024	31,4	29,8	2505,8	0,1	90-2000	100,00	-
T	10	1915-1945	90,0	60,5	605,3	0,0	200-1200	100,00	-
WEHOLITE	48	2004-2025	7,9	32,9	1577,8	0,1	500-1200	97,92	2,08

Notes:

(*) No failure (%) presents the percentage of pipes for each material that have no failures recorded. It is important to note that this percentage is calculated for each material individually, based on the available data for that material.

(**) At least one failure (%) indicates the percentage of pipes of each material that have at least one failure recorded, based on the inspections performed using CCTV by Kretslopp och Vatten (KoV). The failures considered include in this work: Infiltration (INL), Rupture (RBR), Cracks (SPR), Surface Damage (YTS) and Deformation

(DEF). These percentages are calculated for the pipes with available inspection data and represent the proportion of pipes that showed any of these types of failures.

Table 4-9 Materials in the provided dataset, from (Kretslopp och Vatten, 2025).

Swedish Abbreviation	Material	Material Translation
BTG	Betong	Concrete
G	Galvstål	Galvanized Steel
GAP	Glasfiberarmerad Polyester	Glass-Fiber Reinforced Polyester
GJJ	Gjutjärn/Gråjärn	Cast Iron/ Grey Iron
H	Äggformad betong/ Murad tegelsten	Egg-Shaped Concrete/ Brick
K	Koppar	Copper
LER	Lergods	Ceramics
PE	Polyeten	Polyethylene
PLÅT	Plåt	Steel Plate
PP	Polypropylen	Polypropylene
PP_URIB	Polypropylen, strukturförstärkt	Polypropylene, structure reinforced
PVC	Polyvinylklorid	Polyvinyl Chloride
RFS	Rostfritt Stål	Stainless Steel
SGN	Segjärn	Ductile Iron
ST	Stål	Steel
T	Trä	Wood
WEHOLITE	PE dubbelväggigt	Double-Walled PE Pipe

4.2.4.2. Failure distribution by construction period

Figure 4-5 presents the distribution of recorded failures by material type and construction year. In Figure 4-4a), failures are significantly concentrated in concrete (BTG) pipes, particularly those installed during the mid-20th century (1940-1970). This strong material dominance suggests a significant material-specific bias in the dataset, where the predictive model used in this study may become specialized in identifying failure patterns associated with BTG pipes, potentially at the expense of generalizing to other materials.

Figure 4-4b) isolates failures among non-concrete materials, revealing a much lower and more dispersed number of failures. GJJ pipes constructed around the 1960s exhibit a higher failure count among non-concrete materials, while modern materials such as PE and PP show relatively few failures, particularly for pipes installed after 2000. This reflects a historical period bias, where failure patterns from older construction practices dominate the available data. The underrepresentation of modern materials like PE and PVC indicates that the model's ability to predict failures in newer pipes will be limited. Failure predictions will likely be skewed towards older materials and construction practices.

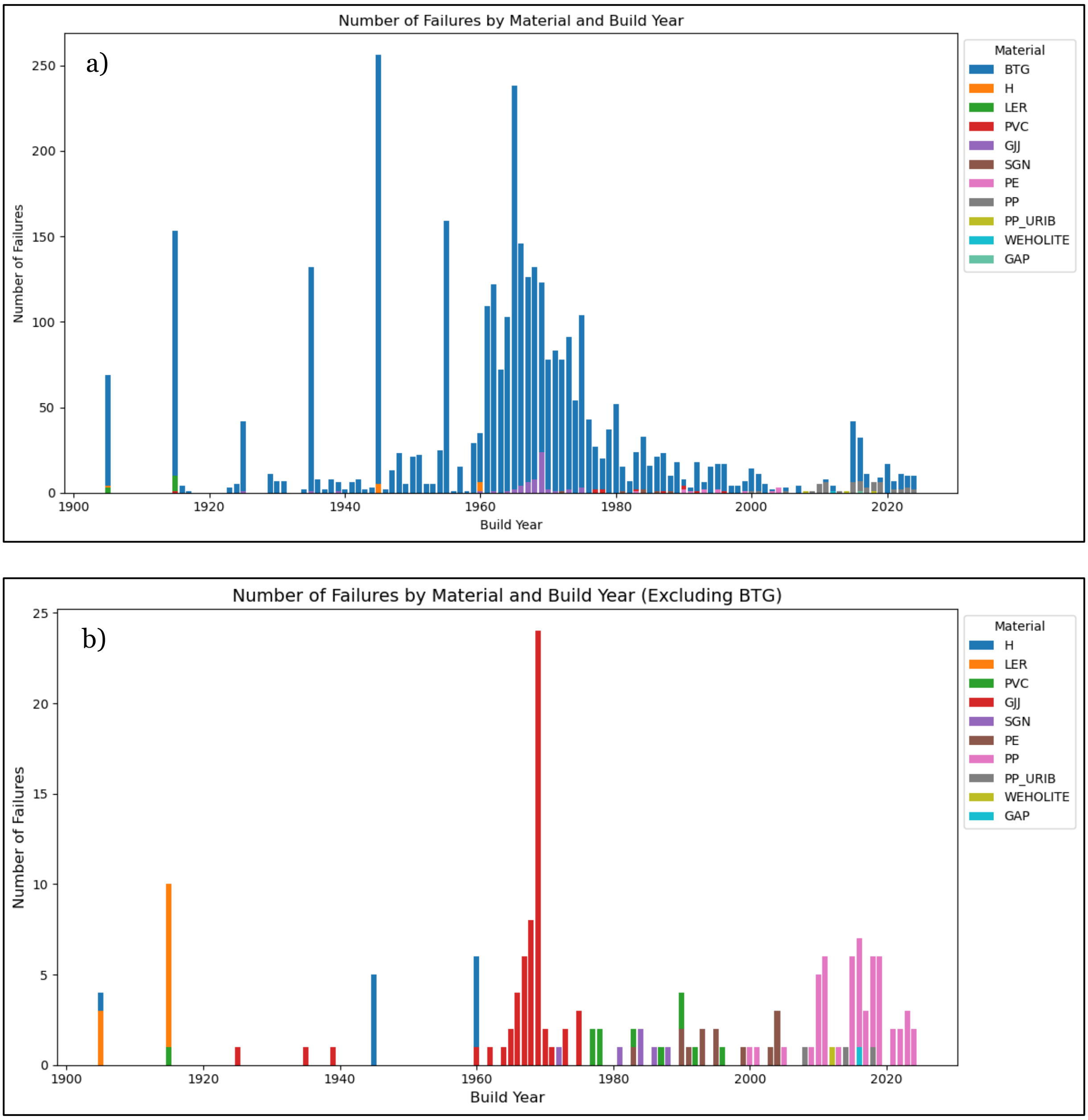


Figure 4-5 Recorded failures by material type. Figure 4-2a) presents the number of failures between Concrete (BTG) with other materials. Figure 4-2b) presents the number of failures between different materials (excluding Concrete (BTG)).

4.2.5 Soil data

A comparison of the primary soil type distribution between the inspected pipes and the entire pipe network reveals both similarities and differences, see Figure 4-6, Table 4-6 and Table 4-10 for more information. In both groups, glacial clay (label 40) and postglacial clay (label 17) are the most common soil types, together accounting for the majority of pipes. However, their relative proportions differ: postglacial clay (17) is slightly more prevalent among the inspected pipes, while glacial clay (40) is marginally more common in the full dataset. Other soil types, such as bedrock (890), silt (31) and sand (200) also appear in both groups but with varying frequencies. The differences suggest that the inspected subset is broadly representative of the overall network, but some soil types may be slightly over- or underrepresented due to the selection of pipes for inspection.

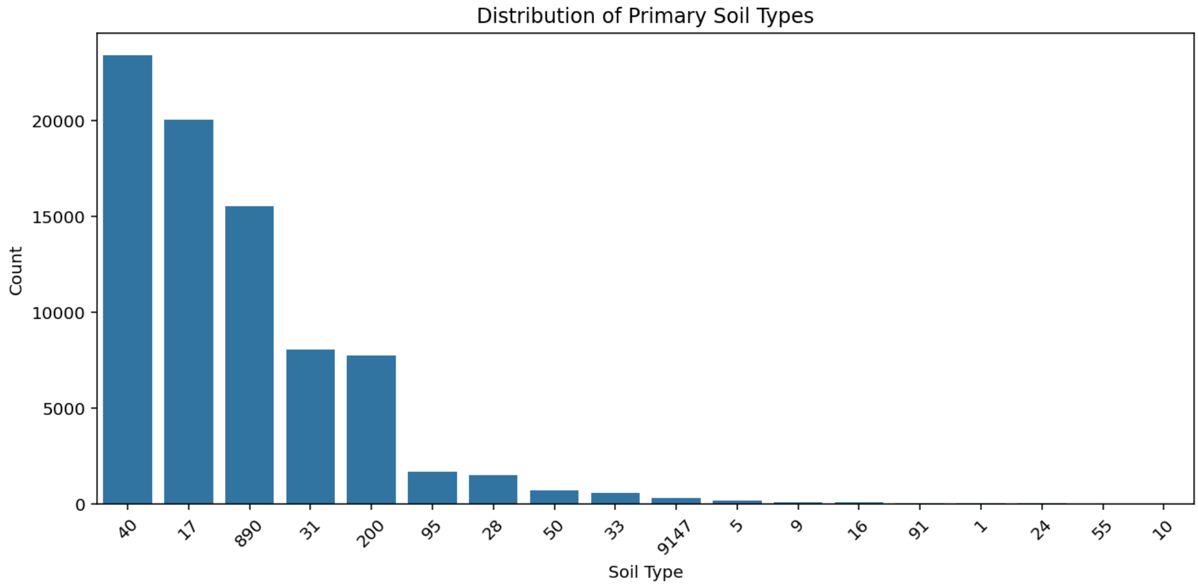


Figure 4-6 Distribution of primary soil types for the whole dataset, including uninspected pipes.

Table 4-10 Soil data in the provided dataset, from Kretslopp och Vatten (2025).

Soil type ID	Explanation (with soil type in Swedish)	Count	Percentage
40	glacial clay (glacial lera)	23375	29,22
17	postglacial clay (postglacial lera)	20031	25,04
890	bedrock (urberg)	15511	19,39
31	postglacial sand	8051	10,06
200	fill (fyllning)	7720	9,65
95	sandy moraine (sandig morän)	1683	2,10
28	postglacial fine sand (postglacial finsand)	1528	1,91
50	glaciofluvial sediment (isälvssediment)	699	0,87
33	washed sediment, gravel (svallsediment, grus)	581	0,73
9147	moraine alternating with sorted sediments (morän omväxlande med sorterade sediment)	294	0,37
5	fen peat (kärrtorv)	189	0,24
9	overbank sediment, clay silt (svämsediment, lersilt)	98	0,12
16	gyttja clay (gyttjelera)	72	0,09
91	water (vatten)	59	0,07
1	bog peat (mossetorv)	49	0,06
24	postglacial silt	35	0,04
55	glaciofluvial sediment, sand (isälvssediment, sand)	25	0,03
10	overbank sediment, sand (svämsediment, sand)	2	0,00

4.3 Data processing

4.3.1 Data preparation

This study focuses on the 11,179 pipes (14% of the total network) with CCTV inspection records for the five failures mentioned earlier in section 4.2.2. By using large datasets from CCTV inspections, model can learn from diverse defect types and therefore improving its generalization capabilities (Moradi, 2020, p. 52). Additionally, including non-inspected pipes would have required assumptions about their condition, potentially introducing bias into the model (Caradot et al., 2020, p. 290).

Missing values in numerical features were imputed using the Iterative Imputer from scikit-learn. This method models each feature with missing values as a function of other features, which provides a more robust and data-driven imputation compared to simple mean or median filling. After imputation, numerical features were scaled using MinMaxScaler to transform them to the range [0,1] to optimize compatibility with the neural network's activation functions and prevent features with larger sizes from dominating the learning process. For categorical features, missing values were filled with the string "missing". This creates a separate category for missing data and allows the model to recognize and potentially learn from the presence of missing information. Categorical variables were converted to numerical representations using target encoding, where each category is replaced by a smoothed average of the target variable for that category to prevent overfitting. This approach was selected over label encoding to avoid introducing artificial ordinal relationship between categories like pipe materials that have no inherent hierarchical order, see section 3.2.1. However, it is possible that during model application (e.g., on validation or test data), the data may contain categories that were not present in the training set. In such cases, including when a value is missing or entirely new, the model cannot compute a category-specific average. To address this, these unseen or missing categories are assigned the global mean of the target variable from the training data. This ensures that the model can still make a reasonable prediction for previously unseen categories, rather than failing or introducing bias, and helps maintain the robustness and generalizability of the model.

Construction year was converted to pipe age (Age = Current Year (2025) – Construction Year). This transformation provided a more intuitive measure directly relevant to deterioration processes. For the target variables representing condition ratings (INL, RBR, SPR, YTS and DEF), a binary transformation was applied, with 0 representing no failures and 1 representing the presence of any failure. While binarizing the condition ratings reduced complexity and helped address issues of class imbalance, it also involved a trade-off. The original multi-class labels could have provided more nuanced insights into varying degrees of deterioration, which may be valuable for long-term planning and targeted interventions. However, given the limited availability of well-distributed class labels in the dataset, a binary classification approach was more suitable for producing stable, interpretable predictions.

The same ANN architecture was used to train five separate models, each given to one failure type. This approach allows each model to learn the specific patterns and risk factors relevant to its respective failure mode, rather than forcing a single model to generalize across all types.

Future studies with more balanced and complete datasets may benefit from revisiting the multiclass output classification approach to capture the full spectrum of pipe condition severity, with these five failures combined in one final model.

An alternative data split of 70-10-20 (training-validation-testing) was employed to ensure sufficient data for model training while reserving adequate samples for unbiased validation. Using 20% of the data for testing can be useful to evaluate the model's performance on unseen data, providing an unbiased estimate of how well the model generalizes. This split also helps identify overfitting, ensuring that the model's performance is not just due to memorizing the training data. The validation set (10%) is kept separate from the training and testing data. It acts as a reference during model tuning to select hyperparameters or to monitor the model's performance during training (e.g., early stopping), which ensures that the data used for tuning does not contaminate the final unbiased test performance.

To assess variable relationships in this imbalanced dataset, Monte Carlo correlation analysis was employed. This technique generated multiple random subsamples and calculated correlation coefficients for each, producing a distribution that better reflected the underlying relationships than a single coefficient calculation (Rickman et al., 2017, p. 26).

4.3.2 Feature selection and engineering

Key factors influencing the deterioration of sewer pipelines were identified based on a thorough review of the literature, see section 2.2.1. These factors include physical characteristics such as pipe age, material, environmental aspects like soil type, and operational conditions reflected through inspection-based condition ratings. The dataset used in this study incorporates a wide range of attributes, including pipe age, material, diameter, length, sewer type, soil data, and condition ratings obtained from CCTV inspections. In this project, condition ratings are predicted using a machine learning model that leverages these parameters as input features.

4.4 ANN model implementation

The development of ANN model for predicting sewer pipe conditions involved a systematic process including data preparation, model training, testing and validation. This workflow follows general practices for ANN implementation in infrastructure modelling as discussed by Atambo et al. (2022, p. 11), Kulandaivel (2004, pp. 107-129) and Kerwin et al. (2023, p. Case study).

4.4.1 Implementation environment

The ANN model was implemented using Python 3.10 within the Spyder IDE, part of the Anaconda distribution. This environment was selected for its comprehensive data science tools and integration with essential libraries (Raschka & Mirjalili, 2019, p. Chapter 1). The implementation utilized TensorFlow 2.15 for constructing the neural network architecture, with NumPy 1.26 and Pandas 2.2 handling data preprocessing tasks. Matplotlib 3.8 provided visualization capabilities for model performance analysis and result interpretation.

Building on the processed data from section 4.3, ANN models were developed for each failure type (INL, RBR/SPR combined, YTS and DEF) to capture the unique deterioration mechanisms affecting sewer pipes.

4.4.2 Model architecture

The ANN framework used in this study was a multilayer perceptron (MLP) designed for binary classification for sewer pipe failures. The architecture consisted of an input layer corresponding to the number of selected features, including both numerical and categorical variables. This was followed by two fully connected hidden layers, where the number of neurons, activation function, regularization, and dropout rates are not fixed but were instead determined through Bayesian hyperparameter optimization using the Optuna library. For the best-performing models, the first hidden layer typically contained between 64-128 neurons and the second hidden layer between 32-64 neurons, with the exact configuration selected to maximize validation performance for each failure type.

Each hidden layer incorporated L2 regularization to prevent overfitting, batch normalization to stabilize and accelerate training, and a non-linear activation function (either LeakyReLU or ELU, as selected by the optimizer). Dropout was also applied after each hidden layer, with the dropout rate optimized during training. The output layer consisted of a single neuron with a sigmoid activation function, producing a probability estimate for pipe failure (0 = no failure, 1 = failure). Model training was performed using an optimizer (Adam, RMSprop or Nadam) selected through Bayesian hyperparameter optimization to allow the pipeline to identify the most effective optimization algorithm for each failure prediction task, see Table 4-11 for the optimal hyperparameters that selected for each failure type after BO. The result shows that the optimal architecture and training parameters varied between failure types, which reflects differences in data characteristics and model complexity requirements. Weighted binary cross-entropy loss was used to address class imbalanced by penalizing misclassification of the minority class more heavily. Early stopping and learning rate scheduling were employed to prevent overfitting and ensure efficient convergence. The network's weights were updated iteratively via the backpropagation algorithm. Further details on the model implementation and hyperparameter optimization process can be found in Section 3.4.

Table 4-11 Optimal hyperparameters for each failure type as determined by Bayesian Optimization.

Failure type	Units 1	Units 2	Dropout 1	Dropout 2	L2 factor	Learning rate	Activation	Optimizer	Batch size
INL	80	32	0.445423	0.368386	0.005251	0.000443495	elu	adam	128
SPR	96	64	0.47184	0.342757	0.005973	0.000493116	elu	adam	64
RBR	128	64	0.429195	0.335566	0.00052	2.05E-05	leakyrelu	rmsprop	32
YTS	64	48	0.334497	0.365993	0.006125	0.000305622	elu	adam	128
DEF	128	48	0.324622	0.316602	0.000162	0.000282513	leakyrelu	rmsprop	32

Figure 4-7 presents an example of basic ANN structure used in this study.

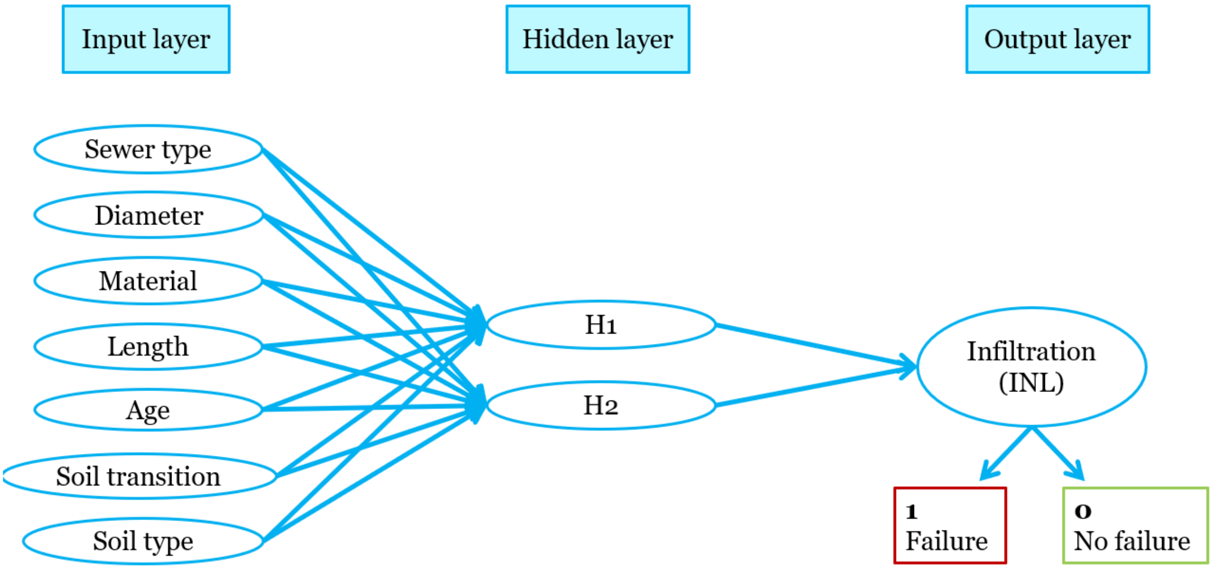


Figure 4-7 An example of ANN structure used in this study for Infiltration (INL).

4.4.3 Training process

The model training process followed a systematic pipeline, as illustrated in Figure 4-8. First, the input data was normalized to ensure all features were on a comparable scale. The neural network was then initialized with small random weights. During each training epoch, the training data was passed forward through the network to generate predictions. The binary cross-entropy loss was calculated by comparing these predictions to the true labels. The model then performed backpropagation, using the optimizer to update the network weights in order to minimize the loss. After each epoch, the model's performance was evaluated on a separate validation set. The learning rate was adaptively adjusted using a scheduler, and training continued until convergence or until early stopping criteria were met, preventing overfitting. This process ensured that the model was both robust and generalized well to unseen data.

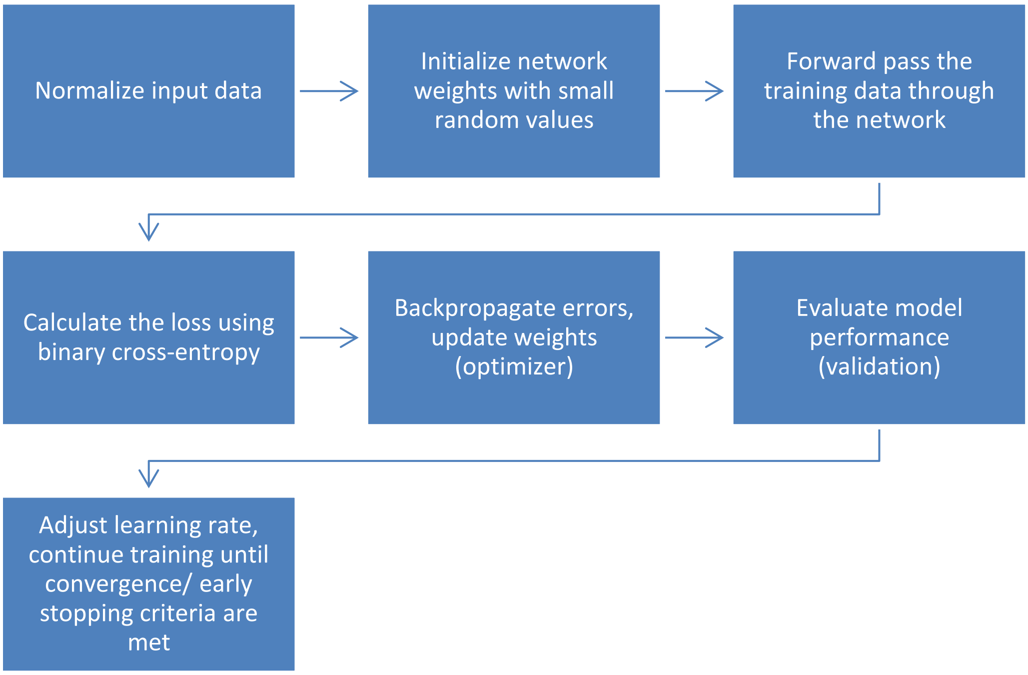


Figure 4-8 Flowchart of model training process in the study.

4.5 Model evaluation

The performance of the trained ANN models was evaluated using a set of metrics and visualization techniques. Model evaluation included the calculation of accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve (ROC AUC) on the test. To provide further insight into model performance, confusion matrices were generated and visualized with both absolute counts and normalized percentages. ROC curves were plotted to assess the models' ability to distinguish between failure and non-failure cases across different thresholds. Feature importance was analyzed by aggregating the absolute weights of the first dense layer and grouping them by feature category, allowing for interpretation of which input variables contributed most to the predictions. The model's predictions were further evaluated by grouping pipes into risk categories based on predicted failure probabilities, as described in section 3.5.

5 RESULTS AND DISCUSSION

This chapter presents the results of the artificial neural network (ANN) models developed for sewer pipe failure prediction, followed by a discussion of their implications. The performance of the models is evaluated using various metrics, and the importance of different input features is analyzed. The practical utility of the model outputs, including risk group stratification, is also discussed in the context of asset management.

The ANN architecture employed for each failure consisted of two hidden layers. The optimal number of units in each layer, along with other hyperparameters such as dropout rates, L2 regularization factor, learning rate, activation function, optimizer, and batch size, were determined through Bayesian optimization for each specific failure type. Detailed optimal hyperparameters for each model are presented in Table 4-11 in section 4.4.2.

Table 5-1 summarizes the key evaluation metrics for each failure type, including accuracy, precision, recall, specificity, F1-score, and ROC-AUC, as measured on the test set. The results demonstrate that the ANN models achieved moderate to strong predictive performance for most failure types. The YTS (Surface damage) model got the highest F1-score (0.5017), recall (0.7636) and ROC-AUC (0.7576) compared to the other models. This indicates a strong ability to identify true failures and distinguish between failed and non-failed pipes. The INL (Infiltration) and SPR (Cracks) models also showed reasonable differences, with F1-scores above 0.32 and ROC-AUC values above 0.69. In contrast, the DEF (Deformation) model achieved high accuracy (0.9821) but failed to identify any positive cases, as reflected by zero precision, recall and F1-score. These zero values indicate that the model completely failed to identify any deformation failures in the test set. This suggests that the model defaulted to predicting the majority (non-failure) class, likely due to extreme class imbalance for this failure type (36 failed pipes vs. 11143 non-failed pipes, as of historical data).

Specificity, which measures the model's ability to correctly identify non-failures, shows relatively high values across all models (ranging from 0.64 to 0.99). This indicates that the models are generally good at correctly identifying pipes that are not at risk of failure. However, the high specificity values, particularly for the DEF model (0.99), should be interpreted in conjunction with the other metrics, as they might reflect the models' tendency to predict the majority class.

Table 5-1 Models' evaluation metrics for each failure type. The best value for each metrics is highlighted as bold number.

	Model 1 (INL)	Model 2 (SPR)	Model 3 (RBR)	Model 4 (YTS)	Model 5 (DEF)
Accuracy	0.65	0.70	0.83	0.66	**0.98**
Precision	0.24	0.22	0.10	**0.37**	0.00
Recall	0.60	0.60	0.29	**0.76**	0.00
Specificity	0.66	0.72	0.86	0.64	**0.99**
F1-score	0.34	0.33	0.14	**0.50**	0.00
ROC-AUC	0.69	0.71	0.6	**0.76**	0.53

A notable trend across all models is the inconsistency between recall and precision. For instance, the YTS model achieved a high recall rate of 0.7636, but a comparatively lower precision rate of 0.3735. In contrast, the RBR model demonstrated high accuracy of 0.8309, however, this was accompanied by very low precision of 0.0931 and an F1-score of 0.1409. This pattern is indicative of models that have been trained on imbalanced datasets, where the number of failure cases is significantly smaller than the number of non-failure cases (Chen et al., 2024, p. Chapter 3). To mitigate the effects of this imbalance, class weighting was applied during model training, assigning higher importance to the minority failure class. In such scenarios, the model may correctly identify most actual failures (high recall) but also produce a higher number of false positives (low precision), as it is more likely to misclassify non-failures as failures. Achieving the right balance between precision and recall is crucial in the context of sewer pipe failure prediction. High recall ensures that most true failures are detected, which is important for minimizing the risk of missed failures that could lead to costly or hazardous incidents. However, if precision is too low, the model will generate many false alarms, potentially leading to unnecessary inspections or interventions and inefficient use of resources. The F1-score provides a single metric that balances these two aspects and makes it particularly valuable for evaluating model performance in imbalanced settings.

The ROC-AUC is a measure of the model's overall ability to discriminate between the positive (failure) and negative (non-failure) classes across all possible classification thresholds. According to the thresholds proposed by Tavakoli et al. (2019, p. 97), ROC-AUC values ranging from 0.7 and 0.8 indicate acceptable classification quality. The models for INL (0.6960), SPR (0.7122), RBR (0.6792), and YTS (0.7576) all achieved ROC-AUC values within or near this acceptable range, with the YTS model showing the strongest overall discriminative power. The DEF model's ROC-AUC of 0.5251 suggests discrimination only slightly better than random chance, suggesting that it is unable to effectively distinguish failures.

Table 5-2 presents a comparison of evaluation metrics obtained in previous studies with the range of results from this study's approach. This includes studies focusing on both drinking water networks (Giraldo-González & Rodríguez, 2020; Robles-Velasco et al., 2021; Winkler et al., 2018) and sewer networks (Goodarzi and Vazirian (2024); (Kizilöz, 2024); Malek Mohammadi et al. (2021). This study's models achieved a range of accuracy from 0.653 to 0.982, recall from 0 to 0.764, and specificity from 0.640 to 0.990 across different failure types.

When comparing to drinking water network studies, Giraldo-González and Rodríguez (2020) reported very high accuracy (0.999) and specificity (0.996), but low recall (0.392). While this study's models generally achieved lower accuracy and specificity, the best-performing model (YTS) achieved a substantially higher recall (0.764). Compared to Winkler et al. (2018), the performance range of this study's models overlaps or falls slightly below their reported ranges for accuracy and specificity, while the recall range is within their reported values. Similarly, the performance range includes accuracy and specificity values comparable to those reported by Robles-Velasco et al. (2021), although the highest recall (0.764) is slightly lower.

Turning to sewer network studies, Goodarzi and Vazirian (2024) and Malek Mohammadi et al. (2021) reported an accuracy of 0.84, while Kizilöz (2024) reported 0.99. This study's accuracy range (0.653–0.982) overlaps with the former and approaches the latter. However, due to differences in model types and the lack of reported recall and specificity in one of these studies, a detailed comparison beyond overall accuracy is not feasible.

It is important to note that direct comparisons of these metrics should be interpreted with caution. Evaluation results are highly dependent on dataset characteristics, including size, failure definitions, and data structure. Notably, the sewer network studies primarily employed regression models in general (see Table 2-5 in section 2.3), which typically report only overall accuracy and often omit classification-specific metrics such as recall and specificity metrics that are particularly important in imbalanced classification tasks like this one.

Differences in data types (e.g., discrete failure events vs. continuous time-series data) and problem definitions (e.g., binary classification vs. remaining useful life prediction) may also explain the divergence in modeling approaches between drinking water and sewer network studies. This study utilizes discrete data on pipe attributes and historical failure events, making a classification approach appropriate for predicting binary failure outcomes. In contrast, regression models are more suitable for continuous data, such as sensor readings or time-series data, which have been the focus of many previous sewer network studies.

Table 5-2 Comparison between quality metrics obtained in previous studies and this study's approach.

Pipe network	Research	Accuracy	Recall	Specificity
Drinking water	**Giraldo-González and Rodríguez (2020)**	0.999	0.392	0.996
	Winkler et al. (2018)	0.830-0.960	0.720-0.808	0.835-0.989
	Robles-Velasco et al. (2021)	0.783	0.817	0.783
Sewer	**Goodarzi and Vazirian (2024)**	0.84	0.88	0.80
	Malek Mohammadi et al. (2021)	0.84	0.57	0.93
	Kizilöz (2024)	0.99		
	This study	0.653-0.982	0.000-0.764	0.640-0.990

5.1 Confusion Matrix

To further interpret the classification performance of the models, confusion matrices were generated for each failure type. These matrices provided a detailed breakdown of four values: true positives (TP = failed pipes correctly identified), false positives (FP = non-failed pipes incorrectly predicted as failures), true negatives (TN = non-failed pipes correctly identified), and false negatives (FN = failed pipes identified as non-failed). For comparison, the confusion matrices for all failure types are presented in Appendix 2. Figure 5-1 shows the confusion matrix for the YTS model, chosen as a representative example due to its overall strong performance (as shown in Table 5-1). To facilitate comparison, Table 5-3 summarizes the key outcomes from the confusion matrices for each failure type, including recall/ sensitivity (TP-rates), and specificity (TN-rates).

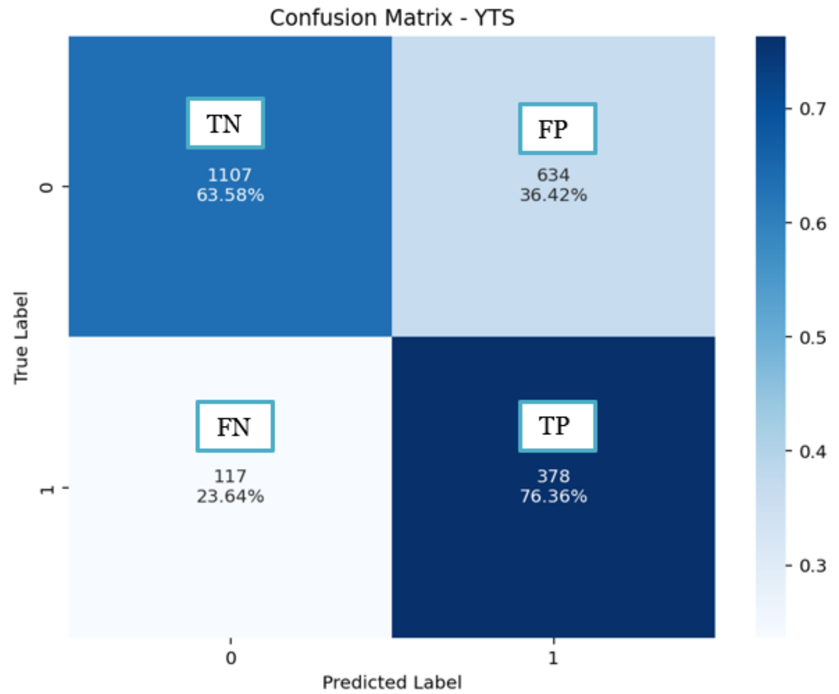


Figure 5-1 Confusion matrix for the YTS failure prediction model. Values represent the proportion of samples in each category.

Table 5-3 Summary table of key outcomes from the confusion matrices for each failure type.

Model	True Positive Rate (Recall)*	True Negative Rate (Specificity)**	Strengths	Weaknesses
1 (INL)	60%	66%	Moderate recall, some precision	Misses ~40% of failures, moderate false positives
2 (SPR)	60%	72%	Moderate recall, good precision	Misses ~40% of failures, moderate false positives
3 (RBR)	29%	86%	High precision, few false alarms	Misses most failures, low recall
4 (YTS)	76%	64%	High recall, best at finding failures	More false positives, lower precision
5 (DEF)	0%	99%	Very few false alarms	Misses most failures, no recall

*,** Calculation of recall and specificity are presented in Appendix 2.

The confusion matrices for the five-failure prediction model present differences in their ability to correctly identify both failure and non-failure cases (Table 5-3). Among all models, the YTS model demonstrates the highest recall at 76%, indicating it was the most effective at detecting actual failures. While its specificity was lower at 64%, this balance made it particularly effective for identifying pipes likely to fail. The INL and SPR models showed moderate performance, both achieving a recall of 60%. The SPR model had slightly higher specificity (72%) compared to the INL model (66%) suggesting a better ability to correctly identify non-failures while maintaining the same failure detection rate. This implies that while these models detected a reasonable proportion of true failures, they also produced a moderate number of false positives. In contrast, the RBR model got a low recall of 29%, meaning it missed a significant majority of actual failures. However, it compensated with a high specificity of 86%, resulting in fewer false alarms. The DEF models struggled the most with failure detection of 0% recall and missing all actual failures. Its very high specificity of 99% meant it produced almost no false positives. This outcome of the DEF model was likely attributable to the challenging nature of predicting this failure type and potential issues related to extreme class imbalance, where the model defaulted to predicting the majority non-failure class and overlooks rare events.

Overall, the Table 5-3 highlights the trade-off between recall and specificity in classification models, especially in the context of imbalanced datasets like sewer pipe failures. Models with high recall (such as YTS) are better at detecting failures but may generate more false positives (FP/ false alarm), which can lead to unnecessary maintenance actions. Conversely, models with high specificity (such as RBR and DEF) minimize false alarms but risk missing critical failures (false negatives, FN), which could have serious consequences for infrastructure management. This trade-off between FP and FN is typical in imbalanced classification problems. In the context of sewer pipe maintenance, missing a true failure (FN) can be more costly than investigating a false alarm (FP). The choice of model should therefore be guided by the operational priorities of the utility, whether it is more important to avoid missed failures or to minimize unnecessary interventions.

5.2 Training vs validation loss

Figure 5-2 presents the training and validation loss and accuracy curves for the YTS model, while Figure 5-3 shows the corresponding curves for the DEF model. These figures illustrate the distinct learning processes observed for the worst-performing (DEF) and best-performing (YTS) models, serving as representative examples of the range of learning behaviors observed across all failure types. The training and validation plots for the remaining failure types (INL, SPR, RBR) are provided in Appendix 2.

For the YTS model, both training and validation loss decreased significantly during the initial epochs, reflecting rapid convergence and efficient learning. The validation loss then plateaued and remained relatively stable, closely following the training loss before a slight divergence. This stabilization of validation loss with the plateauing validation accuracy suggests that the model converged and generalized reasonably well to unseen data without significant overfitting, which is indicative of effective training and the appropriate application of regularization and early stopping. The observation that validation accuracy is higher than training accuracy is a positive sign in this context. This suggests that the model's performance on unseen data is robust and that the regularization strategies are effectively preventing overfitting while maintaining good generalization capabilities.

In contrast, the DEF model has a significantly different learning curve. The validation loss decreased very rapidly and stabilized at a very low value, while the training loss fluctuated at a much higher level. Correspondingly, both training and validation accuracy quickly reached and remained at a very high level, nearly 100%. This unusual behavior, particularly the large gap between training and validation loss and the consistently high validation accuracy despite zero recall (as seen in the performance metrics), suggests that the DEF model primarily learned to predict the majority class (non-failures) on the validation set. While this resulted in minimal validation loss and high validation accuracy for non-failures, it indicates a failure to effectively learn or detect the minority (failure) class, highlighting the extreme challenge posed by the imbalance and nature of the DEF failure type rather than effective learning of the overall prediction task.

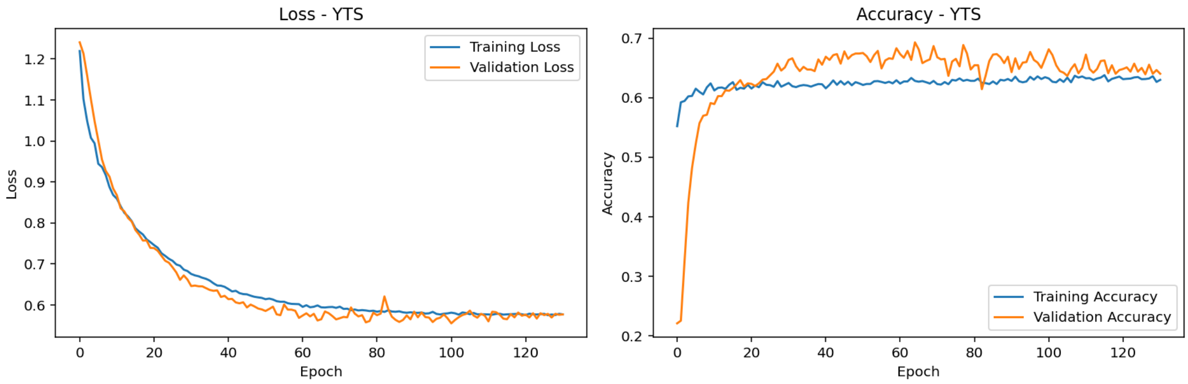


Figure 5-2 Training and validation loss and accuracy curves for the YTS model.

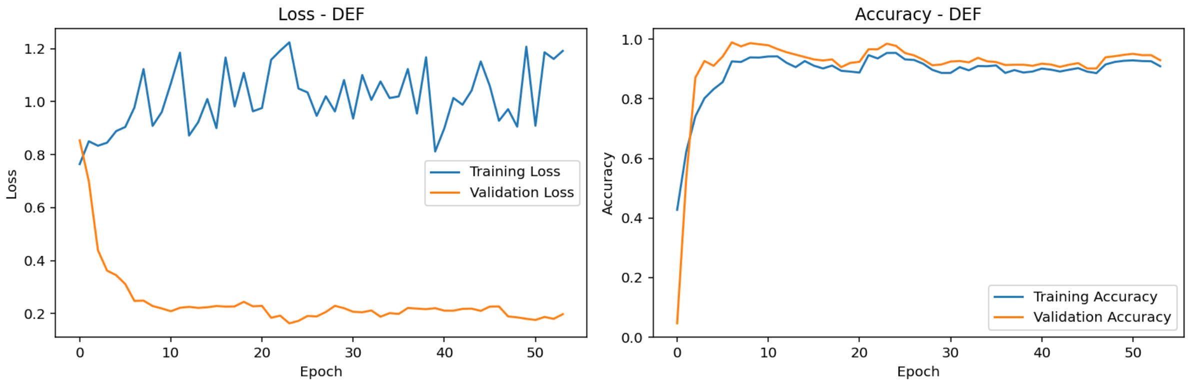


Figure 5-3 Training and validation loss and accuracy curves for the DEF model.

5.3 Feature importance

Figure 5-4 presents the normalized feature importance by category for each of the five failure types, derived from the weights of the ANN models. This analysis highlights which input features were most influential in the models' predictions.

Overall, the feature importance varies significantly across the different failure types that reflects distinct predictive patterns for each failure type. Among the most consistently important feature categories are Age, Soil Change and Material Type. These features are fundamentally linked to the physical degradation processes and external stresses pipes experience over time and due to environmental interactions, making them intuitively influential factors for various failure mechanisms. Age demonstrates particularly high importance for predicting INL and SPR. Soil Change is most influential for SPR and RBR, which aligns with the understanding that changes in soil conditions can significantly impact pipe structural integrity. Material Type is a strong predictor for SPR and YTS, which is consistent with domain knowledge as different pipe materials have varying susceptibilities to cracking and surface degradation. Other features like Diameter, Length and Pipe Type also show considerable importance for specific failure types, such as Diameter for SPR and Length and Pipe Type for YTS.

The importance scores observed for features like Age, Soil Change and Material Type are largely consistent with existing domain knowledge and findings from the previous studies on pipe failure prediction, which often identify these as key factors influencing deterioration and failure risks (see section 2.2.1). Conversely, the features have relatively low importance for the DEF model, which aligns with its poorer predictive performance metrics observed earlier (Table 5-1), which suggests that the current set of features may be less effective in modeling this specific failure type compared to others. These findings highlight the importance of targeted data collection and dedicated feature engineering efforts to both increase understanding of the model's working and improve its predictive power. This is particularly important for features identified as less dominant, as their influence may become critical under certain specific conditions.

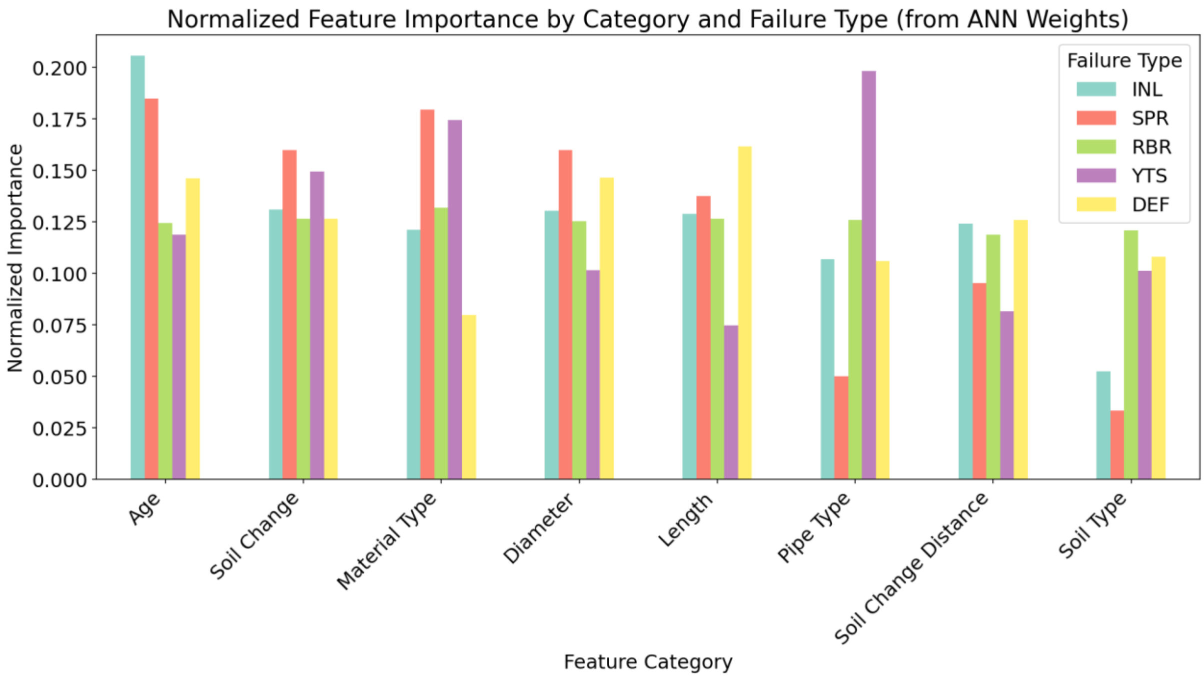


Figure 5-4 Normalized feature importance by category and failure type, derived from the weights of the trained ANN models. The y-axis represents the normalized importance score, while the x-axis shows different feature categories. The colored bars indicate the importance of each feature category for predicting each specific failure type (INL-Infiltration, SPR-Cracks, RBR-Rupture, YTS-Surface damage, DEF-Deformation).

An attempt to combine "soil_change" (categorical) and "soil_change_distance" (numerical) into a single "soil_cross" feature, encoded using one-hot encoding, was explored. However, this approach resulted in high input dimensionality and complicated the interpretation of feature importance. After evaluating the outcomes of these trials and through discussions

with supervisors, the “soil_cross” feature was ultimately discarded.
Instead, “soil_change” and “soil_change_distance” were retained as separate features. The “soil_change” value was set to 0 for all instances, where the “soil_change_distance” is greater than 50 meters. This method ensured that only nearby soil transitions, regarded most likely to affect pipe condition, were fed into the model, while distant transitions were effectively excluded from influencing predictions. By focusing on local soil changes, this 50-meter threshold helped to minimize noise, update the feature representation and reduce the likelihood of model overfitting. This feature engineering step was guided by engineering judgement and preliminary data analysis and could be further investigated in future work. The impact of these feature engineering decisions, including a comparison of feature importance with and without the “soil_cross” and renovation data, is presented in Appendix 2 (Figure 2L and 2M).

5.4 Failure frequency per prediction group

In line with recommendation from the KoV utility, risk group stratification was employed as an intuitive and practical approach to interpret the ANN model predictions for asset management. The idea behind this method is to move beyond simply looking at raw predicted failure probabilities, which can be difficult to interpret directly across a large network. Instead, the predicted probabilities for each pipe segment are used to assign that pipe to one of several discrete risk groups. This typically involves defining thresholds that divide the range of predicted probabilities into a fixed number of groups, five in this case, see Table 5-4.

For each group, the observed failure frequency is calculated. As highlighted by KoV utility, this calculation standardizes the number of observed failures by a common denominator - the total pipe length within that prediction group. This results in the metric “number of observed failures per kilometer of pipe per prediction group” (failures/km). This standardization is crucial because individual pipe segments in the dataset can have varying lengths, and reporting failures per unit length makes the failure rates within each group directly comparable.

Table 5-4 Five prediction groups and their corresponding risk levels, interpretation based on predicted probability and suggested actions for a utility.

Prediction group	Risk level	Interpretation	Suggested action for utility
1	Very low risk	Very low predicted probability of failure	Routine monitoring, defer maintenance and focus resources elsewhere
2	Low risk	Low predicted probability of failure	Less frequent inspection cycles, consider proactive maintenance
3	Medium risk	Moderate predicted probability of failure	Prioritized for proactive maintenance or detailed inspections

4	High risk	High predicted probability of failure	Prioritized for detailed inspection, targeted rehabilitation/replacement
5	Very high risk	Very high predicted probability of failure	Urgent inspection, major rehabilitation and expedited replacement

Figure 5-5 presents the failure frequency per prediction group for the YTS model on a test set covering approximately 77.28 km of pipes. The plot demonstrates a strong positive correlation between the assigned risk group and the observed failure frequency. As the prediction group increases from 1 to 5, the failure frequency steadily rises, with prediction group 5 showing a significantly higher failure rate (9.87 failures/km over 20 km of pipe in this group) compared to the lowest risk group (group 1 with 0.78 failures/km over 11 km of pipe in this group). This increasing trend validates the model's ability to stratify pipes according to their actual historical failure risk for YTS, within this test set, and provides an actionable output for prioritizing inspection and maintenance efforts on the pipe segments categorized into higher risk groups.

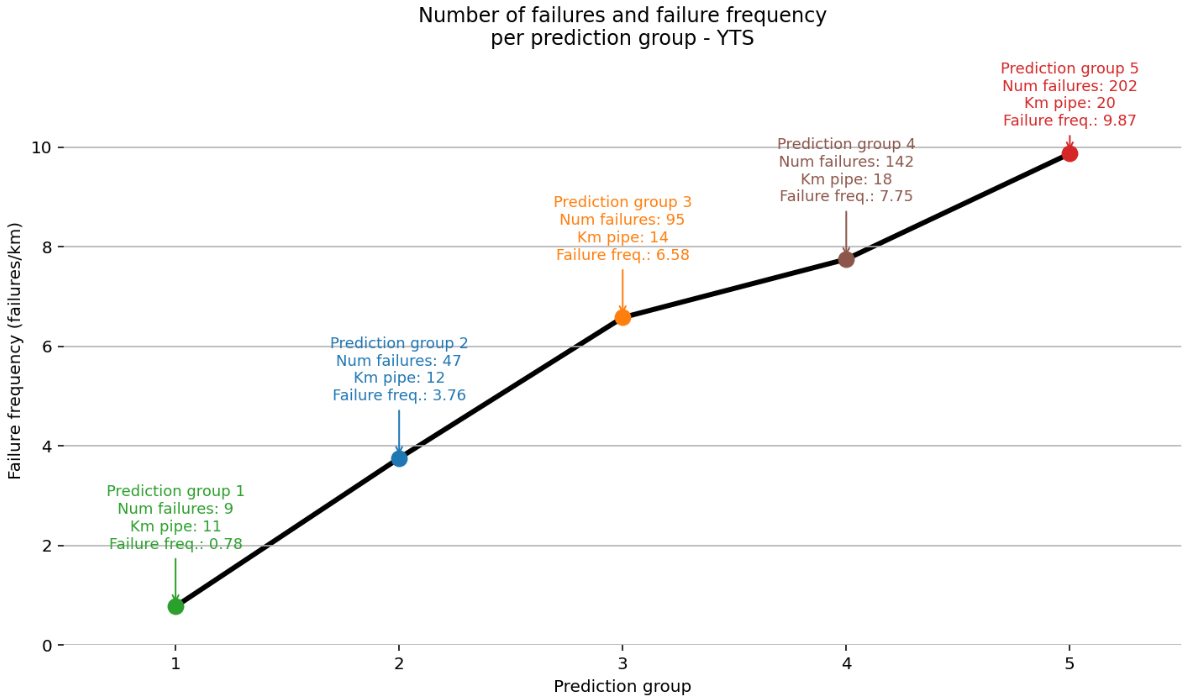


Figure 5-5 Failure frequency per prediction group for the YTS model on the test set.

Similar analyses were also conducted for the other failure types on their respective test sets (INL: 77.53 km, SPR: 77.84 km, RBR: 77.50 km, DEF: 80.62 km), with their figures presented in Appendix 2. For both the SPR and RBR models, the risk stratification demonstrates a positive correlation, with observed failure frequency increasing significantly from the lower to the highest risk groups, which indicates effective differentiation of failure risk across the test sets. The INL model also shows an increasing trend in failure frequency across lower and medium risk groups, though the trend plateaus at the highest risk levels, suggesting less distinct stratification compared to YTS, SPR and RBR. In contrast, the DEF model does not have a clear or consistent increasing trend across the prediction groups as it reflects its difficulty in effectively stratifying pipes by deformation risk, which is consistent with its overall low predictive performance.

6 CONCLUSIONS

This degree project aimed to develop and evaluate Artificial Neural Network (ANN) models for predicting different types of sewer pipe failures using available inspection and static data from Gothenburg municipality – Kretslopp och Vatten. Five separate binary classification models were developed, each targeting a specific failure type: infiltration (INL), cracks (SPR), rupture (RBR), surface damage (YTS), and deformation (DEF). The study involved comprehensive data preprocessing, feature engineering, model training, and evaluation using standard classification metrics (accuracy, precision, specificity, recall, F1-score, ROC-AUC), as well as feature importance analysis and risk group stratification.

The following research questions were addressed:

RQ1: What are the main challenges and limitations when implementing ANN models for predictive maintenance in sewer networks, and how can these be addressed through improved data collection or modeling techniques?

The primary challenges included limited inspection coverage (only 14% of the whole pipe network), data quality issues (e.g., missing values, inconsistent formats), and severe class imbalance, particularly for rare failure types like deformation (DEF). For instance, the DEF dataset had only 36 failed pipes compared to 11,143 non-failed pipes which highlighting the extreme skew. These limitations affected model generalizability and performance. To mitigate these issues, the study applied advanced preprocessing techniques (e.g., iterative imputation, normalization, target encoding), class weighting during training, and Bayesian hyperparameter optimization. Despite these efforts, the findings highlight the need for more representative and balanced datasets, especially for underreported failure types, to improve model robustness and transferability.

RQ2: How does ANN model performance vary across different sewer pipe failure types, and what role do input features and data imbalance play in this variation?

Model performance varied significantly across failure types. The ANN model for surface damage (YTS) achieved the best results, with an F1-score of 0.50, recall of 0.76 and ROC-AUC of 0.76, indicating strong predictive capability. Models for infiltration (INL) and cracks (SPR) also performed moderately well with F1-scores above 0.32 and ROC-AUC values above 0.69. In contrast, the model for deformation (DEF) failed to identify any positive cases, as reflected by zero precision, recall and F1-score. This suggested that the model defaulted to predicting the majority (non-failure), which directly attributable to the extreme class imbalance mentioned earlier (e.g. 36 DEF failures vs. 11,143 non-failures).

The varying influence of input features across failure types also contributed to the performance variation. Feature importance analysis revealed that pipe age, material, and soil transition were consistently influential predictors for more predictable failure types like YTS, INL, and SPR, and these features had lower importance for the DEF model. This suggests that the current set of input features available in the current dataset for such types underscores why performance varied so drastically. Future work may benefit from incorporating different types of data or features that are more specifically correlated with mechanisms behind failures like deformation.

7 SUGGESTIONS FOR FURTHER WORK

To further enhance the predictive performance and practical utility of sewer pipe failure models, the following recommendations for future work are suggested:

- **Data collection and management**

To mitigate the impact of data limitations and potential bias, future data collection efforts should be strategically directed. This involves proactively targeting the inspection and recording of pipe segments and failure types currently underrepresented in the dataset, with particular emphasis on rare failure mechanisms such as Deformation (DEF). Implementing stratified or targeted sampling strategies during field inspections can help ensure a more balanced representation of all failure classes across the network, providing the models with sufficient examples to learn the distinct patterns associated with less common events. Collaborative efforts with operational teams, such as those conducting routine inspections or maintenance, could also facilitate prioritizing data collection in areas or on pipe types where historical failure data is sparse, thereby enriching the dataset with more diverse and representative information.

Beyond increasing data volume, maintaining and improving data quality is crucial. This includes implementing regular data audits and validation checks to systematically identify and correct inconsistencies, missing values, and potential errors within the inspection records and static pipe attributes.

Furthermore, expanding the feature set available for modeling is recommended. Incorporating additional relevant data, such as chemical properties like pH, soil corrosivity, or other environmental factors, alongside collecting more detailed information on failure cases themselves, could significantly improve model input diversity and enhance predictive power by capturing more complex influences on pipe deterioration.

- **Model development and evaluation**

In terms of modeling techniques, future work could explore alternative or advanced approaches. Experimenting with multi-output (or multi-label) classification models could provide a more comprehensive view by predicting multiple failure types or condition grades simultaneously. This approach may better reflect real-world scenarios where a single pipe segment might exhibit several issues concurrently, allowing for a more integrated assessment of its overall condition.

Further refinement of the model training process is also recommended. While Bayesian optimization was utilized in this study, exploring a wider range of hyperparameters for the chosen ANN architecture (including learning rate schedules, various batch sizes, alternative network configurations, and different regularization techniques) using complementary methods such as grid search or random search could potentially lead to further optimization of model performance.

A critical area for future investigation is the generalizability of the developed models. Evaluating model performance on uninspected pipes within the same network or on data from different geographic regions is essential to assess how well the models transfer to unseen data and diverse environments. If significant performance degradation is observed, exploring techniques such as transfer learning or domain adaptation could be beneficial. These methods allow models trained on one dataset (the inspected pipes) to be adapted and potentially perform better when applied to new datasets (uninspected pipes or different cities) with potentially different data distributions or characteristics.

Beyond exploring variations within the ANN framework, it would be valuable to investigate the use of other machine learning models to compare their performance and interpretability. Evaluating models such as Random Forests, Support Vector Machines (SVMs), or various ensemble methods could reveal alternative approaches that offer competitive predictive accuracy or provide different insights into feature relationships compared to ANN models.

Finally, to ensure the long-term relevance and effectiveness of predictive maintenance models in practice, it is crucial to develop a continuous model updating pipeline. This involves establishing a systematic process where newly acquired inspection and failure data are regularly incorporated into the dataset used for retraining and refining the models. Implementing such a feedback loop supports adaptive learning, allowing the models to evolve with the changing conditions of the network and ensuring their predictions remain accurate and relevant as new information becomes available over time.

REFERENCES

Achim, D., Ghotb, F., & McManus, K. (2007). Prediction of water pipe asset life using neural networks. *Journal of infrastructure systems*, *13*(1), 26-30.

Agahi, M., & Kim, D. (2021). Rank and Linear Correlation Differences in Monte Carlo Simulation. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, *7*, 04020058. https://doi.org/10.1061/AJRUA6.0001115

Ahn, J., Lee, S., Lee, G., & Koo, J. (2005). Predicting water pipe breaks using neural network. *Water Science and Technology: Water Supply*, *5*(3-4), 159-172.

Ahrary, A., Nassiraei, A., & Ishikawa, M. (2007). A study of an autonomous mobile robot for a sewer inspection system. *Artificial Life and Robotics*, *11*, 23-27. https://doi.org/10.1007/s10015-006-0392-x

Al-Barqawi, H., & Zayed, T. (2006). Condition Rating Model for Underground Infrastructure Sustainable Water Mains. *Journal of Performance of Constructed Facilities*, *20*(2), 126-135. https://doi.org/doi:10.1061/(ASCE)0887-3828(2006)20:2(126)

Alda-Vidal, C., Browne, A. L., & Hoolohan, C. (2020). “Unflushables”: Establishing a global agenda for action on everyday practices associated with sewer blockages, water quality, and plastic pollution. *WIREs Water*, *7*(4), e1452. https://doi.org/https://doi.org/10.1002/wat2.1452

Allweyer, T. (2016). *BPMN 2.0: introduction to the standard for business process modeling*. BoD–Books on Demand.

Alsaqqar, A. S., Khudair, B. H., & Jbbar, R. K. (2017). Rigid trunk sewer deterioration prediction models using multiple discriminant and neural network models in Baghdad City, Iraq. *Journal of Engineering*, *23*(8), 70-83.

Amaitik, N. M., & Amaitik, S. M. (2010). Prediction of pipe failures in water mains using artificial neural network models. Proceedings of the 11th International Arab Conference of information Technology (ACIT’2010), University of Garyounis, Benghazi, Libya,

Ana, E., W., B., M., P., C., T., S., S., I., B., & and De Gueldre, G. (2009). An investigation of the factors influencing sewer structural deterioration. *Urban Water Journal*, *6*(4), 303-312. https://doi.org/10.1080/15730620902810902

Anwar, A., Liu, X., & Zhang, L. (2022). Biogenic corrosion of cementitious composite in wastewater sewerage system–A review. *Process Safety and Environmental Protection*, *165*, 545-585. https://doi.org/https://doi.org/10.1016/j.psep.2022.07.030

Asnaashari, A., McBean, E. A., Gharabaghi, B., & Tutt, D. (2013). Forecasting watermain failure using artificial neural network modelling. *Canadian Water Resources Journal*, *38*(1), 24-33.

Atambo, D., Najafi, M., & Kaushal, V. (2022). Development and Comparison of Prediction Models for Sanitary Sewer Pipes Condition Assessment Using Multinomial Logistic Regression and Artificial Neural Network. *Sustainability*, *14*, 5549. https://doi.org/10.3390/su14095549

Belcic, I., & Stryker, C. (2024). *What is hyperparameter tuning?* IBM. Retrieved 04-19 from https://www.ibm.com/think/topics/hyperparameter-tuning

Bergstra, J., Bardenet, R., Kégl, B., & Bengio, Y. (2011). *Algorithms for Hyper-Parameter Optimization*.

Bishara, A. J., & Hittner, J. B. (2015). Reducing Bias and Error in the Correlation Coefficient Due to Nonnormality. *Educ Psychol Meas*, *75*(5), 785-804. https://doi.org/10.1177/0013164414557639

Blomkvist, P., Karpouzoglou, T., Nilsson, D., & Wallin, J. (2023). Entrepreneurship and alignment work in the Swedish water and sanitation sector. *Technology in Society*, *74*, 102280. https://doi.org/https://doi.org/10.1016/j.techsoc.2023.102280

Bui, D., & Seidu, R. (2022). Comparison of Machine Learning Techniques for Condition Assessment of Sewer Network. *IEEE Access*, *100*, 124238-124258. https://doi.org/10.1109/ACCESS.2022.3222823

Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, *36*(3, Part 1), 4626-4636. https://doi.org/https://doi.org/10.1016/j.eswa.2008.05.027

Caradot, N., Mathias, R., Pascale, R., Antoine, C., Nic, L., Elke, E., Alexander, R., François, C., & and Cherqui, F. (2020). The influence of condition assessment uncertainties on sewer deterioration modelling. *Structure and Infrastructure Engineering*, *16*(2), 287-296. https://doi.org/10.1080/15732479.2019.1653938

Caradot, N., Sampaio, P. R., Guilbert, A. S., Sonnenberg, H., Parez, V., & Dimova, V. (2021). Using deterioration modelling to simulate sewer rehabilitation strategy with low data availability. *Water Science and Technology*, *83*(3). https://doi.org/10.2166/wst.2020.604

Charniak, E. (1985). *Introduction to Artificial Intelligence*. Pearson Education. https://books.google.se/books?id=HACaS635bYcC

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321-357. https://doi.org/10.1613/jair.953

Chen, W., Yang, K., Yu, Z., Shi, Y., & Chen, C. L. P. (2024). A survey on imbalanced learning: latest research, applications and future directions. *Artificial Intelligence Review*, *57*(6), 137. https://doi.org/10.1007/s10462-024-10759-6

Cheng, J. C. P., & Wang, M. (2018). Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques. *Automation in Construction*, *95*. https://doi.org/10.1016/j.autcon.2018.08.006

Chughtai, F., & Zayed, T. (2007). Sewer Pipeline Operational Condition Prediction Using Multiple Regression. https://doi.org/10.1061/40934(252)18

Clementson, I., Alenius, E. and Gustafsson, L.-G. (2020). *Tillskottsvatten i avloppssystemet - nya tankar om nyckeltal*. Svenskt Vatten

Clementson, I., & Charlesworth, S. (2025). *Machine learning as a method of reducing leakage at Örbyfältet. Implementations of new technologies within water production*. Svenskt Vatten AB

Curran, P. A. (2014). Monte Carlo error analyses of Spearman's rank test. *arXiv preprint arXiv:1411.3816*.

Davies, J. P., Clarke, B. A., Whiter, J. T., & Cunningham, R. J. (2001). Factors influencing the structural deterioration and collapse of rigid sewer pipes. *Urban Water*, *3*(1), 73-89. https://doi.org/https://doi.org/10.1016/S1462-0758(01)00017-6

Denha, R. F. (2023). *Theoretical Analysis of Sewer Pipe Inspections andOptimization of CCTV Inspection for EffectiveMaintenance: Moving from Reactive to Proactive : A Guide Towards Selecting and Implementing Advanced Technologies forEnhanced Sewage Infrastructure Management* (Publication Number 23178) [Student thesis, DiVA. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-330511

Denison, I. A., & Ewing, S. P. (1935). CORROSIVENESS OF CERTAIN OHIO SOILS. *Soil Science*, *40*(4). https://journals.lww.com/soilsci/fulltext/1935/10000/corrosiveness_of_certain_ohio_soils.3.aspx

Developers Google. (2025). *Datasets: Dividing the original dataset*. Google Machine Learning. Retrieved 05-02 from https://developers.google.com/machine-learning/crash-course/overfitting/dividing-datasets

Elmasry, M., Al Hawari, A., & Zayed, T. (2017). Cost benefit analysis for failure of sewer pipelines. *MATEC Web of Conferences*, *120*, 08006. https://doi.org/10.1051/matecconf/201712008006

Enfinger, K., & Stevens, P. (2007). *Scattergraph Principles and Practice: Characterization of Sanitary Sewer and Combined Sewer Overflows*. https://doi.org/10.1061/40934(252)75

EPA. (2009). *Condition Assessment of Wastewater Collection Systems*. Retrieved from https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P1003ZQY.txt

EPA. (2022). *Wastewater treatment in Sweden 2020*. Stockholm Retrieved from https://www.naturvardsverket.se/publikationer/8800/978-91-620-8896-5/

F. Alqahtani, A. A., Ghalib M. Hommadi, M. A. Alammari. (2023). Assessment Framework for the Maintainability of Sewer Pipeline Systems. *Applied Sciences*, *13*(21). https://doi.org/10.3390/app132111828

Fernandez, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *Journal of Artificial Intelligence Research*, *61*, 863-905. https://doi.org/10.1613/jair.1.11192

Geem, Z. W., Tseng, C.-L., Kim, J., & Bae, C. (2007). Trenchless water pipe condition assessment using artificial neural network. In *Pipelines 2007: Advances and experiences with trenchless pipeline projects* (pp. 1-9).

Gerlin, V., Priya, Y. G., k, R. S., Kaladevi, R., Shanmugasundaram, H., & Bhanuprasad, A. (2023, 23-25 Jan. 2023). Sewage Monitoring System Using IoT. 2023 International Conference on Computer Communication and Informatics (ICCCI),

Ghavami, S. M., Borzooei, Z., & Maleki, J. (2020). An effective approach for assessing risk of failure in urban sewer pipelines using a combination of GIS and AHP-DEA. *Process Safety and Environmental Protection*, *133*, 275-285. https://doi.org/https://doi.org/10.1016/j.psep.2019.10.036

Giraldo-González, M. M., & Rodríguez, J. P. (2020). Comparison of statistical and machine learning models for pipe failure modeling in water distribution networks. *Water*, *12*(4), 1153.

Goodarzi, M. R., & Vazirian, M. (2024). A machine learning approach for predicting and localizing the failure and damage point in sewer networks due to pipe properties. *Journal of Water and Health*, *22*(3), 487-509. https://doi.org/10.2166/wh.2024.249

Grigg, N. S. (2010). Water, Wastewater, and Stormwater Infrastructure Management | Neil S. https://doi.org/10.1201/9781420032338

Haghighatafshar, S., Nordlöf, B., Roldin, M., Gustafsson, L. G., & Jonsson, K. (2018). Efficiency of blue-green stormwater retrofits for flood mitigation - Conclusions drawn from a case study in Malmö, Sweden. *Journal of Environmental Management*, *207*, 60-69. https://doi.org/10.1016/j.jenvman.2017.11.018

Hahn, M. A., Palmer, R. N., Merrill, M. S., & Lukas, A. B. (2002). Expert system for prioritizing the inspection of sewers: Knowledge base formulation and evaluation [Article]. *Journal of Water Resources Planning and Management*, *128*(2), 121-129. https://doi.org/10.1061/(ASCE)0733-9496(2002)128:2(121)

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, *73*, 220-239. https://doi.org/https://doi.org/10.1016/j.eswa.2016.12.035

Hawari, A., Alkadour, F., Elmasry, M., & Zayed, T. (2017). Simulation-based condition assessment model for sewer pipelines. *Journal of Performance of Constructed Facilities*, *31*(1), 04016066.

Hawari, A., Alkadour, F., Elmasry, M., & Zayed, T. (2020). A state of the art review on condition assessment models developed for sewer pipelines. *Engineering Applications of Artificial Intelligence*, *93*. https://doi.org/10.1016/j.engappai.2020.103721

He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263-1284. https://doi.org/10.1109/TKDE.2008.239

Huang, S.-C., & Le, T.-H. (2021). Chapter 2 - Neural networks. In S.-C. Huang & T.-H. Le (Eds.), *Principles and Labs for Deep Learning* (pp. 27-55). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-323-90198-7.00006-9

Huynh, N. (2023, 05-04). Understanding Loss Functions for Classification. https://medium.com/@nghihuynh_37300/understanding-loss-functions-for-classification-81c19ee72c2a

Isik, F., Ozden, G., & Kuntalp, M. (2012). Importance of data preprocessing for neural networks modeling: The case of estimating the compaction parameters of soils.

Energy Education Science and Technology Part A: Energy Science and Research, *29*, 871-882.
Ismail, R., Alsadi, J., Hatamleh, R., Telfah, D., Jaradat, A., Aljamal, M., Trrad, I., & Al-Mattarneh, H. (2024). Employing CNN and black widow optimization for sustainable wastewater management in an environmental engineering context [Article]. *Asian Journal of Civil Engineering*, *25*(5), 3973-3988. https://doi.org/10.1007/s42107-024-01024-w
Jafar, R., Shahrour, I., & Juran, I. (2010). Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. *Mathematical and Computer Modelling*, *51*(9-10), 1170-1180.
Jeong, H., Baik, H.-S., & Abraham, D. (2005). *An Ordered Probit Model Approach for Developing Markov Chain Based Deterioration Model for Wastewater Infrastructure Systems*. https://doi.org/10.1061/40800(180)52
Jiang, G., Keller, J., Bond, P. L., & Yuan, Z. (2016). Predicting concrete corrosion of sewers using artificial neural network. *Water Research*, *92*, 52-60. https://doi.org/https://doi.org/10.1016/j.watres.2016.01.029
Kadhum, M., Eng, K., Abdulrazzaq, H., & Agha. (2016). Developing Artificial Neural Network and Multiple Linear Regression Models to Predict the Ultimate Load Carrying Capacity of Reactive Powder Concrete Columns. *8*.
Kargar, K., & Joksimovic, D. (2024). Analysis of sewer blockage causes using open data [Article]. *Water Practice and Technology*, *19*(9), 3855-3866. https://doi.org/10.2166/wpt.2024.218
Kerwin, S., Borja, G. d. S., Bryan, A., Kleio, S., & and Heller, H. (2023). Combining recorded failures and expert opinion in the development of ANN pipe failure prediction models. *Sustainable and Resilient Infrastructure*, *8*(1), 86-108. https://doi.org/10.1080/23789689.2020.1787033
Khan, S. (2025). The Application of Artificial Intelligence and Machine Learning in Civil Engineering. *Doupe Journal of Top Trending Technologies*, *1*, 1-11. https://doi.org/10.71063/DJTTT.2025.1103
Khan, Z., Zayed, T., & Moselhi, O. (2010). Structural condition assessment of sewer pipelines. Journal of Performance of Constructed Facilities,
Khan, Z., Zayed, T., & Moselhi, O. (2010). Structural Condition Assessment of Sewer Pipelines. *Journal of Performance of Constructed Facilities*, *24*(2), 170-179. https://doi.org/doi:10.1061/(ASCE)CF.1943-5509.0000081
Kingma, D., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*.
Kizilöz, B. (2024). Prediction of failures in sewer networks using various machine learning classifiers. *Urban Water Journal*, *21*(7), 877-893. https://doi.org/10.1080/1573062X.2024.2360184
Koo, D.-H., & Ariaratnam, S. T. (2006). Innovative method for assessment of underground sewer pipe condition. *Automation in Construction*, *15*(4), 479-488. https://doi.org/https://doi.org/10.1016/j.autcon.2005.06.007
Kretslopp och Vatten. (2025). *Dataset for ANN sewer model*.
Kulandaivel, G. (2004). *Sewer Pipeline Condition Prediction Using Neural Network Models*. Michigan State University. Department of Building Construction Management. https://books.google.se/books?id=VU56oAEACAAJ
Kutyłowska, M. (2017). Comparison of two types of artificial neural networks for predicting failure frequency of water conduits. *Periodica Polytechnica Civil Engineering*, *61*(1), 1-6.
Laakso, T., Ahopelto, S., Lampola, T., Kokkonen, T., & Vahala, R. (2018). Estimating water and wastewater pipe failure consequences and the most detrimental failure modes [Article]. *Water Science and Technology: Water Supply*, *18*(3), 901. https://doi.org/10.2166/ws.2017.164
Laakso, T., Kokkonen, T., Mellin, I., & Vahala, R. (2019). Sewer Life Span Prediction: Comparison of Methods and Assessment of the Sample Impact on the Results. *Water*, *11*(12), 2657. https://www.mdpi.com/2073-4441/11/12/2657

Larsson, D. (2015). Kolera. Samhället, idéerna och katastrofen 1834.
Latifi, M., Beig Zali, R., Javadi, A. A., & Farmani, R. (2024). Customised-sampling approach for pipe failure prediction in water distribution networks. *Scientific Reports*, *14*(1), 18224. https://doi.org/10.1038/s41598-024-69109-9
Li, X., Khademi, F., Liu, Y., Akbari, M., Wang, C., Bond, P. L., Keller, J., & Jiang, G. (2019). Evaluation of data-driven models for predicting the service life of concrete sewer pipes subjected to corrosion. *Journal of Environmental Management*, *234*, 431-439. https://doi.org/https://doi.org/10.1016/j.jenvman.2018.12.098
Li, Y., Wang, H., Dang, L. M., Song, H.-K., & Moon, H. (2023). Attention-guided multiscale neural network for defect detection in sewer pipelines. *Computer-Aided Civil and Infrastructure Engineering*, *38*(15), 2163-2179. https://doi.org/https://doi.org/10.1111/mice.12991
Lundberg, R. (2021). *Risk-based Rehabilitation of Wastewater Pipe* Chalmers]. Gothenburg.
Mahamud, A. H. (2023). Moving towards a proactive sewer pipe inspection approach: A state-of-the-art analysis. In.
Malek Mohammadi, M. (2019). *Development of Condition Prediction Models for Sanitary Sewer Pipes*
Malek Mohammadi, M., Najafi, M., Kermanshachi, S., Kaushal, V., & Serajiantehrani, R. (2019). Factors Influencing the Condition of Sewer Pipes: A State-of-the-Art Review. *Journal of Pipeline Systems Engineering and Practice*, *11*. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000483
Malek Mohammadi, M., Najafi, M., Serajiantehrani, R., Kaushal, V., & Hajyalikhani, P. (2021). *Using Machine Learning to Predict Condition of Sewer Pipes*. https://doi.org/10.1061/9780784483602.022
Malm, A., Horstmark, A., Jansson, E., Larsson, G., Meyer, A., & Uusijärvi, J. (2011). *Guideline for renewal planning of water and wastewater networks*. Svenskt Vatten AB
Malm, A., Horstmark, A., Larsson, G., Uusijärvi, J., Meyer, A., & Jansson, E. (2011). *Pipe material in Swedish water and wastewater pipes - properties and useful lifetime*. Svenskt Vatten Utveckling
Malm, A., & Svensson, G. (2011). *Material and age distribution of Sweden's water and wastewater networks and future renewal needs*. Svenskt Vatten Utveckling
Malm, A., Svensson, G., Bäckman, H., & Morrison, G. M. (2013). Prediction of water and wastewater networks rehabilitation based current age and material distribution. *Water Supply*, *13*(2), 227-237. https://doi.org/10.2166/ws.2013.011
Mehmood, H., Mukkavilli, S. K., Weber, I., Koshio, A., Meechaiya, C., Piman, T., Mubea, K., Tortajada, C., Mahadeo, K., & Liao, D. (2020). Strategic Foresight to Applications of Artificial Intelligence to Achieve Water-related Sustainable Development Goals.
Micevski, T., Kuczera, G., & Coombes, P. (2002). Markov Model for Storm Water Pipe Deterioration. *Journal of Infrastructure Systems - J INFRASTRUCT SYST*, *8*. https://doi.org/10.1061/(ASCE)1076-0342(2002)8:2(49)
Mohammadagha, M., Najafi, M., Kaushal, V., & Jibreen, A. (2025). *Machine Learning Models for Reinforced Concrete Pipes Condition Prediction: The State-of-the-Art Using Artificial Neural Networks and Multiple Linear Regression in a Wisconsin Case Study*. https://doi.org/10.48550/arXiv.2502.00363
Mohammadi, M. M., Najafi, M., Kaushal, V., Serajiantehrani, R., Salehabadi, N., & Ashoori, T. (2019). Sewer pipes condition prediction models: A state-of-the-art review [Review]. *Infrastructures*, *4*(4), Article 64. https://doi.org/10.3390/infrastructures4040064
Moradi, S. (2020). Defect Detection and Classification in Sewer Pipeline Inspection Videos Using Deep Neural Networks. https://spectrum.library.concordia.ca/986945/
Najafi, M. (2005). DESIGN CONSIDERATIONS FOR TRENCHLESS PIPELINE CONSTRUCTION METHODS. In *Trenchless Technology: Pipeline and Utility Design, Construction, and Renewal* (1st Edition ed.). McGraw-Hill Education. https://www.accessengineeringlibrary.com/content/book/9780071422666/chapter/chapter4

Najafi, M., & Kulandaivel, G. (2005). Pipeline condition prediction using neural network models. Proceedings of the Pipeline Division Specialty Conference,

Najar, N., & Persson, K. M. (2023). Status improvement in water and wastewater fixed facilities: Success and challenges of 11 Swedish water utilities as case studies. *Water Policy*, *25*(7), 656-679. https://doi.org/10.2166/wp.2023.263

Nguyen, L., & Seidu, R. (2022). Application of Regression-Based Machine Learning Algorithms in Sewer Condition Assessment for Ålesund City, Norway. *Water*, *14*(24). https://doi.org/10.3390/w14243993

Noshahri, H., olde Scholtenhuis, L. L., Doree, A. G., & Dertien, E. C. (2021). Linking sewer condition assessment methods to asset managers' data-needs. *Automation in Construction*, *131*, 103878. https://doi.org/https://doi.org/10.1016/j.autcon.2021.103878

O'Reilly, M. P., Transport, & Laboratory, R. R. (1989). *Analysis of Defects in 180 Km of Pipe Sewers in Southern Water Authority*. Ground Engineering Division, Structures Group, Transport and Road Research Laboratory. https://books.google.se/books?id=vMInMwAACAAJ

Ojha, D. (2024). Advancements and Challenges in the Application of Artificial Intelligence in Civil Engineering. *International Journal of Contemporary Research in Multidisciplinary*, *3*.

Okwori, E. (2021). *Data-driven approaches for proactive maintenance planning of sewer blockage management* [Licentiate thesis, comprehensive summary, Luleå University of Technology]. DiVA. Luleå. http://urn.kb.se/resolve?urn=urn:nbn:se:ltu:diva-83891

Okwori, E., Viklander, M., & Hedström, A. (2020). Performance assessment of Swedish sewer pipe networks using pipe blockage and other associated performance indicators [Article]. *H2Open Journal*, *3*(1), 46-57. https://doi.org/10.2166/h2oj.2020.027

Okwori, E., Viklander, M., & Hedström, A. (2024). Data integration in asset management of municipal pipe networks in Sweden: Challenges, gaps, and potential drivers. *Utilities Policy*, *86*, 101689. https://doi.org/https://doi.org/10.1016/j.jup.2023.101689

Onorato, G. (2024). *Bayesian Optimization for Hyperparameters Tuning in Neural Networks*. https://doi.org/10.48550/arXiv.2410.21886

Pargent, F., Pfisterer, F., Thomas, J., & Bischl, B. (2022). Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features. *Computational Statistics*, *37*, 1-22. https://doi.org/10.1007/s00180-022-01207-6

Phuenaree, B., & Sanorsap, S. (2017). An Interval Estimation of Pearson's Correlation Coefficient by Bootstrap Methods.

Prabhu, O. (2024, 05-08). Artificial Neural Networks. https://techz.vcet.edu.in/2024/05/04/artificial-neural-networks/

Pramoditha, R., & Oak, O. (2024). Training ANNs: A deep dive into Backpropagation and Gradient descent. https://medium.com/@omkarsoak/training-anns-a-deep-dive-into-backpropagation-and-gradient-descent-bea84ff5273c

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

Rallabandi, S. (2023, 05-04). Activation functions: ReLU vs. Leaky ReLU. https://medium.com/@sreeku.ralla/activation-functions-relu-vs-leaky-relu-b8272dc0b1be#:~:text=The%20main%20advantage%20of%20Leaky,addressing%20the%20dying%20ReLU%20problem.

Raschka, S., & Mirjalili, V. (2019). *Python Machine Learning*. Packt Publishing Limited.

Rickman, J. M., Wang, Y., Rollett, A. D., Harmer, M. P., & Compson, C. (2017). Data analytics using canonical correlation analysis and Monte Carlo simulation. *npj Computational Materials*, *3*(1), 26. https://doi.org/10.1038/s41524-017-0028-9

Robles-Velasco, A., Ramos-Salgado, C., Muñuzuri, J., & Cortés, P. (2021). Artificial Neural Networks to Forecast Failures in Water Supply Pipes. *Sustainability*, *13*(15).

Rodríguez, J. P., McIntyre, N., Díaz-Granados, M., & Maksimović, Č. (2012). A database and model to support proactive management of sediment-related sewer blockages. *Water Research*, *46*(15), 4571-4586. https://doi.org/https://doi.org/10.1016/j.watres.2012.06.037

Rokstad, M. M., & Ugarelli, R. M. (2015). Evaluating the role of deterioration models for condition assessment of sewers. *Journal of Hydroinformatics*, *17*(5), 789-804. https://doi.org/10.2166/hydro.2015.122

Rolf, K., & Stål, Ö. (1994). Tree Roots in Sewer Systems in Malmo, Sweden. *Arboriculture & Urban Forestry*.

Ruby, U., & Yendapalli, V. (2020). Binary cross entropy with deep learning technique for Image classification. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*. https://doi.org/10.30534/ijatcse/2020/175942020

Russell, & Norvig, S. (2010). Artificial Intelligence: A Modern Approach. *Prentice Hall, Englewood Cliffs, NJ*.

Sacluti, F. R. (1999). Modelling water distribution pipe failures using artificial neural networks.

Salihu, C., Mohandes, S. R., Kineber, A., Hosseini, M., Elghaish, F., & Zayed, T. (2023). A Deterioration Model for Sewer Pipes Using CCTV and Artificial Intelligence. *Buildings*. https://doi.org/10.3390/buildings13040952

Saravanan, R., & Vipulanandan, C. (2014). Testing wastewater pipe-joints for infiltration. Pipelines 2014: From Underground to the Forefront of Innovation and Sustainability - Proceedings of the Pipelines 2014 Conference,

Sattar, A. M., Ertuğrul, Ö. F., Gharabaghi, B., McBean, E. A., & Cao, J. (2019). Extreme learning machine model for water network management. *Neural Computing and Applications*, *31*, 157-169.

SCB. (2022). *Befolkningen ansluten till det kommunala nätet efter typ av anslutning och år*. https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__MI__MI0902__MI0902C/MI0902T05/chart/chartViewColumn/

SCB. (2025). *Population and population change - Second half of 2024*. Retrieved 04-11 from https://www.scb.se/hitta-statistik/statistik-efter-amne/befolkning-och-levnadsforhallanden/befolkningens-sammansattning-och-utveckling/befolkningsstatistik/pong/tabell-och-diagram/folkmangd-och-befolkningsforandringar---manad-kvartal-och-halvar/folkmangd-och-befolkningsforandringar---andra-halvaret-2024/

Semadeni-Davies, A., Hernebring, C., Svensson, G., & Gustafsson, L.-G. (2008). The impacts of climate change and urbanisation on drainage in Helsingborg, Sweden: Suburban stormwater. *Journal of Hydrology*, *350*(1), 114-125. https://doi.org/https://doi.org/10.1016/j.jhydrol.2007.11.006

Singh, A., & Adachi, S. (2013). Bathtub curves and pipe prioritization based on failure rate. *Built Environment Project and Asset Management*, *3*(1), 105-122. https://doi.org/10.1108/BEPAM-11-2011-0027

SMHI. (2021). *Dataserier med normalvärden för perioden 1991-2020*. Sveriges meteorologiska och hydrologiska institut. Retrieved 04-11 from https://www.smhi.se/data/temperatur-och-vind/temperatur/dataserier-med-normalvarden-for-perioden-1991-2020

Smith, W. (1968). Soil Evaluation in Relation to Cast-Iron Pipe. *Journal American Water Works Association*, *60*, 221-227. https://doi.org/10.1002/J.1551-8833.1968.TB03536.X

Sousa, V., Matos, J. P., & Matias, N. (2014). Evaluation of artificial intelligence tool performance and uncertainty for predicting sewer structural condition. *Automation in Construction*, *44*, 84-91. https://doi.org/https://doi.org/10.1016/j.autcon.2014.04.004

StormTac. (2025). *Guide, StormTac Web* https://app.stormtac.com/_dwl/Guide%20StormTac%20Web%20Eng.pdf

SWWA. (2007). *Avloppsteknik 1*. Svenskt Vatten AB

SWWA. (2018). Results for VASS pipe networks 2016 [Resultat för VASS Rörnät 2016].

SWWA. (2019). *Avledning av dag-, drän- och spillvatten P110*. Svenskt Vatten AB
SWWA. (2021a). *Effektivt Underhåll av VA-System*. Bromma, Sweden: Svenskt Vatten AB
SWWA. (2021b). *P122 weight factors card rating*. Svenskt Vatten AB
SWWA. (2023). *Investeringsbehov och framtida kostnader för kommunalt vatten och avlopp*. (R2023-02). Retrieved from https://vattenbokhandeln.svensktvatten.se/wp-content/uploads/2023/05/SvensktVatten_Investeringsrapport_2023.pdf
SWWA. (2024). *Resultatrapport för VASS Drift 2023*. Retrieved from https://www.svensktvatten.se/globalassets/dokument/organisation--styrning/va-statistik/r2024-06-resultatrapport-for-vass-drift-2023-3.pdf
Sörensen, J., Nilsson, E., Nilsson, D., Gröndahl, E., Rehn, D., & Giertz, T. (2024). Evaluation of ANN model for pipe status assessment in drinking water management. *Water Supply*, *24*(5). https://doi.org/10.2166/ws.2024.104
Tabesh, M., Soltani, J., Farmani, R., & Savic, D. (2009). Assessing pipe failure rate and mechanical reliability of water distribution networks using data-driven modeling. *Journal of Hydroinformatics*, *11*(1), 1-17.
Taheri, S., Ams, M., Bustamante, H., Vorreiter, L., Bevitt, J. J., Withford, M., & Clark, S. M. (2020). Characterizing concrete corrosion below sewer tidal levels at chemically dosed locations. *Water Research*, *185*, 116245. https://doi.org/https://doi.org/10.1016/j.watres.2020.116245
Taiwo, R., Zayed, T., Bakhtawar, B., & Adey, B. T. (2025). Explainable deep learning models for predicting water pipe failures [Article]. *Journal of Environmental Management*, *379*, Article 124738. https://doi.org/10.1016/j.jenvman.2025.124738
Tavakoli, R., Sharifara, A., & Najafi, M. (2019). *Prediction of Sewer Pipe Deterioration Using Random Forest Classification*. https://doi.org/10.48550/arXiv.1912.04194
Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, *17*(1), 168-192. https://doi.org/10.1016/j.aci.2018.08.003
Tran, D. H., M., N. A. W., C., P. B. J., S., B., & and Davis, P. (2006). Application of probabilistic neural networks in modelling structural deterioration of stormwater pipes. *Urban Water Journal*, *3*(3), 175-184. https://doi.org/10.1080/15730620600961684
Tran, D. H., Perera, B. J. C., & Ng, A. W. M. (2009). Comparison of Structural Deterioration Models for Stormwater Drainage Pipes. *Computer-Aided Civil and Infrastructure Engineering*, *24*(2). https://doi.org/10.1111/j.1467-8667.2008.00577.x
Tran, H. D. (2007). *Investigation of deterioration models for stormwater pipe systems* Victoria University].
Tscheikner-Gratl, F., Caradot, N., Cherqui, F., Leitão, J. P., Ahmadi, M., Langeveld, J. G., Gat, Y. L., Scholten, L., Roghani, B., Rodríguez, J. P., Lepot, M., Stegeman, B., Heinrichsen, A., Kropp, I., Kerres, K., Almeida, M. d. C., Bach, P. M., Vitry, M. M. d., Marques, A. S., . . . Clemens, F. (2019). Sewer asset management – state of the art and research needs. *Urban Water Journal*, *16*(9). https://doi.org/10.1080/1573062X.2020.1713382
United Nations. (2025a). *SDG Goal 6 Clean water and sanitation*. United Nations. Retrieved 04/03 from https://sdgs.un.org/goals/goal6
United Nations. (2025b). *SDG Goal 9 Industry, Innovation and Infrastructure*. United Nation. Retrieved 04/03 from https://sdgs.un.org/goals/goal9
United Nations. (2025c). *SDG Goal 11 Sustainable Cities and Communities*. United Nation. Retrieved 04/03 from https://sdgs.un.org/goals/goal11
United Nations. (2025d). *SDG Goal 13 Climate Action*. Retrieved 04/03 from
Vinayak, H. (2024). Review of AI in Civil Engineering. *Annals of Civil and Environmental Engineering*. https://doi.org/10.29328/journal.acee.1001066
Vipulanandan, C., & Liu, J. (2005). Sewer pipe-joint infiltration test protocol developed by CIGMAT. Proceedings of the Pipeline Division Specialty Conference,
Vyas, P., Sharma, M., Rasool, A., & Dubey, A. (2023). Comparative Study of Loss Functions for Imbalanced Dataset of Online Reviews. In (pp. 115-126). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-0047-3_11

Walczak, S., & Cerpa, N. (2003). Artificial Neural Networks. In R. A. Meyers (Ed.), *Encyclopedia of Physical Science and Technology (Third Edition)* (pp. 631-645). Academic Press. https://doi.org/https://doi.org/10.1016/B0-12-227410-5/00837-1

Wang, A.-J., Li, H., He, Z., Tao, Y., Wang, H., Yang, M., Savic, D., Daigger, G. T., & Ren, N. (2024). Digital Twins for Wastewater Treatment: A Technical Review. *Engineering*, *36*, 21-35. https://doi.org/https://doi.org/10.1016/j.eng.2024.04.012

Wang, W., Robert, D., Zhou, A., & Li, C. (2016). Effect of corrosion influencing factors of cast iron pipes in clay soil. 357-362. https://consensus.app/papers/effect-of-corrosion-influencing-factors-of-cast-iron-pipes-wang-robert/068bdae2d0785fdebdac7049557daef1/

Wear, S. L., Acuña, V., McDonald, R., & Font, C. (2021). Sewage pollution, declining ecosystem health, and cross-sector collaboration. *Biological Conservation*, *255*. https://doi.org/10.1016/j.biocon.2021.109010

Weerasinghe, D., Kodikara, J., & Bui, H. (2015). *Impact of Seasonal Swell/Shrink Behavior of Soil on Buried Water Pipe Failures*.

Wei, H., Wang, Y., & Wang, M. (2018). Characteristic and pattern of urban water cycle: theory. *Desalination and Water Treatment*, *110*, 349-354. https://doi.org/10.5004/dwt.2018.22342

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.

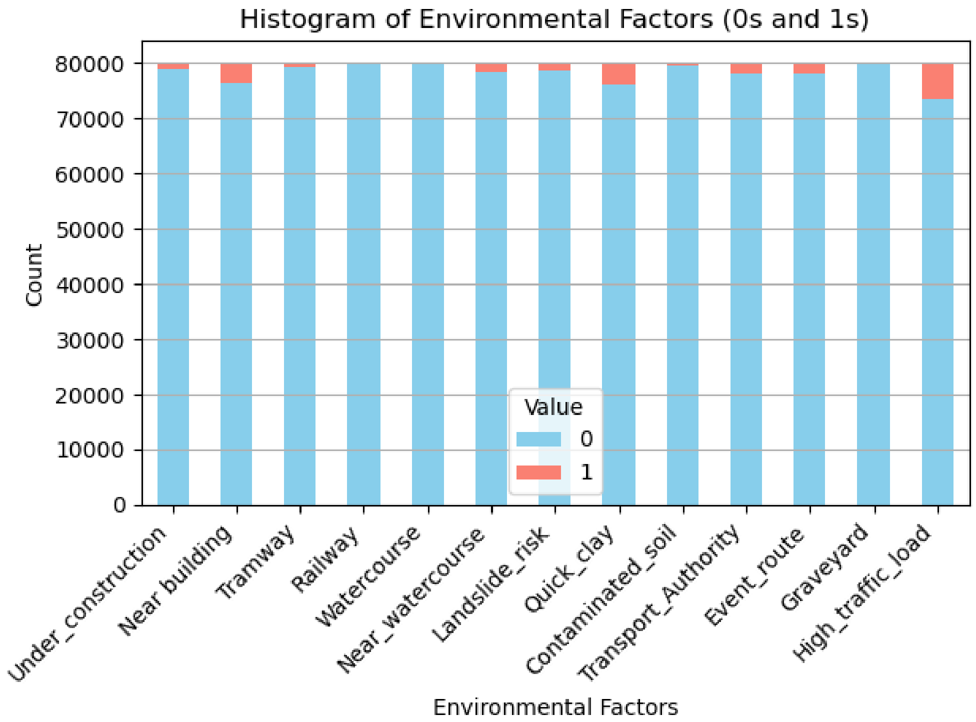
Zamanian, S., Rahimi, M., & Shafieezadeh, A. (2020). Resilience of Sewer Networks to Extreme Weather Hazards: Past Experiences and an Assessment Framework. Pipelines 2020: Utility Engineering, Surveying, and Multidisciplinary Topics - Proceedings of Sessions of the Pipelines 2020 Conference,

Zeydalinejad, N., Javadi, A. A., & Webber, J. L. (2024). Global perspectives on groundwater infiltration to sewer networks: A threat to urban sustainability. *Water Research*, *262*, 122098. https://doi.org/https://doi.org/10.1016/j.watres.2024.122098

Ziemba, E., Duong, E., Gonzalez-Perez, J., Kazlauskaitė, M., Mazurek, P., Wach, J., Ziemba, E., Cong, D., Duong, D., Ejdys, J., Gonzalez-Perez, M. A., Kazlauskaite, R., Korzynski, P., Mazurek, G., Paliszkiewicz, J., Stankevičienė, J., & Wach, K. (2024). Leveraging artificial intelligence to meet the sustainable development goals. *Journal of Economics and Management*, *46*. https://doi.org/10.22367/jem.2024.46.19

Åhnberg, H., Larsson, R., & Holmén, M. (2013). Degradation of clay due to cyclic loadings and deformations. Proceedings of the 18th international conference on soil mechanics and geotechnical engineering ICSMGE, Paris,

APPENDIX 1: IMBALANCE IN ENVIRONMENTAL FACTORS



Histogram of Environmental factors (0s and 1s) presents the dataset that contains a disproportionately high number of 0s compared to 1s, which may influence model performance.

APPENDIX 2: MODEL EVALUATION

Confusion matrices for INL, SPR, RBR, and DEF.

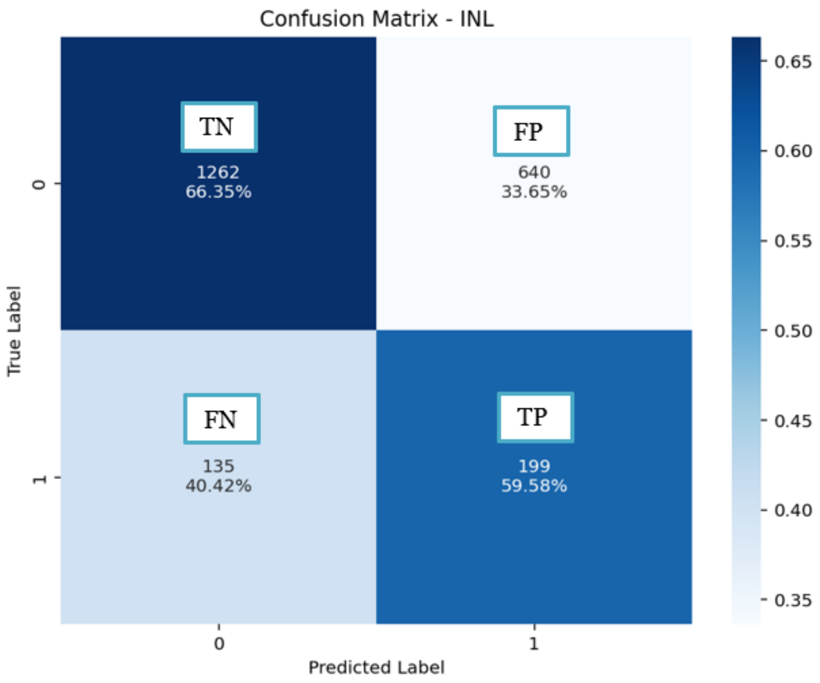


Figure 2A Confusion matrix for INL (infiltration).

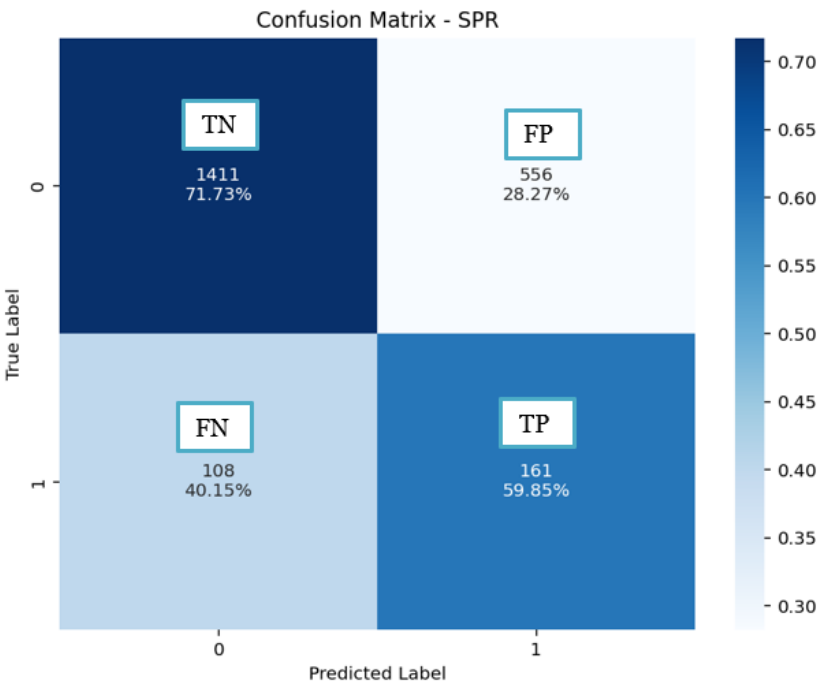


Figure 2B Confusion matrix for SPR (cracks).

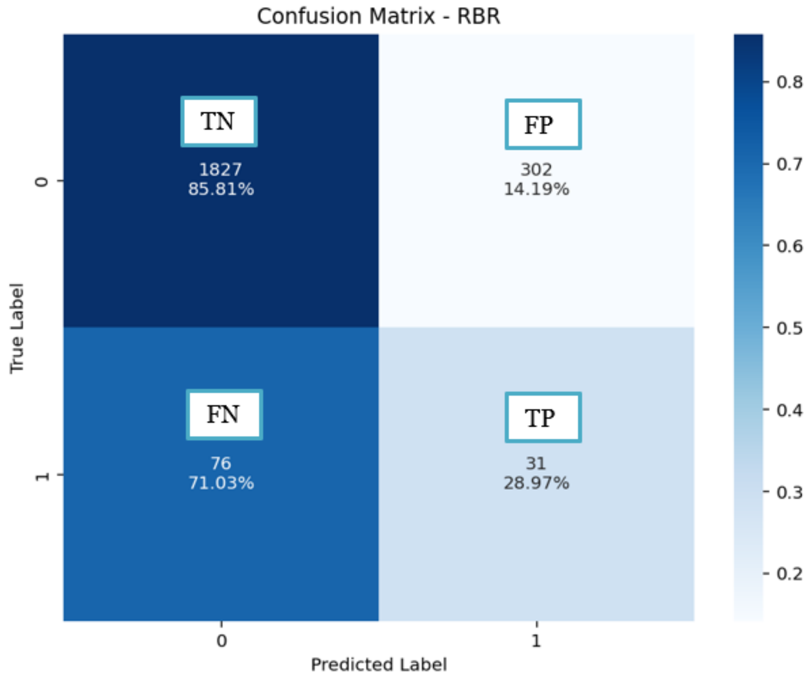


Figure 2C Confusion matrix for RBR (rupture/collapse).

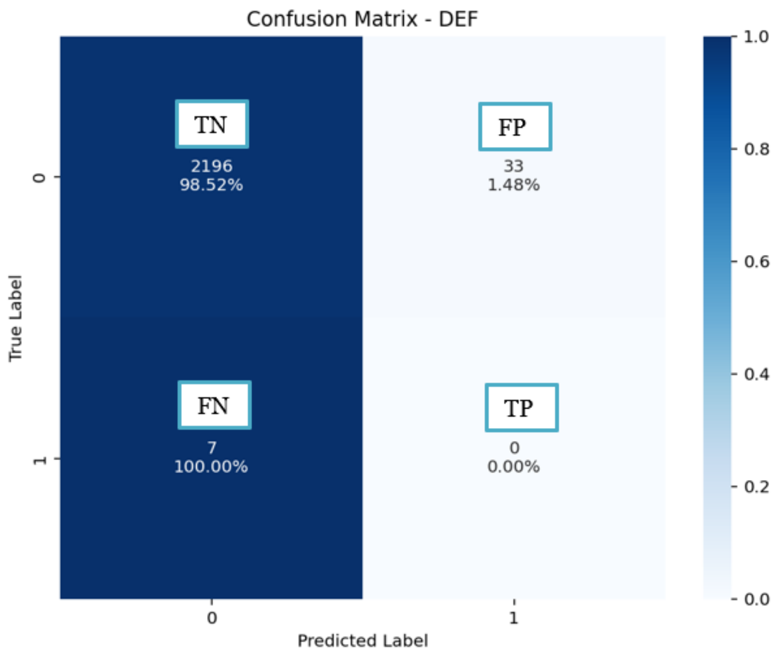


Figure 2D Confusion matrix for DEF (deformation).

Calculation of recall and specificity:

Table 2A Recall and specificity calculation

Model	**Recall = TP/(TP+FN)**	**Specificity = TN/(TN+FP)**
INL	199/(199+135) = 0.60	1262/(1262+640) = 0.66
SPR	161/(161+108) = 0.60	1411/(1411+556) = 0.72

RBR	31/(31+76) = 0.29	1827/(1827+302) = 0.86
YTS	378/(378+117) = 0.76	1107/(1107+634) = 0.64
DEF	0/(0+7) = 0	2196/(2196+33) = 0.99

Training and validation loss curves for INL, SPR and RBR:

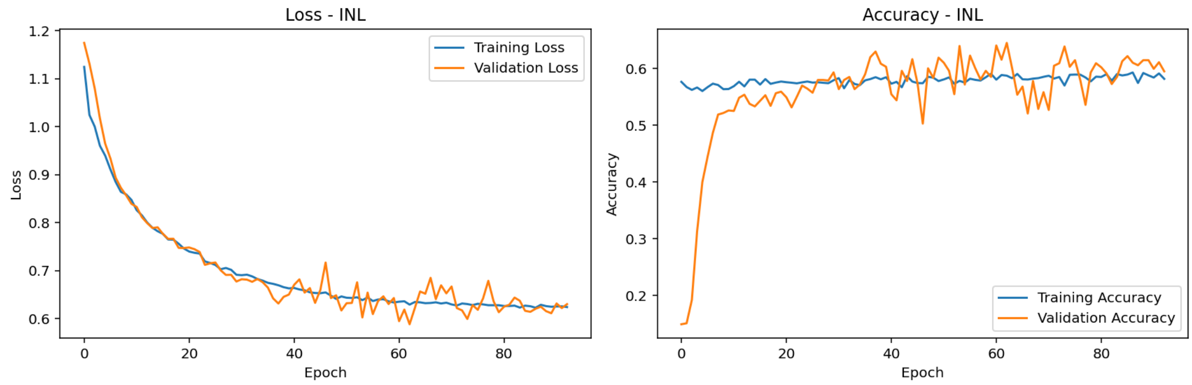


Figure 2E Training and validation loss curves for INL.

Figure 2F Training and validation loss curves for SPR.

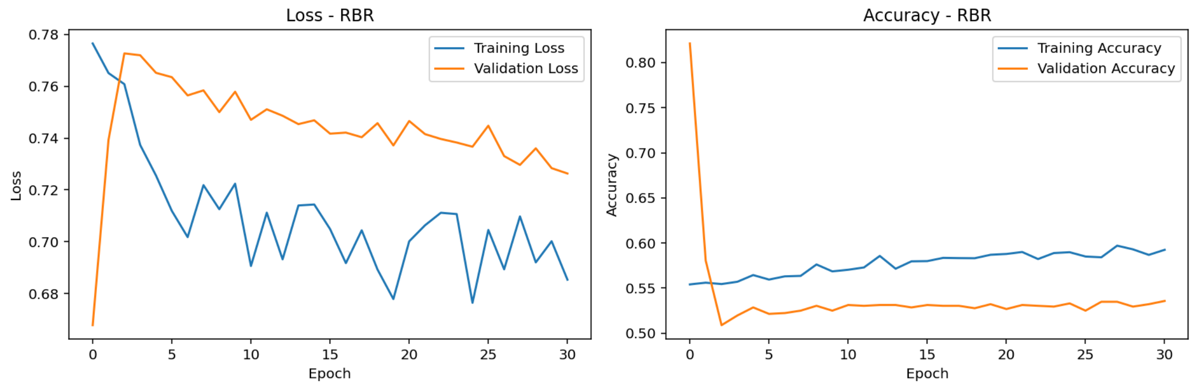


Figure 2G Training and validation loss curves for RBR.

Prediction group evaluation (KoV's recommendation) for INL, SPR, RBR, and DEF.

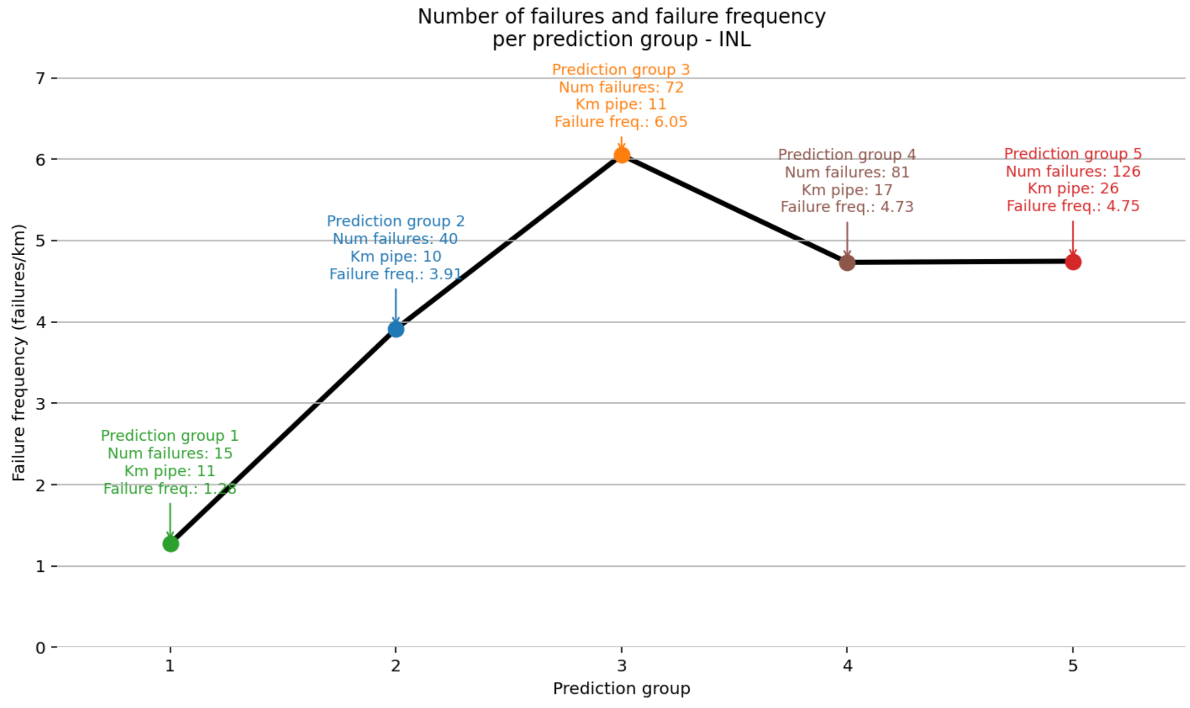


Figure 2H Failure frequency per prediction group for the INL model on the test set.

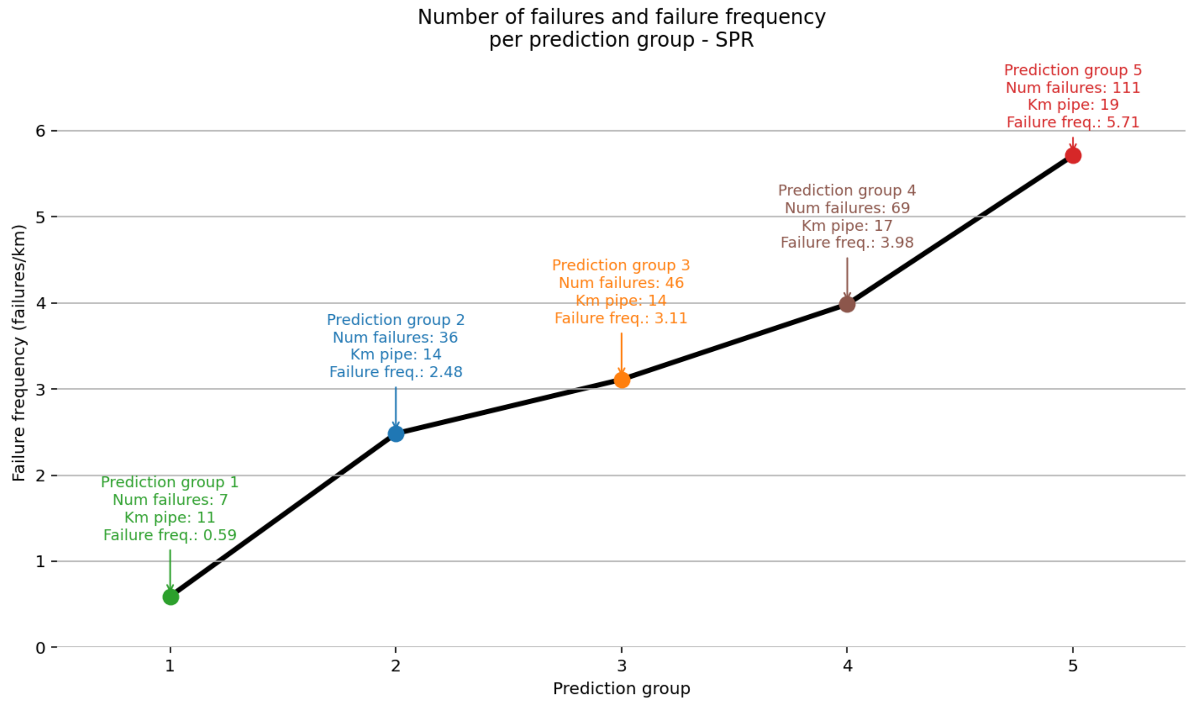


Figure 2I Failure frequency per prediction group for the INL model on the test set.

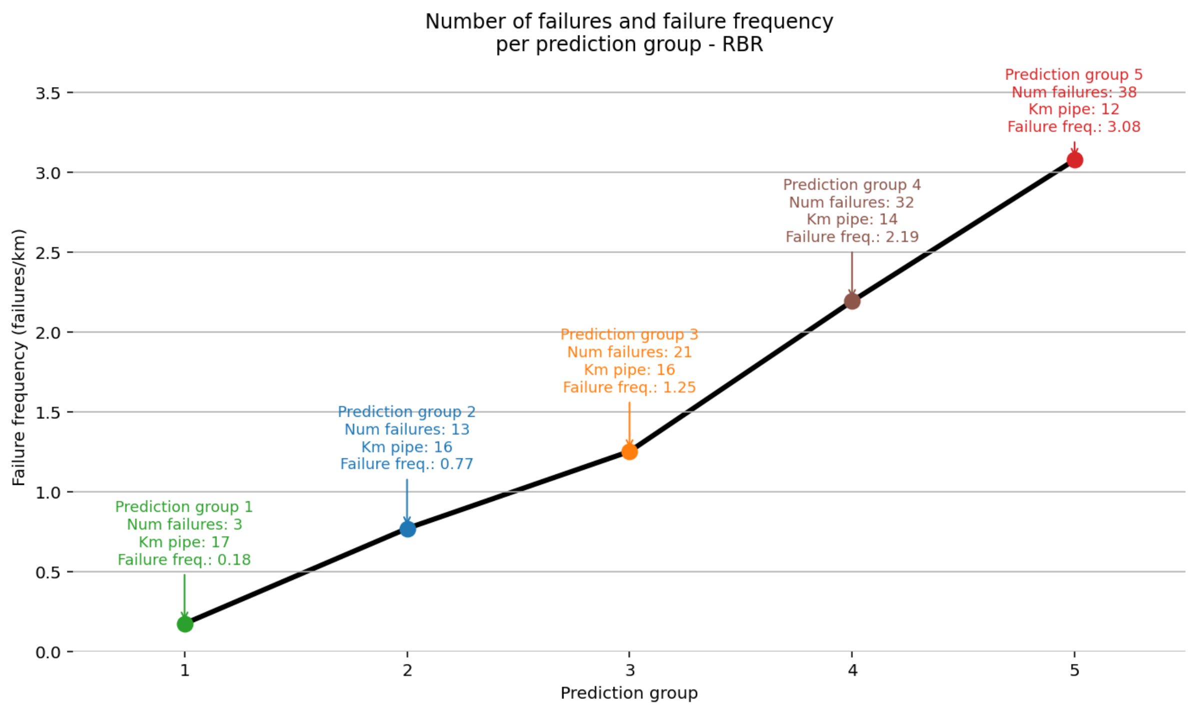


Figure 2J Failure frequency per prediction group for the RBR model on the test set.

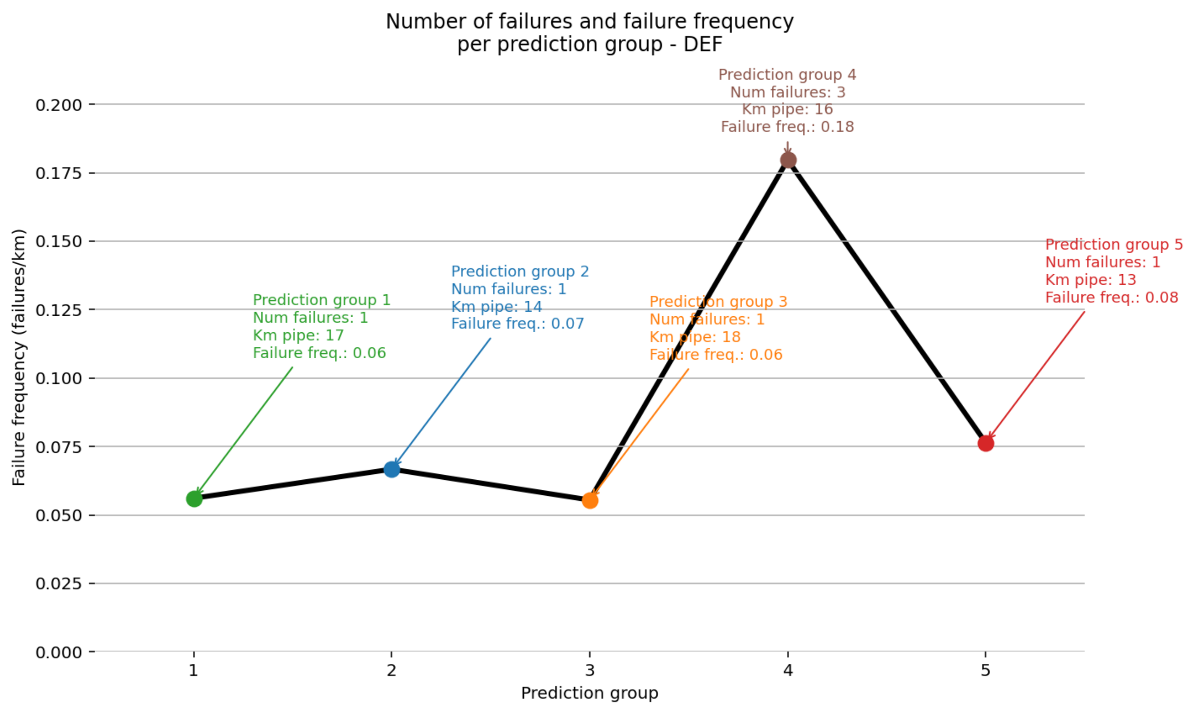


Figure 2K Failure frequency per prediction group for the DEF model on the test set.

A comparison of feature importance before and after removing renovation data & soil transition type (soil_cross):

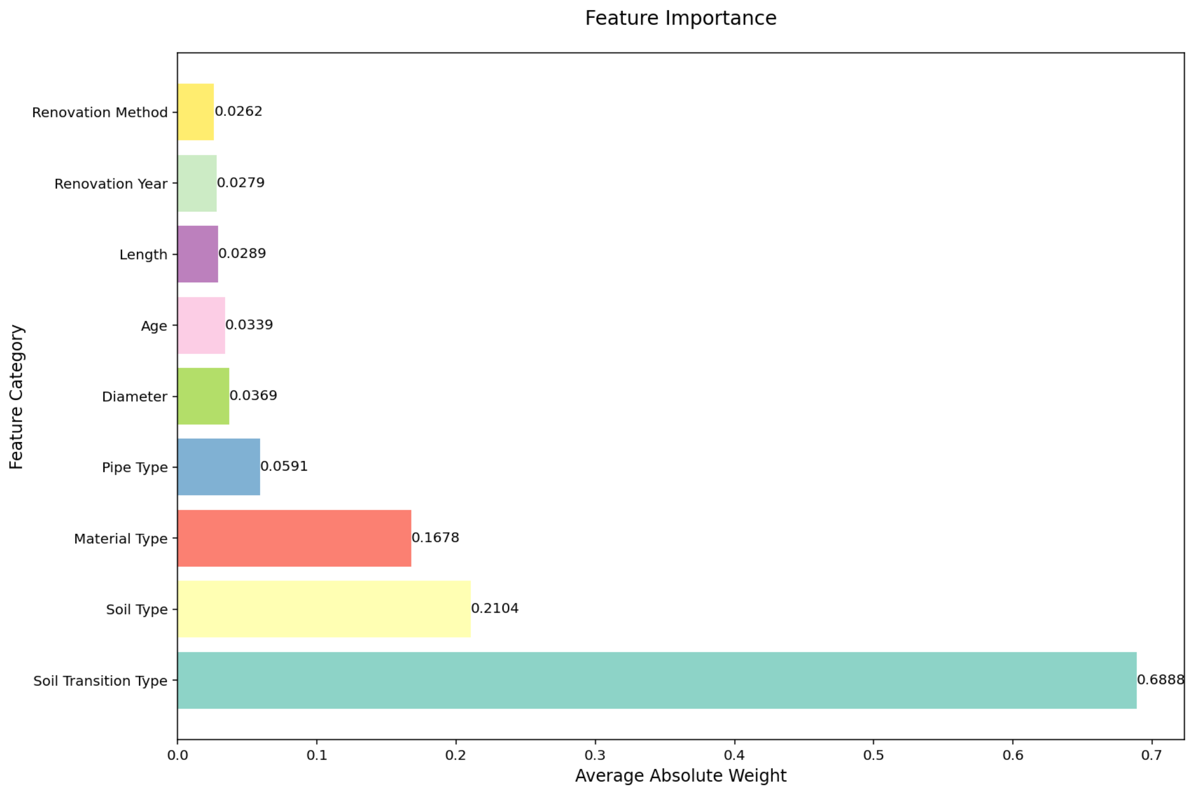


Figure 2L Feature importance for SPR, with renovation data and soil transition type (encoded by one-hot encoder).

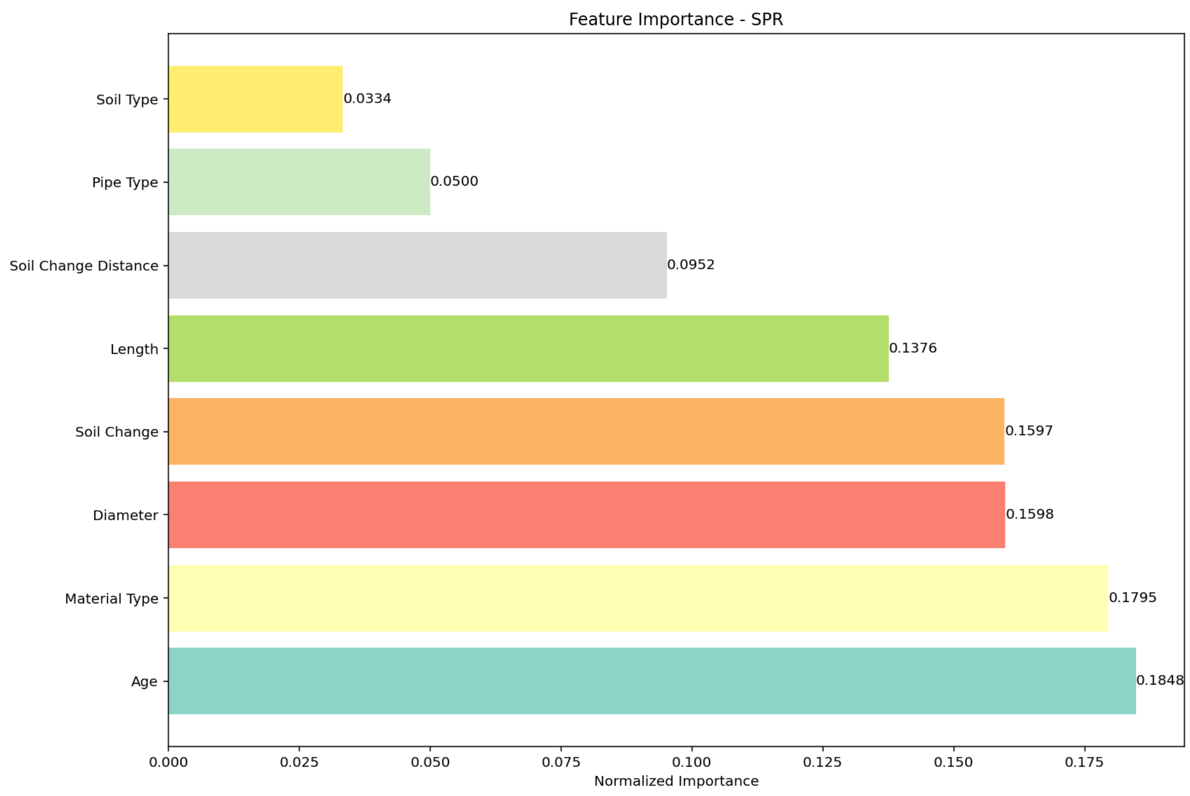


Figure 2M Feature importance for SPR, without renovation data and soil transition type. Soil change and soil change distance replaced soil transition type. Soil change was encoded with target encoding.