DEVELOPMENT AND COMPARISON OF PREDICTION MODELS FOR SANITARY SEWER PIPES CONDITION ASSESSMENT USING MULTINOMIAL LOGISTIC REGRESSION AND ARTIFICIAL NEURAL NETWORK

by

DANIEL OGARO ATAMBO

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2021

Acknowledgements

I am highly indebted and thankful to my supervisor Dr. Mohammad Najafi, P.E., F. ASCE, Professor and Director of the Center for Underground Infrastructure Research & Education (CUIRE), Department of Civil Engineering, The University of Texas at Arlington, for teaching and guiding me diligently in this research work. Dr. Najafi mentored and motivated me in all the steps of my research work.

Many thanks go to all members of my PhD Dissertation Committee, Dr. Ardeshir Anjomani, Dr. Vinayak Kaushal, Dr. Sharareh Kermanshachi, and Dr. Melanie Sattler for their valuable comments, suggestions, and advice. I cannot forget Dr. Mostafa Ghandehari for his critique and suggestions on statistical analysis.

I am grateful to my wife Florence and children Eugine, Ruth, and Sandra for their moral support and encouragement as I pursued and completed this work. This research work is dedicated to them.

Many thanks to the City of Dallas Water Utilities department for providing me with sanitary sewer condition assessment data that I used in this research work. In a special way I thank Sophia Harvey, Wastewater Collection Division Manager, Dallas Water Utilities, City of Dallas for accepting and approving me to use the sanitary sewer pipe condition assessment data in this research work. I appreciate Paul Sill, GIS Technical Manager and Rob Street, GIS Analyst of Dallas Water Utilities, City of Dallas for helping me extract the data from the City of Dallas GIS web database.

I would like to thank Rishi Bhattarai, Program Manager Pipeline Project Management Division and Matt Penk, Assistant Director Capital Improvement, Dallas Water Utilities, the City of Dallas for their support.

Finally, I sincerely thank my colleagues in CUIRE laboratory for their support and suggestions. Dr. Zahra Kouchesfehani encouraged me during my research work. Dr. Amin Darabnoush Tehrani helped me in learning MATLAB software. Roz Javadi Zadeh and Elham Ahmadi helped me in scheduling meetings with Dr. Najafi and collecting signatures from committee members.

August 16, 2021

Abstract

DEVELOPMENT AND COMPARISON OF PREDICTION MODELS FOR SANITARY SEWER PIPES CONDITION ASSESSMENT USING MULTINOMIAL LOGISTIC REGRESSION AND NEURAL NETWORK

Daniel Ogaro Atambo, Ph.D.

The University of Texas at Arlington, 2021

Supervising Professor: Mohammad Najafi

Sanitary sewer pipes infrastructure system in good condition is essential in providing safe conveyance of the wastewater from homes, businesses, and industries to the wastewater treatment plants. For sanitary sewer pipes to deliver the wastewater to the treatment plants, they must be in good condition. Most of the water utilities have aged sanitary sewer pipes. Water utilities inspect sewer pipes to decide which segments of the sanitary sewer pipes need rehabilitation or replacement. The process of inspecting the sewer pipes is described as condition assessment. This condition assessment process is costly and necessitates developing a model that predicts the condition rating of sanitary sewer pipes. The objective of this dissertation is to develop Multinomial Logistic Regression (MLR) and Artificial Neural Network (ANN) models to predict sanitary sewer pipes condition rating using inspection and condition assessment data. MLR and ANN models are developed from the City of Dallas' data. The MLR model is built using 80% of randomly selected data and validated using the remaining 20% of data. Similarly, the ANN model is trained, validated, and tested. The results of this research reveal that MLR and ANN models are acceptable. The significant physical factors influencing sanitary pipes condition rating include diameter, age, pipe material, and segment length. Soil type is the most important environmental factor that influences sanitary sewer pipes condition rating. The accuracy of the performance of the MLR and ANN is found to be 75% and 85%, respectively. This dissertation contributes to the body of knowledge by developing models to predict sanitary sewer pipes condition rating that enables policymakers and sanitary sewer utilities managers to prioritize the sanitary sewer pipes to be rehabilitated and/or replaced.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction and Background

1.1 Introduction

The provision of wastewater services to communities and municipalities is essential for public health, safety, and socioeconomic development. This requires wastewater infrastructure systems that collect wastewater from homes, businesses, and industry and convey the sewer to the treatment plants. The components of wastewater infrastructure systems include service laterals, sewer pipelines, manholes, force mains, siphons, combined sewer overflow regulations, pumping stations, and wet wells. According to the 2021 American Society of Civil Engineers (ASCE) Infrastructure Report Card, wastewater infrastructure was given a D+ score. It is unpredictable to know where or when an accidental pipeline failure would occur. Consequently, and to mitigate this, regulating agencies demand wastewater collection systems conduct periodic sewer inspections to comply with legal requirements. Due to limited budgets, however, not all segments of sewer pipes in wastewater collections systems can be inspected and assessed in a short time. To address this shortcoming of assessing criticality of the sewer pipes, utilities need pipe condition estimation models.

There are three types of sewer infrastructure condition prediction models: physical, artificial intelligence, and statistical (Hawari et al. 2020). Sanitary sewer pipes prediction models are classified into different categories (Mohammadi et al. 2019). The category of the models are statistical models and artificial intelligence models. The statistical models are discriminant analysis, logistic regression, binary regression, linear regression, exponential regression, Markov-chain, semi Markov-chain, ordinal regression, and cohort survival. The artificial intelligence models are classified into two. The two categories of artificial intelligence models are neural network and genetic algorithms and machine learning. The neural network and genetic algorithms include Artificial Neural Network (ANN) and fuzzy logic. Machine learning models include support vector machine, decision trees, random forest, and Bayesian networks.

In this dissertation, Multinomial Logistic Regression (MLR) and ANN prediction models are developed. MLR is a statistical tool that establishes relationships of independent and dependent variables. ANN is a computer mathematical tool that processes data as input and produces results as the output. The data is processed in layers. The input data is processed in the input layer. Between the input and output layers there is a hidden layer. The hidden layer is like a black box. The layers are connected to each other by weighted connections. The weights are calculated through the training process. The basic elements of the layers are nodes or neurons. The nodes or neurons are interconnected to each other to perform numerical manipulations that produces results as output. Then physical and environmental factors influencing sanitary sewer pipe condition were evaluated. First, the physical variables evaluated are pipe material, pipe diameter, pipe length, age, depth, and pipe slope. Second, the environmental variables are surface condition, soil type, corrosivity concrete, corrosivity steel, and soil pH. The physical and environmental factors are used in the development of sanitary sewer pipe condition rating prediction models. To build estimation models, extensive inspections and condition assessment past data set are required.

1.2 Background

Many water utilities in the U.S. have aging sewer pipeline and critical utility infrastructure. Critical utility and sewer pipeline infrastructure includes water, power, transportation, telecommunication, and wastewater management. Access to high levels of drinking water and wastewater services is fundamental for the protection of public health, the comfort and well-being of the population, sustainable development of the community, and environmental protection (Alegre, 2010). Most of the existing wastewater pipelines are deteriorating. In fact, the 2021 ASCE Infrastructure Report Card gave wastewater infrastructure a D+. The aging of wastewater pipes can result in an increase in pipe failure rates. Failed pipes are costly to repair and replace and can result in social and economic consequences (Opila, 2011). Water utilities are faced with challenges in operating and maintaining aging wastewater pipelines; the older the network gets, operation and maintenance (O&M) expenses increase (Ugarelli et al., 2008). Inspection, condition assessment, renewal of the sewer pipes is not fast enough to keep up with failure of sewer pipes. With limited budgets, policy makers and utility managers must make rational decisions in replacing and/or rehabilitating the pipelines. Asset managers need to make decisions regarding the selection of optimal rehabilitation action for each sewer condition (Wirahadikusumah et al., 1999). Managing these assets rationally is, therefore, fundamental for the sustainability of the services and to the economy of societies (Alegre, 2010). To be able to make effective decisions, most water utilities are implementing asset management concepts. This involves mapping and condition assessment of the wastewater collection systems. In reviewing the sustainability of urban water systems, Bruaset et al., 2018 concluded that identifying and implementing sustainable rehabilitation interventions in the long-term is essential for the survival of a high service level urban water system.

It is costly to perform condition assessment for all sanitary sewer collection system pipes. Most of the time, wastewater utilities set a goal of the percentage of sewer pipe segments to be inspected and replaced annually. Generally, condition check of the pipes is conducted when there is already a pipe failure.

This dissertation is focused on development and comparison of MLR and ANN models. MLR and ANN prediction models were developed to determine sewer pipes condition rating.

1.3 Problem Statement

The development of models requires large amounts of data to be collected. In addition, sewer pipes condition assessment must be periodically repeated to capture changing conditions. The variables used to develop the models vary from one region to another. There is a high demand to develop models that can be applicable and accurate in all regions that have similar characteristics.

In their review on sewer pipes condition prediction models Mohammadi et al. (2019) stated that more investigation is required to identify the influence of physical and environmental factors that affect deterioration of sewer pipes. To date, they noted, few studies had considered the effect of independent variables on condition of sewer pipes. Additionally, they recommended future research to investigate more pipe material such as steel and concrete pipes in sewer networks and compare the results, and that results of prediction models should be developed for different cities.

Salman’s and Salem’s (2004) review on sewer pipes condition prediction models observed that there is a need for more research to predict condition of sewer pipes with higher accuracy and confidence level. Following his research on advanced sewer asset management using dynamic deterioration models Syachrani (2010) discovered there was still room for improvement. Accordingly, he recommended, in future research, a more comprehensive model be developed by incorporating additional location related attributes such as soil type, water table, among others. Further, Syachrani recommended it be necessary to improve the accuracy of the consequence of failure estimate by using quantitative measures. He maintained it would be interesting to determine how the consequence of failure components could change depending on the size of utility (small, medium, or large) and its setting (rural or urban). In a later study Syachrani et al.,

(2013) advised municipalities to develop and implement risk assessment models for their utilities to get the best utility of their limited budgets available for replacing deteriorating assets.

Vahidi et al., (2016) in their research on infrastructure management and deterioration risk assessment of wastewater collection systems advised that the deterioration models can be improved by addition or consideration of other independent variables such as soil type, groundwater level, and initial quality of construction. In line with Vahidi et al's recommendation, environmental factors including surface condition, soil type, corrosivity concrete, corrosivity steel, and pH considered in building the models were included in this dissertation. Surface condition and corrosivity variables have not been studied more by others.

According to Caradot et al., (2018), the improvement of technical asset management and the use of digital solutions to improve the efficiency of inspection and rehabilitation strategies is the promising leverage of utilities. Caradot et al., (2018), further stated that, most metrics are based on statistical and do not provide understanding of deteriorations for sewer operators. Accordingly, there is need to utilize artificial intelligence methods and compare the results to those of statistical methods.

1.4 Objectives

1.4.1 Main objective

The main objective of this dissertation is to develop MLR and ANN models to predict sanitary sewer pipe condition rating using inspection and condition assessment data. The secondary objectives of this research include:

i. To identify, evaluate, categorize, and develop relationships of different factors affecting sewer pipes condition rating.

ii. To compare the performance of MLR and ANN models for predicting sewer pipes condition rating.

1.5 Scope of Work

Table 1-1 presents scope of this dissertation.

Table 1-1 Scope of Work

Included	Not Included
• The condition scores and pipe material, diameter, age, slope, depth, surface condition, soil type, corrosivity concrete, corrosivity steel, and pH variables obtained from condition assessment. • Data extracted from GIS files for the City of Dallas GIS Database.	• Pretreatment and Wastewater Treatment Plants (WTPs). • Financial infrastructure of wastewater collection system. • Use, storage, or handling of chemicals. • Stormwater pipes • Force main sewer pipes

1.6 Methodology

The methodology in this research involves five (5) steps, introduction, and background; literature review; methodology, data collection and model development; results and discussions; conclusions and recommendations for future research. These steps are presented in Figure 1-1.

First, utilizing engineering journals, databases, and Google Scholar, a thorough literature review is conducted to mainly study current sewer pipe prediction models, modes of sewer pipe failure, and variables for sewer pipe failure. In addition to physical, environmental, and operational factors influencing sewer pipes failure, literature on the failure of sewer pipe and risk process evaluation is reviewed. Second, data is collected from Geographical Information System (GIS) shape files for the City of Dallas Water Utilities (DWU) GIS Database. The GIS data originates from condition assessment and CCTV inspection records. The data is comprised of pipe segments/locations, length (manhole to manhole), pipe material, pipe diameter, pipe age (current year minus year of installation), depth (depth of backfill over the crown of pipe in ft), soil conditions, corrosivity, slope, surface condition – highway/street, and PACP condition rating. Third, the data was prepared, processed, and analyzed. Condition rating was designated as the dependent variable while physical and environmental factors were used as independent variables.

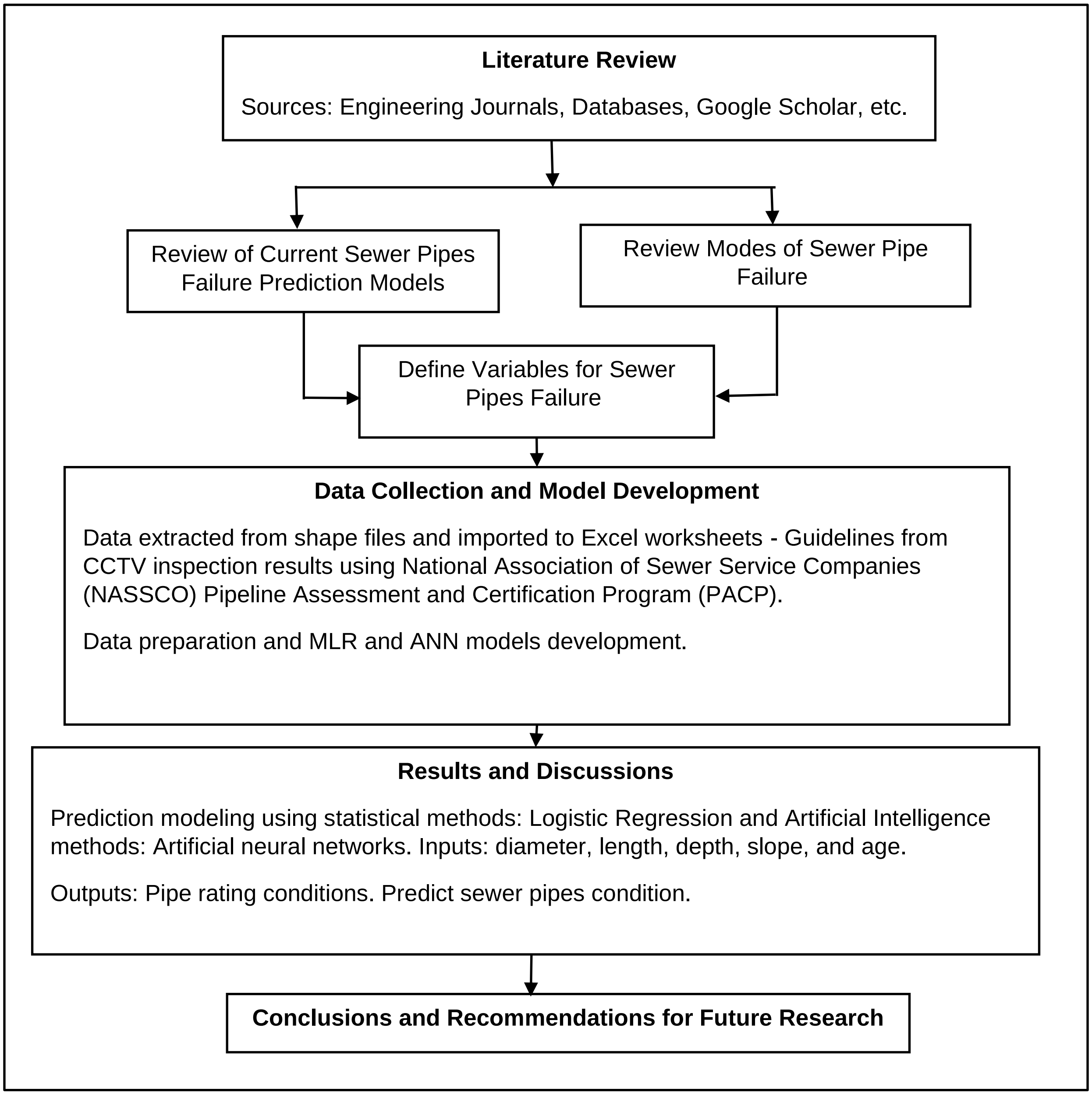

Figure 1-1 Dissertation Methodology

In this research, statistical (MLR) and artificial intelligence (ANN)models are developed to predict sewer pipes condition rating. The MLR and ANN models are applied to the Dallas Water Utilities' wastewater collection System. It is expected the models could be generalized and applied to regions with similar characteristics as that of Dallas Fort Worth metropolitan area (DFW).

1.7 Expected Outcome

In this this dissertation, MLR statistical and ANN Artificial Intelligence methods are used to build a model that predict sewer pipes condition rating. Utilizing the physical and environmental as independent variables the two methods are evaluated and compared. The comparison of the performance of the models assists in the selection of the best prediction model.

1.8 Contributions to the Body of Knowledge

In this dissertation wide range of physical and environmental factors are used to develop MLR and ANN pipes condition prediction models. All types of pipe materials, pipe sizes, and surface conditions are considered to estimate the condition of sewer pipe. In previous studies the comparison of MLR and ANN is not commonly conducted. Predicting and knowing the sanitary sewer pipes condition rating score enables policy makers and sanitary sewer utilities managers in decision making in prioritizing segments of sanitary sewer pipes to be rehabilitated and/ or replaced.

1.9 Future Research

The future research is discussed in chapter 7 section 7.2.

1.10 Hypotheses

1.10.1 Hypothesis 1

Null hypothesis (H_0): Pipe material, diameter, age, slope, and depth variables do not influence sewer pipes condition rating.

Alternative hypothesis (H_A): Pipe material, diameter, age, slope, and depth variables influence sewer pipes condition rating.

1.10.2 Hypothesis 2

Null hypothesis (H_0): Surface condition, soil type, corrosivity concrete, corrosivity steel, and pH are insignificant variables in sewer pipes condition rating.

Alternative hypothesis (H_A): Surface condition, soil type, corrosivity concrete, corrosivity steel, and pH are significant in sewer pipes condition rating.

1.11 Organization of Dissertation

This dissertation is divided into seven chapters. Chapter one included introduction and background information about condition of sanitary sewer pipes as well as the importance of sewer pipes condition assessment and prediction models. In addition, the problem statement, objectives, scope of work, methodology, expected outcome and contribution to the body of knowledge are discussed.

Chapter two presents a comprehensive review of literature on risk factors associated with sanitary sewer pipe asset failures. In addition to sewer pipe classification by material, diameter, and other attributes, sewer pipe failure predictions methods are reviewed. Steps of acquiring a risk model is included as well.

Chapter three present logistic regression and neural network methods. Regression assumptions test and test of significance of the coefficients of logistic regression models are discussed. In addition, the area under the receiver operating characteristic curve (ROC), ANN architecture backpropagation neural network (BPNN), validation, testing, and optimization of the performance of ANN model are discussed.

Chapter four present data collection, preparation, and preliminary data analysis. In this chapter data collected is described. The description of dependent and independent variables is provided.

Chapter five discusses model development. The MLR and ANN models are developed, validated, and tested. Logistic regression and ANN models are validated and tested.

Chapter six presents results and discussions. Results of MLR and ANN models are discussed. Performance of MLR and ANN models, MLR classification table, ROC, and justification of results are discussed in this chapter. The MLR and ANN are compared, and best prediction model is selected.

Chapter seven presents the dissertation conclusions, and recommendations for future research. The significant factors influencing sanitary sewer pipes condition rating are identified. The practical applications of the dissertation are discussed.

1.12 Chapter Summary

This chapter discussed background information about sanitary sewer pipes condition. The importance of sewer pipes condition assessment and prediction models were discussed. The statement of the problem, objectives, scope of work, research methodology, expected outcome, and contribution to the body of knowledge were examined.

Literature Review

2.1 Introduction

This chapter presents the background of sewer pipes. A detailed literature review is presented in this chapter on sewer pipes classification, risk of sewer pipe, risk process evaluation and modelling methods, rehabilitation, and renewal strategies. Modes of sewer pipes failure are reviewed, and the parameters that affect the performance of sewer pipes are comprehensively analyzed. The mode and frequency of failure is dependent on the physical condition of pipe, performance (O&M) condition, and the effect of environmental condition. Each of these variables is discussed in detail. Figure 2-1 illustrates flow chart for the outline of this chapter.

2.2 Sewer Pipe Classification

Sewer pipe are classified based on several attributes. These include pipe material, pipe diameter, pipe age, pipe length, pipe depth Pipe slope, surface condition, soil type, and corrosivity. Burn et al. (2010), identified the pipe characteristics that are significant in sewer pipes failure as pipe material, diameter, soil type, operating pressure, road type, road surface condition, and density of service connection.

2.2.1 Sewer Pipe Material

Cement-based pipes, Vitrified clay pipe (VCP), plastic pipes, and metallic pipes are four categories of pipe material. Najafi (2016), discussed types of these pipe categories as: Cement based pipes include concrete pipes and asbestos-cement (AC) pipe. Concrete pipes are nonreinforced concrete pipe (CP), reinforced concrete pipe (RCP), prestressed concrete cylinder pipe (PCCP), reinforced concrete cylinder pipe, bar-wrapped steel-cylinder concrete pipe, and polymer concrete pipe (PCP).

The plastic pipes are polyvinyl chloride (PVC) pipe, polyethylene (PE) pipe, glass reinforced pipe (GRP or Fiberglass Pipe). Metallic pipes are ductile iron (DI) pipe and steel pipe. Burn et al., (2010), stated that sewer material ranges from brick, iron, vitrified clay, pitch fiber, plastic, composite materials, and concrete.

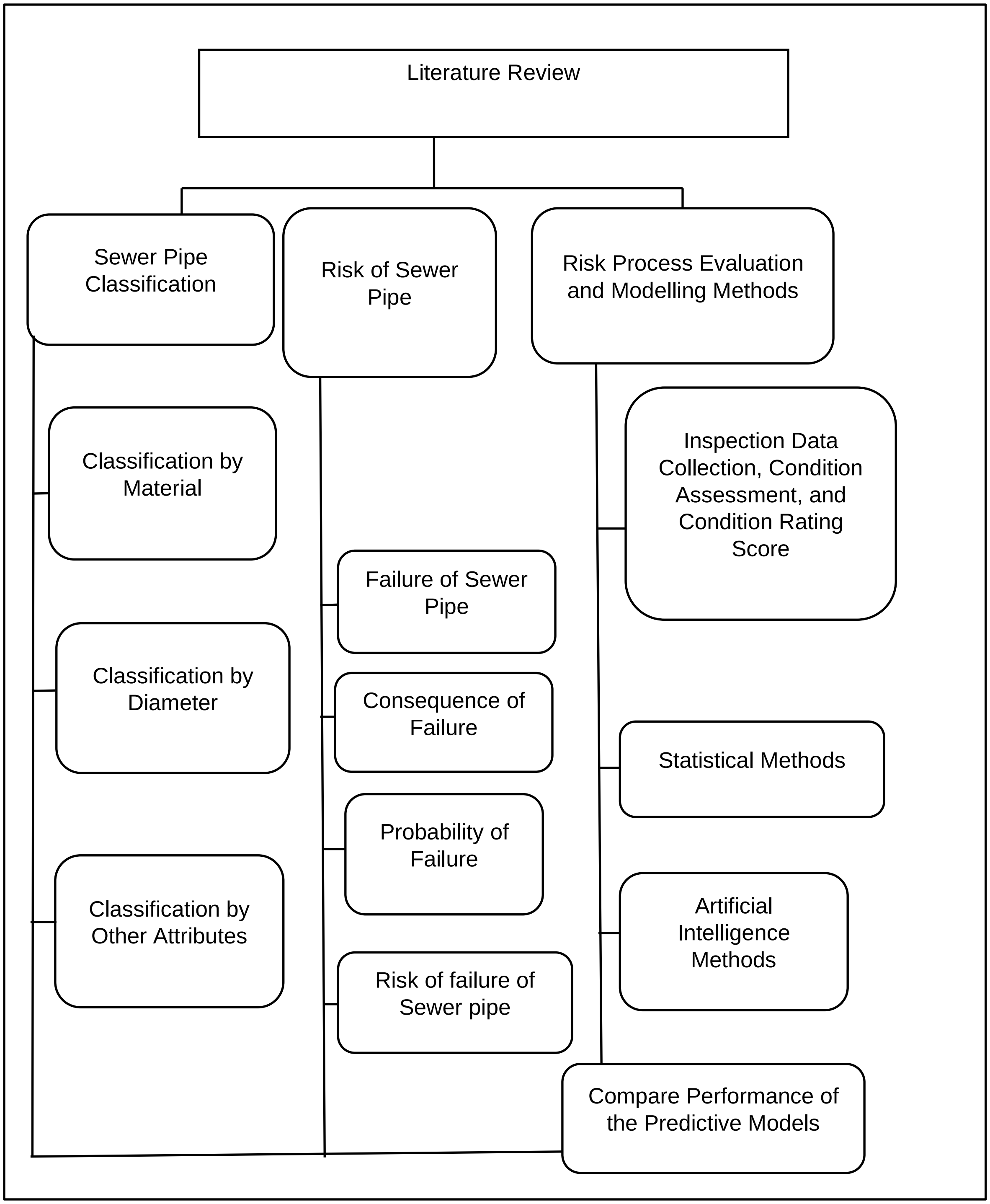

Figure 0-2 Literature Review Flow Chart

2.2.2 Sewer Pipe Diameter

Most water collection systems use a minimum of 6 in. sewer pipes while, a few, others use a minimum of 8 in. These two common sizes notwithstanding, pipes can be greater than 96 in. Pipe diameter is one of the factors affecting deterioration of sewer pipes. Malek Mohammadi et al. (2020), stated that some condition prediction models identified that sewer deterioration rate decreases with increasing diameter. “With occurrence of obstacles in the conduit, segments with small diameters are more likely to experience a hydraulic performance drop than large diameter ones” (Lubini and Fuamba, 2011)

2.2.3 Sewer Pipe Age

Pipe age is one of the pipe characteristics that is a significant variable in sewer pipes failure. Pipe age is defined as the difference between the year the pipe was installed, and the date pipe was inspected. (Lubini and Fuamba, 2011) found that pipe age was a significant parameter in the sewer systems deterioration model. Muhlbauer (2004) established that the pipe age influenced sewer pipe condition and found that poor pipe sewer condition is higher for pipes more than 50 years. “Most of the condition prediction models developed in previous studies show that pipe age has a significant relationship with deterioration of sewer pipes” (Hou et al., 2020).

2.2.4 Sewer Pipe Length

Sewer pipe length is a segment of a pipe that is measured from manhole-to-manhole. Caradot et al., 2018, noted that length is relevant in describing sewer deterioration even though it is secondary to pipe material. “Typically, longer manhole-to-manhole sewer pipe segments have higher deterioration rates because the probability of defects is greater in longer pipes” (Najafi, 2016).

2.2.5 Sewer Pipe Depth

Pipe depth is the distance from ground surface to the top of the installed pipe in the ground. Syachrani et al. (2013) , determined that pipes buried in depths between 2 m (6 ft) and 3 m (9 ft) were least connected to poor sewer pipe condition. Najafi (2016) stated that shallowly buried pipes are subjected to more defects and higher deterioration rate due to surface load, illegal connections, and tree root intrusion.

2.2.6 Sewer Pipe Slope

Slope is the gradient of pipe installed from one manhole to manhole. Lubini and Fuamba (2011), described segment slope in percentage per length of a segment. Slope will determine the velocity of flow in the sewer. Flat slope will encourage deposition of debris inside the pipe. Muhlbauer (2004), stated that negative slopes and extremely low slopes lead to debris accumulation and blockages. Laakso et al., (2018) found that negative and very low slopes were the most harmful conditions for sewer pipes, whereas steep slopes high velocities cause erosion in the pipe walls.

2.2.7 Environmental Factors

Environmental factors associated with sewer pipe failure include, but not limited to, surface condition, soil type, soil pH, and corrosivity. Surface condition is the ground surface beneath which a sewer pipe is located. Najafi (2016), stated that the location of pipe affect the magnitude of surface loading to which it is subject. The most common types of soil classified as soil texture are sand, loam, clay, and rock. Najafi and Gokhale (2005), stated that the type of soil is a factor that affects ground loss and stability of the sewer pipeline. The interaction of the soil with the sewer pipes determines the deterioration of pipes. Najafi and Gokhale (2005) found that environmental features such as soil, tree density, groundwater level show little or no influence in sewer pipe failure.

Table 0-2 Characteristics of Common Piping Materials

Adapted from Salman and Salem (2004)

Material	Application	Key Advantages	Key Disadvantages
Ductile Iron	High pressure Available sizes 4-54 in.	Good resistance to pressure surges	More expensive than reinforced concrete and fiberglass
Reinforced Concrete	Moderate pressure Available sizes 12-72 in.	Low corrosion rate	Relatively brittle, heavy, and high transportation cost
Vitrified Clay	Low Pressure for larger-diameter applications	Low thermal expansion, long life cycle, raw materials availability, corrosion resistant	High transportation cost, heavy and labor-intensive to work
PVC	Low Pressure for up to 36-inch pipe sizes	Light weight, no corrosion	Suitable for small pipe sizes and low pressure only
Reinforced Plastic Mortar	Moderate pressure available sizes up to 72 inches	Light weight, no corrosion	Expensive
HDPE	Moderate pressure available sizes of 4-63 in.	Light weight and flexible, leak-free joints	Sensitivity to temperature changes and mechanical loading

2.3 Risk

According to Najafi (2016), risk is commonly evaluated as the product of likelihood of occurrence and the impact severity of occurrence of an event.

$$\mathrm{RISK}\left(\frac{\mathrm{Consequence}}{\mathrm{Time}}\right) = \mathrm{LIKELIHOOD}\left(\frac{\mathrm{Event}}{\mathrm{Time}}\right) \times \mathrm{IMPACT}\left(\frac{\mathrm{Consenquence}}{\mathrm{Event}}\right) \quad \text{Eq. 2-1}$$

Syachrani et al. (2013), stated that risk is the probability of an event that causes a loss and the potential magnitude of that loss. Hence, Syachrani et al., (2013) expressed risk as the following mathematical relationship:

Risk = (event likelihood) x (event consequence) Eq. 2-2

NASSCO (2018) and Muhlbauer (2004), described risk as the product of the probability and consequences of failure occurrence presented in the relationship below:

$$\mathrm{Risk} = \mathrm{Probability\ x\ Consequence} \quad \text{Eq. 2-3}$$

Najafi (2016), described risk as the probability of pipe failure happening, or likelihood of a negative impact occuring. Mathematically, Najafi 2016 expressed risk as a product of Likelihood of Failure (LoF) and Consequence of Failure (CoF):

$$\text{Risk} = \text{LoF x CoF} \qquad \text{Eq. 2-4}$$

Vladeanu and Matthews (2019), posited that risk score can be calculated as a product of probability of failure and the consequence of failure. Najafi (2016), described Likelihood of failure as a curated numerical representation (rating) denoting the probability of failure based upon an asset's physical condition.

2.3.1 Failure of Sewer Pipe

According to NASSCO (2018), failure can be defined as a pipeline system not accomplishing its agreed level of service (LOS). NASSCO (2018), described pipe failure as the inability to convey flow. Pipe failure occurs when pipes are exposed to stresses. NASSCO (2018) suggested, internal pressure, soil overburden, extreme temperatures, external forces, and fatigue are examples of stresses that must be resisted by pipelines.

An asset can be consindered to have failed when it no longer achieves the required levels of service or when it is no longer providing the most cost effect means of providing the service (NASSCO, 2018)

2.3.2 Probability of Failure or Likelihood of Failure

Likelihood of failure considers two different modes of failure, physical and performance. The physical condition is similar to structural condition.

Likelihood of failure = Performance Score + Structural Score Eq. 2-5

Age of the pipe is one pipe attribute that contributes to pipe failure. Khazraeializadeh (2012), stated that classical survival function relating the age of the pipeline to the failure rate is denoted by a bathtub. As the pipes tend toward the end of their useful life the failure rate increases exponentially.

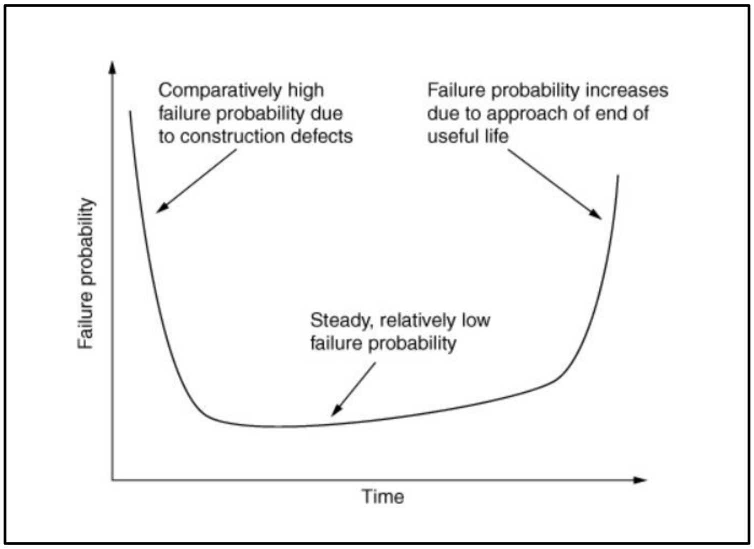

Figure 0-3 Bathtub of Pipe Performance with Age

(Najafi and Gokhale, 2005)

According to Najafi (2005), the factors that accelerate pipe aging process are pipe size, pipe section length, pipe gradient, pipe depth, surface loading and surface type, frost heave, frost load, sewage characteristics, soil-pipe interaction, pipe wall temperature gradients, corrosion, differential pipe temperature, soil type, soil pH, groundwater level, overburden pressure, temperature, and precipitation (snow or rain).

The most common predictor of estimating the probability of failure for sewer pipes is age of the asset. The bathtub illustrates three zones of pipe failure with age. The first zone is part of the curve with failure probability due to construction defects. In this portion the curve defects that developed during initial manufacture of a component cause failures. The curve levels off into the second zone. This is a constant failure rate zone. The third zone has failure rate increase. USEPA (2009), stated this is a zone where things begin to wear out as they reach the end of their useful service life.

2.3.4 Consequence of Failure

USEPA (2010), indicated that CoF is associated with cost of repair, social cost, collateral damage cost, and legal cost. Syachrani et al. 2013, specified that environmental damage, threat to public health, regulatory compliance, and impact to public relations, are among those which are hard to quantify CoF. Najafi (2016), grouped CoF into direct and indirect categories, where direct costs include property damage, damages to human health, environmental damages, loss of product, repair cost, and cleanup and remediation costs. NASSCO (2018), defined critical assets as assets with a high probability of failure and high CoF).

It is important to consider all the possible CoFs. Traditionally, CoF is always associated with cost or dollar value such as cost of repair, social cost, collateral damage cost, and legal cost (Chung, Jeong, & Syachrani, 2013). However, not every CoF component can be valued easily. According to NASSCO (2018), environmental damage, threat to public health, regulatory compliance, and impact to public relations, are among those which are hard to quantify.

According to Khazraeializadeh (2012), CoF is a combination of direct and indirect impact on the vicinity and community due to a potential asset failure. This type of impact is expressed in "Triple Bottom Line (TBL) terms. TBL goes beyond considering organization's finance but also focus on social and environmental aspects. TBL concept focusses on direct economic costs, social costs, and environmental costs with an aim of sustainability.

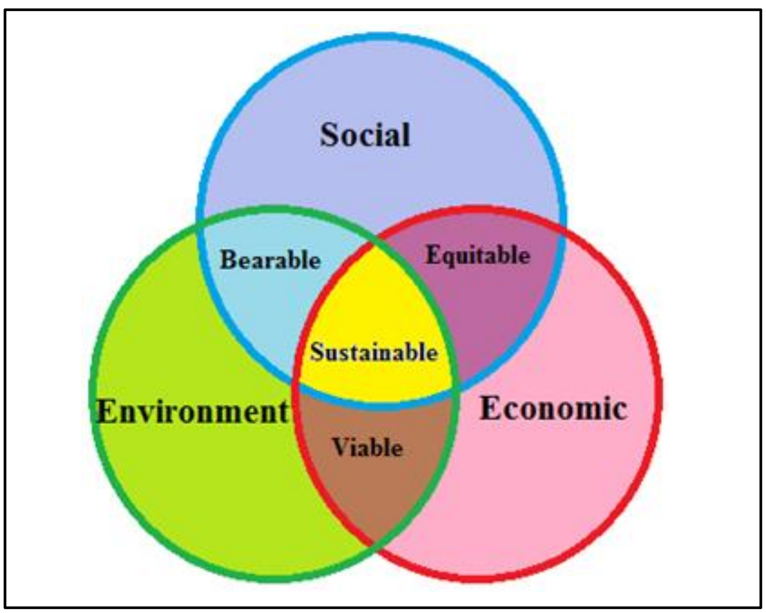

Figure 0-4 Consideration of Sustainability in Social, Environmental, and Economic Costs (Kaushal, 2019)

NASSCO (2018) identified factors that are used to determine a sewer segment's TBL COF. These factors are pipe diameter, depth of burial, location of the pipe, relative network position of the pipe, proximity to environmentally sensitive features, type of customers served, and pipe accessibility.

The three types of costs are considered when assigning weighting factors. CoF is relative to factors such as diameter of pipe, depth, relative network position, location of pipe, proximity to environmental sensitivity features, significant customer service, and accessibility for maintenance and inspection.

The features considered critical are infrastructure, water bodies, hospitals, and central business district (CBD). The features are categorized into causing either economic, social, or environmental consequences in case of failure.

The economic feature is developed based on previous emphasis on pipe diameter to assign COF. Larger size indicates higher capital, operating, and replacement costs incurred because of failure. Economic costs can be direct and/or indirect. Direct costs could include repairs, legal fees, or fines. Indirect costs could include property values, increased insurance rates

Social category is based on the cost of community impact resulting from a failure. These costs are indirect. Consequences include duration of failure, public health and safety, road closures, damage to property, and affecting critical services.

Environmental category is based on the impact to ecological conditions because of an asset failure. The costs include a violation of regulatory statures resulting in a fine, proximity to wetlands and waterways, possible contamination of water source, and sensitivity of nearby soils.

According to NASSCO, 2018, the CoF is assigned a scale from 1 to 6. Six (6) is the highest consequence and 1 is the least.

Table 0-3 Consequences of Failure Factor

(Yang and Su, 2006)

Diameter	CoF Factor	Depth	CoF Factor	Relative network position of pipe	CoF Factor	Location of pipe (Class of road)	CoF Factor
Less than 8"	1	Less than 6"	1	10 or less	1	Unpaved	1
≥ 8" - < 10"	2	≥ 6" - < 10"	2	11-30	2	Minor Local	2
≥ 10" - < 15"	3	≥ 10" - < 14"	3	31-70	3	Major Local	3
≥ 15" - < 21"	4	≥ 14" - < 18"	4	71-120	4	Collector	4
≥ 21" - < 30"	5	≥ 18" - < 24"	5	121-150	5	Arterial/Building/Pool	5
≥ 30"	6	≥ 24"	6	>150	6	Highway/Waterway	6

Table 0-4 Consequence of Failure Factor

(NASSCO, 2018)

Distance between pipe and waterway (LF)	CoF Factor	Distance between downstream pipe to a service lateral for customer with high importance (LF)	CoF Factor	Accessibility of Pipe	CoF Factor
150 LF or more	1	20,000 LF or more	1	On Right-of-way - no traffic control	1
100-150 LF	2	15,000-20,000 LF	2	On Right-of-way - with traffic control	2
75-150 LF	3	10,000-15,000 LF	3	On public lands with vehicle access	3
50-75 LF	4	5,000-10,000 LF	4	On public lands without vehicle access	4
25-50 LF	5	1,000-5,000 LF	5	On Private lands without vehicle access	5
Less than 25 LF	6	Less than 1,000 LF	6	Behind built structures and no vehicle access	6

The overall CoF score is calculated as weighted average of individual CoF. Table 2-4 illustrates CoF computation.

Table 0-5 Overall Consequence of Failure Factor

(NASSCO, 2018)

	Economic	Social	Environmental
Weighting Factor	**1/4**	**1/4**	**1/2**
	Network Position		
Diameter (12")	3	3	
Depth (9")	2		
	Location of Pipe		
Collector road	4	4	
	Proximity to Sensitive Environment		
40" from creek			5
Serves Important Customer			
2 miles downstream of hospital		3	
Accessibility			
Needs traffic control	2		
Total	**11**	**10**	**5**
Total/Possible (6*#)	11/24 = 0.458	10/18 = 0.555	5/6 = 0.833
Weighted (Total * Weighting Factor)	0.458*1/4 = 0.115	0.555*1/4 = 0.139	0.833*1/2 = 0.417
CoF = SUM (of Weighted)**/*6**			4.03

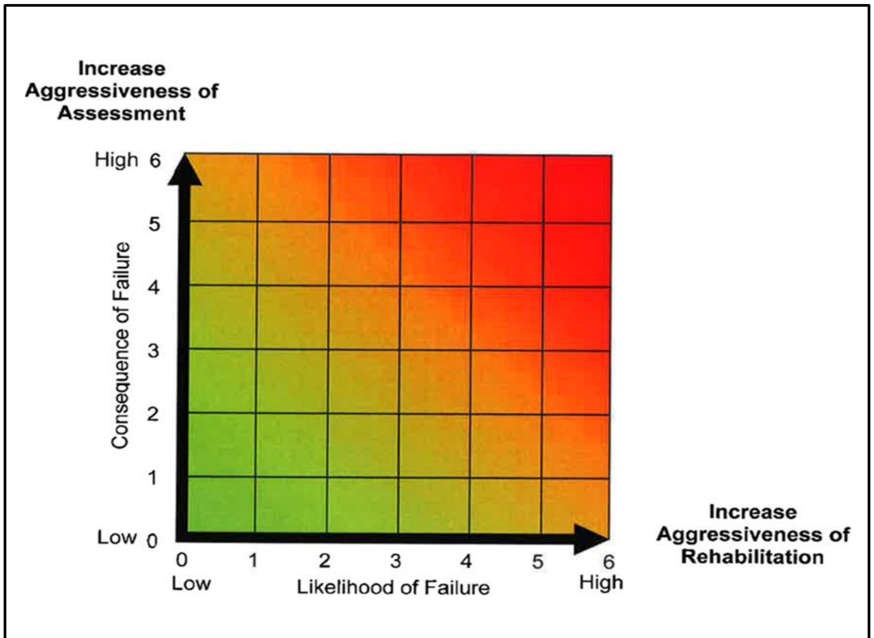

Figure 0-5 Risk Management Strategy

(NASSCO, 2018)

Figure 2-4 shows a relationship between LoF and its CoF. Pipe renewal priority has direct relationship with CoF.

2.4 Risk Process Evaluation

The process of risk evaluation includes data collection, condition assessment, and condition rating.

2.4.1 Inspection Data Collection

There is a variety of pipeline inspection methods utilized by water utilities. These methods include leak detection, electromagnetic, ultra/infra spectrum, physical, and visual. Closed Circuit Television (CCTV) camera inspection is the most common visual method used. Tscheikner-Gratl et al., (2017), in their journal article stated that pipeline management programs look at asset's condition with high sensitivity in the case of failure.

The costs associated with aftereffects on municipal utilities and its users can run high immediately, plus longstanding environmental and health concerns. "Condition monitoring of sewers by means of closed-circuit television (CCTV) inspection is a labor-intensive source of data, because inspections are performed and interpreted manually, but the condition assessment information retrieved is nevertheless one of the most important diagnostic sources when planning the renewal of sewer infrastructures" (Caradot et al., 2018).

2.4.2 Condition Assessment

Wastewater utilities operators conduct pipeline condition checks. This technique involves gathering data on the likelihood of the pipeline failure and the current structural, operational, capacity, and conditions of the pipeline. Pipeline condition assessment is a primary task in running an efficient urban asset management program because in the case of system failure, the consequences can be significant to both municipalities and users (Khazraeializadeh, 2012).

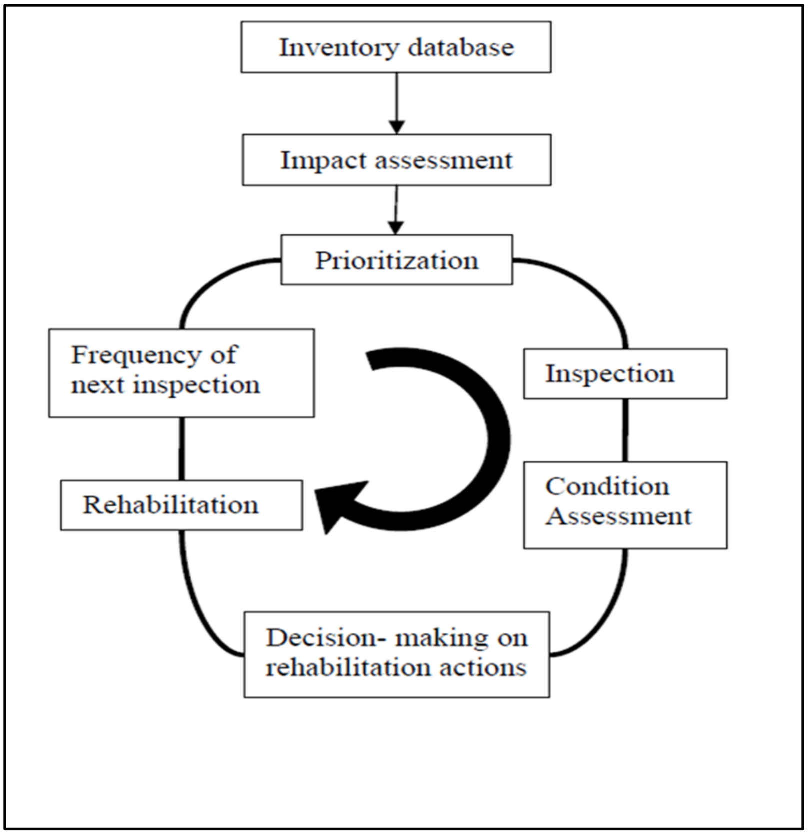

Figure 0-6 Condition Assessment Flow Chart (USEPA, 2009)

Figure 2-5 illustrates the process of condition assessment. The term condition assessment relates to establishing the existing physical condition, identifying the deterioration pattern, and determining the potential of asset collapse or failure (Khazraeializadeh, 2012).

“Understanding the deterioration trends and the remaining asset life could help in identifying underperforming assets from which decisions regarding required future expenses can be made efficiently” (Rokstad and Ugarelli, 2015). Table 2-5 summarizes condition assessment technologies and their applications.

Table 0-6 Condition Assessment Technologies and their Applications
(USEPA, 2010)

	Camera				Acoustic			Electrical & Electro-magnetic			Laser	Innovative Technologies						
	CCTV	Zoom camera	Digital scanning	Push-camera inspection	In-line leak detectors	Acoustic monitoring systems	Sonar	Electrical leak location	Remote field eddy current	Magnetic flux leakage	Laser profiling	Gamma-gamma logging	Ground penetrating radar	Infrared thermography	Micro-deflection	Impact echo/SASW	Ultrasonic pulse velocity	Guided Wave Ultrasonic
Application																		
Pipe type	G	G	G	S	G, F	F	G, F	G, F, S	G,F,S	G,F,S	G, F	G,F,S	G,F,S	G,F,S	G	G	G	F
Pipe material	Any	Any	Any	Any	Any	PCCP	Any	NF	F	F, PCCP	Any	C	Any	Any	B	B, C	C	F
Pipe diameter (in.)	>6	>6	6-60	1-12	≥4	≥18	≥12	3-60	≥2	2-56	> 4	TBD	18-30	TBD	N/A	>6	TBD	>2
Defects Detected																		
Sediment, debris, roots	•	•	•	•			•				•							
Pipe sags & deflections	•	•	Partial	•			•				•							
External pits & voids												•			Partial	•		•
Corrosion & metal loss			Partial				•		•	•	•							•
Off-set joints	•	•	Partial	•														
Pipe cracks	•	•	•	•			•	•	•	•						•	•	•
Leaks	•	•	•	•	•			•	•				•	•				
Broken pre-stressed						•			•									
Wall thickness									•							•	•	•
Service connections	•	•		•								•						
Bedding condition												•	•					
Bedding voids												•	•	•	Partial			
Deteriorated insulation														•				
Overall condition															•			

Pipe type: G – Gravity line F – Force main S – Service lateral TBD – To be decided N/A – Not applicable
Pipe material: NF – Nonferrous F – Ferrous B – Brick C – Concrete PCCP – Pre-stressed concrete cylinder pipe

2.4.3 Risk and Condition Rating

The sewer system is one element of urban infrastructure that is expected to operate without interruptions. A proactive approach to sewer asset management is of key importance for preventing uncontrolled deterioration and for reducing both direct and indirect costs (i.e. social, environmental, and third-party damage) associated with sewer failures. Existing guidelines for condition assessment and rehabilitation of sewer assets suggest that prioritization of inspections should precede the decision-making process (Berardi et al., 2009).

In a risk-based decision-making process for selecting the set of pipes to be inspected, it is necessary to consider the trade-off between economic, technical, and management criteria. Accordingly, an inspection program could be developed using a multi-objective optimization approach where decision variables are the sewer pipes to inspect, and the objectives represent distinct selection criteria. Berardi et al, 2009 concluded that inspection strategies are less effective for decision makers who must select from a large set of different feasible pipe combinations. Muhlbauer (2004), described two methods of rating a pipe asset – index overall rating and quick rating. Index overal total rating by dividing it by the total number of particular observations that contributed to the total. Another method is that of rating a buried pipe is a “quick rating,” the quantities of observations associated with the most significant ratings. An example is a quick rating of 4232. This figure is associated with Two grade-4 and two grade-3 defects, with no grade-5 defects.

According to NASSCO (2018), the PACP quick Rating expreses the number of occurrences for the two highest severity grades as shown in Figure 2-6.

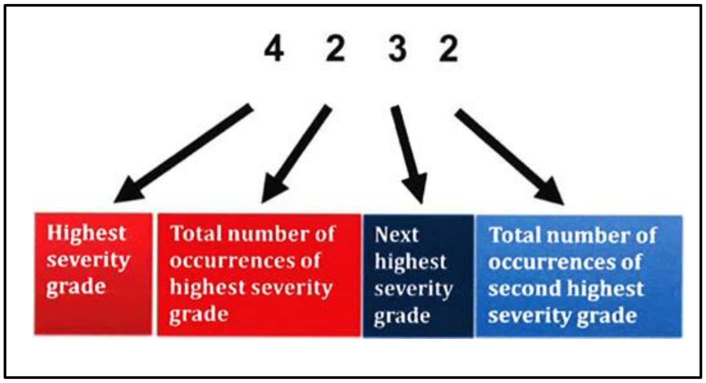

Figure 0-7 PACP Quick Rating

(NASSCO, 2018)

The first characater represents the highest severity grade occurring along the pipe length. The second character represents the number of occurrences of that highest severity grade. When the number exceeds 9, then alphabetical characters are used as follows:

10 to 14 =A, 15 to 19 = B and 20 to 24 = C etc.

According to NASSCO, 2018, the quick rating is calculated separately for Structural (OSR) and (OMR) coded defects.

Pipe segments are gven separate segment grade score. Each condition grade number is multiplied by the number of occurances in a pipe segment.

$$\mathrm{SG_N} = \text{number of grade N defects } \times \text{condition grade N} \qquad \text{Eq. 2-6}$$

For example, if a pipe has 6 structural defects of grade 5, 2 defects of grades 3 and 4 defect of grade 2, the segment grade scores are respectively SG_5 = 30, SG_3 = 6 and SG_2 = 8. The O & M defects grades are separately calculated. Overall pipe rating (OR) is calculated by adding five individual segment grade scores as in equation 2-7.

$$\mathrm{OR} = \mathrm{SG}_1 + \mathrm{SG}_2 + \mathrm{SG}_3 + \mathrm{SG}_4\,\mathrm{SG}_5 \qquad \text{Eq. 2-7}$$

For instance, the overall pipe rating is 44 for structural defects in previous example. Structural and O&M defect grades are used separately to calculate the overall pipe index for each pipe segment. Table 2-6 provides an example of overall pipe rating and segment grade score calculation.

Table 0-7 Overall Rating

Source: (NASSCO, 2018)

Overall Rating	Defects		Segment Grade	
	Structural	O&M	Structural	O&M
5	6	0	30	0
4	0	0	0	0
3	2	2	6	6
2	4	4	8	8
1	0	0	0	0
Total Defects =	**12**	**6**		
Overall Rating =			**44**	**14**

NASSCO (2018) defined Pipe Rating Index (RI) as an indicator defect severity of pipe segment. PRI is expressed in equation 2.8.

$$\mathrm{RI} = \frac{\text{Overall Rating}}{\text{Total number of defects}} \qquad \text{Eq. 2.8}$$

The pipe rating indices are calculating separately for structural and O&M conditions. For example, in previous case the $\mathrm{RI}_{\mathrm{Structural}}$ is 3.7 and $\mathrm{RI}_{\mathrm{O\&M}}$ is 2.3. A pipe in with condition 1 is in excellent condition and pipe with condition 5 pipe has failed or is likely to fail. The pipe segment that is in condition 5 needs immediate renewal.

The LoF is determined from PACP Quick rating.

Table 0-8 Common Pipe Material Defects

(USEPA, 2010)

Defect	Concrete			Ferrous		Ceramic		Plastic	
	Concrete pipe	Asbestos cement	PCCP/CCP	Cast iron/ ductile iron	Steel	VCP	Brick	PVC	HDPE
Internal pipe surface									
Root intrusion	•	•	•	•	•	•	•		•
Grease build-up	•	•	•	•		•	•	•	•
Pipe wall condition									
Cracks/ broken pipe	•	•				•			
Internal corrosion		•	•	•	•				
External corrosion			•	•	•				
Leakage									
General	•	•		•		•		•	
Joint leakage			•		•				
Leaking laterals				•					•
Alignment/grade									
Alignment				•				•	•
Joint misalignment	•	•		•		•			
Excessive deflection					•			•	•
Grade								•	•
Other	1						2	3	4

1 – Liner separation, weld failure
2 – Missing bricks, soft mortar, vertical deflection, collapse
3 – Lateral connections
4 – Pressure capacity (force mains only)
Data from Thomson et al. (2004). Reprinted with permission.

The table 2-7 summarizes the common pipe materials defects. Water utilities use National Association of Sewer Service Companies (NASSCO) Pipeline Assessment and Certificate Program (PACP) standard for defect coding and collection of data. Lubini and Fuamba (2011), stated that the PACP assigns a grade of one to five to each of the structural defects and features. These grades are defined as:

1- Excellent; 2- Good; 3- Fair; 4- Poor and 5- Immediate.

Tables 2-7 and 2-8 presents some examples of structural and operational defect codes.

Table 0-9 Structural Defects

(NASSCO, 2018)

Group	Description	Code	Structural Grade
Crack	Crack Longitudinal	CL	2
	Crack Circumferential	CC	1
	Crack Multiple	CM	3
	Crack Spiral	CS	2
Fracture	Fracture Longitudinal	FL	3
	Fracture Circumferential	FC	2
	Fracture Multiple	FM	4
	Fracture Spiral	FS	3
Broken	Broken	B	4
Hole	Hole	H	< 2 clock pcs →4,>2 clock pcs 5
Deformed	Deformed Rigid	DR	≤ 5% →4, >5% → 5,
	Deformed Flexible Bulging Round	DFBR	≤ 5% →3, >5% to ≤ 10 →4, >10% →5
Collapse	Collapse	X	5
Joint	Joint Offsite Medium Defective	JOMD	3
	Joint Offsite Large Defective	JOLD	4
Weld Failure	Weld Failure Circumference	WFC	2
	Weld Failure Longitudinal	WFL	2

Sousa et al., (2014) stressed that inspectors categorize the structural failures into 9 classes, such as open joint, displaced joint, cracked, fractured, broken, hole, collapsed, spalling, wear, and deformation.

Table 0-10 Operation and Maintenance (O&M) Defects

(NASSCO, 2018)

Description	Code	O&M Grade
Deposits Attached Encrustation	DAE	≤ 10% →2, >10% to ≤ 20 →3, >20% to ≤ 30 →4, >30% →5
Deposits Attached Grease	DAGS	≤ 10% →2, >10% to ≤ 20 →3, >20% to ≤ 30 →4, >30% →5
Deposits Attached Ragging	DAR	≤ 10% →2, >10% to ≤ 20 →3, >20% to ≤ 30 →4, >10% →5
Deposits Attached Other	DAZ	≤ 10% →2, >10% to ≤ 20 →3, >20% to ≤ 30 →4, >10% →5
Obstruction Brick or Masonry	OBB	≤ 10% →2, >10% to ≤ 20 →3, >20% to ≤ 30 →4, >30% →5
Obstruction Pipe Material in invert	OBM	≤ 10% →2, >10% to ≤ 20 →3, >20% to ≤ 30 →4, >30% →5
Obstruction intruding through Wall	OBI	≤ 10% →2, >10% to ≤ 20 →3, >20% to ≤ 30 →4, >10% →5
Obstruction External Pipe or Cable	OBP	≤ 10% →2, >10% to ≤ 20 →3, >20% to ≤ 30 →4, >10% →5
Obstruction Built into Structure	OBS	≤ 10% →2, >10% to ≤ 20 →3, >20% to ≤ 30 →4, >30% →5
Obstruction Construction Debris	OBN	≤ 10% →2, >10% to ≤ 20 →3, >20% to ≤ 30 →4, >30% →5
Obstacle/Obstruction Rocks	OBR	≤ 10% →2, >10% to ≤ 20 →3, >20% to ≤ 30 →4, >10% →5
Obstacle/Obstruction Other Objects	OBZ	≤ 10% →2, >10% to ≤ 20 →3, >20% to ≤ 30 →4, >10% →5

Yang and Su, (2006), study revealed that cracks, joint displacement, cross sectional reductions, infiltration, deposits are the most common defects in pipes.

The final condition rating is defined from two major categories which are structural and operation and maintenance (O&M). The list below presents the grades and definitions of grades respectively (Moselhi and Shehab-Eldeen, 2000):

5 - Most significant defect grades

4 - Significant defect grade

3 - Moderate defect grade

2 - Minor to moderate defect grade

1 - Minor defect grade

The pipe networks are divided into pipe sections. Pipe sections are defined as a homogeneous part of the street section with constant parameters like material, diameter, and construction year. A street section consists of several pipe sections. For each network influencing factors or criteria are assigned which represent different elements of their rehabilitation decision (e.g., condition of the pipe section). To evaluate the characteristics within the criteria, score scales between zero (0) and hundred (100) are used where 0 denotes no influence and 100 means significant influence (Lubini and Fuamba, 2011).

2.5 Sewer Pipe Condition Prediction Models

2.5.1 Introduction

Sanitary sewer pipes condition rating prediction models are utilized to determine the condition of non-inspected pipes. These assists operators of water utilities to develop renewal strategies of the pipes and to forecast the evolution of the condition of the sewer network under different investment strategies. Caradot et al. (2018), emphasized the need for model outputs to provide key information to operators and municipalities in inspection and the planning of rehabilitation budgets. Tscheikner-Gratl et al., (2014), stated that the combination of inspection data and model predictions form the analysis, which will help in identifying rehabilitation needs and support Infrastructure Asset Management (IAM) decisions. Vladeanu and Matthews (2019), stated that condition modelling at pipe level is the optimal way of supporting inspection decisions.

According to Selvakumar et al., (2015), there are 5 steps to risk management. The steps are acquiring a risk assessment model, data collection and preparation, segmentation, assessing risks, and managing risks.

Step 1: Acquire a risk assessment model: This refers to selecting commercially available models or building a required model from the available data.

Step 2: Data collection and preparation: Data preparation are the processes that result in data sets that are ready to be read into and used by the risk assessment model.

Step 3: Segmentation: Risks are assigned along the pipe sections.

Step 4: Assessing Risks: Each pipeline segment will get a unique risk score that reflects its current condition, environment, and the operating/maintenance activities.

Step 5: Managing risks: Determining appropriate actions given risk assessment results.

According to Muhlbauer (2004), there are three general types of models namely, matrix, probabilistic, and indexing models. Matrix models rank pipeline risks according to the likelihood and the potential consequences of an event by simple scale, such as high, medium, or low, or a numerical scale: from 1 to 5.

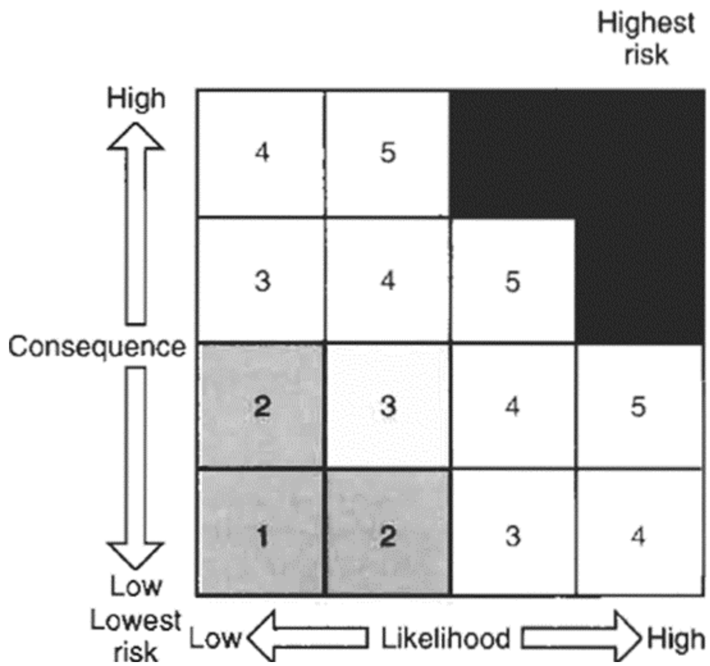

Figure 0-8 Simple Risk Matrix

(Muhlbauer, 2004)

The index matrix in Figure 2-7, numerical values (scores) are assigned to important conditions and activities on the pipeline system that contribute to the risk. Each pipeline section is scored based on all its attributes. The various pipe segments may then be ranked according to their relative risk scores to prioritize repairs, inspections, and other risk mitigating efforts. A risk assessment model can be acquired commercially or be built from the scratch. Statistical models and Artificial intelligence are the two main approaches of predicting pipe condition.

2.5.2 Statistical Methods

Statistical models establish relations between known pipe variables and the sewer pipe condition based on the condition assessment inspection data. Statistical models include discriminant analysis, logistic regression, binary regression, exponential regression, Markov, Gomptiz, and Bayesian. “The model is calibrated using maximum likelihood fitting methods to provide the best match between model predictions and recorded failure data. Goodness of fit between model forecasts and actual observations is then demonstrated by comparison with a blind data set that was not part of the calibration process” (Burn et al., 2010)

Multiple linear regression analysis allows many observed factors to affect y. The general multiple linear regression model can be written as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 + \cdots + \beta_k x_k + u \quad \text{(Wooldridge, 2013)} \qquad \text{Eq. 2-9}$$

β_0 is the intercept

β_1 is the parameter associated with x_1,

β_2 is the parameter associated with x_2, and so on

The equation below is a multiple regression where Y is a predicted outcome for individual based on (a) the Y intercept, a, the value of Y when all predictor values are 0, (b) the product of the independent variables, X_s, and the regression coefficients, b_k; and (c) the residual, ε_i

$$Y_i = a + b_1 X_1 + \cdots + b_k X_k + \varepsilon_i \quad \text{(Lomax \& Hahs-Vaughn, 2012)} \qquad \text{Eq. 2-10}$$

Odds and Logit

$$\mathrm{Odds}(Y = 1) = \frac{P(Y=1)}{1-P(Y=1)} \text{ Eq. 2-11} \qquad \text{Eq. 2-11}$$

Log odds

$$\ln\frac{P(Y=1)}{1-P(Y=1)} = \mathrm{Logit}(Y) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m \qquad \text{Eq. 2-12}$$

Lubini and Fuamba (2011), used age, diameter, length, slope, and material and built a logistic regression model as follows.

$$Y^* = \propto_0 + \beta_1 \times \mathrm{Age} + \beta_2 \times \mathrm{diameter} + \beta_3 \times \mathrm{length} + \beta_4 \times \mathrm{slope} + \beta_5 \times \mathrm{material} + \varepsilon \qquad \text{Eq. 2-13}$$

Where Y^* is the unobservable conduit condition, α_0 is the threshold β_1…. β_5 are regressor coefficients.

2.5.3 Artificial Intelligence System

An artificial intelligence system (ANN) is one of the modeling techniques of artificial intelligence modeling techniques. According to Elmasry et al., (2017), some emerging techniques for artificial intelligence systems seek to make better use of human reasoning to solve problems involving incomplete knowledge and use of descriptive terms. ANN predict output from input information in a manner that simulates the operation of the human central nervous system (Syachrani et al., 2013). Burn et al., (2010), stated that ANN is being increasingly used to solve complex problems, it is also often treated as 'black box' solution. Burn et al., (2010) stated that ANN has layers of nodes, which provide a functional relationship between input information and predicted output. The layers are trained on historical data sets. These data sets demonstrate the actual relationship between input and output information.

According to Ward and Savić (2012), ANN can learn the patterns of the underlying process from past data, capturing the relations between the inputs and the outputs. According to Peponi et al., 2019, ANN is a set of independent neurons linked together in the same way as the synapses, neurons, and dendrites of our brain (Rokstad and Ugarelli, 2015). The neural network learn and execute tasks. During training, the network modifies the weights of the links among the neurons in a way that each input produces the expected outputs. The output is the dependent variable, and the inputs are independent variables.

Machine learning is another type of artificial intelligence modeling technique. Machine Learning include Support Vector Machine, Decision Trees, Random Forest, and Bayesian Forest.

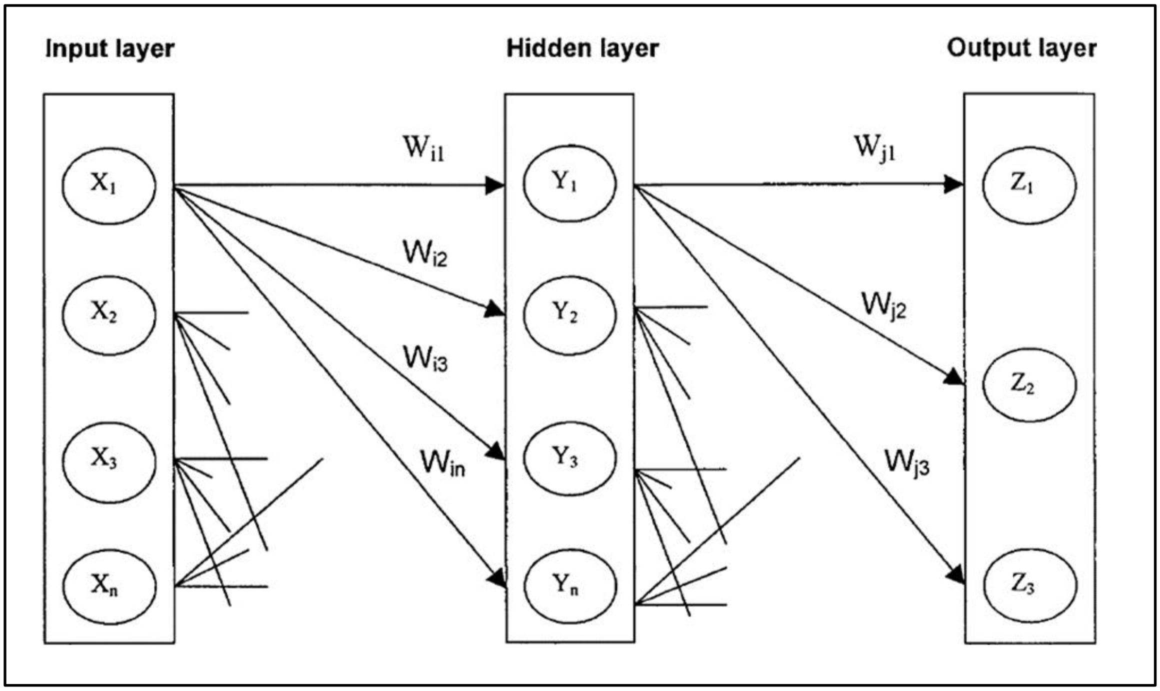

Figure 0-9 General Model for Back-propagation Neural Network

adapted from (Yang and Su, 2006)

Figure 2-8 shows that ANN has input layer, an output layer, and one or more hidden layers. These layers are connected to each other by weighted connections. The weights associated with these connections (i.e., Wi1 and Wj1) are calculated through the training process and generally represent the network's state of knowledge (Moselhi and Shehab-Eldeen 2000).

In this study, ANN will be developed and compared with the logistic regression.

Table 0-11 Advantages and Limitations of Sewer Deterioration Models

(Adopted from Salman and Salem, 2004)

Method	Advantages	Limitations
Multiple linear regression	• Simple method • Linear regression equations can be generated by using common spreadsheet applications.	• Validity of the model depends on satisfying several assumptions. • The functional relationship between condition rating and independent variables is assumed to be linear.
Binary logistic regression	• Probability of a pipe segment to be in a deficient state can be obtained.	• This method is applicable for identifying the odds ratio associated with dichotomous (such as fail – not fail) variables.
Ordinal regression	• Probability of a pipe segment to be in different condition states can be obtained based on the relationship between ordinal dependent variable and independent variables.	• Poor fit to the observed condition ratings may be obtained if a cross-sectional dataset is used. • Proportionality of odds assumption must be satisfied.
Markov chains	• Probability of a pipe segment to be in different condition states can be obtained. • Transition matrices can be generated based on experience of the personnel.	• Markov chains assumes the deterioration rate to be time-independent (unless different transition matrices are used for different time zones). • Dataset should be divided into cohorts and a new Markov-chain deterioration curve must be generated for each cohort unless ordinal regression is used to estimate transition probability values.
Artificial neural networks	• Exact functional form does not have to be identified beforehand. • Complex nonlinear relationships can be modeled.	• An extensive dataset is required for the model to learn all possible combinations.
Survival functions	• Allows the user to analyze the percentages of pipe segments in each condition state with respect to time.	• Cohorts of sewer pipes must be generated.
Simulation	• Condition ratings and confidence levels can be estimated based on limited data.	• Data points with limited information are assumed to have the same deterioration pattern.

Table 0-12 Summary of Literature Review with Models

Author/References	Title	Methodology	Variables	Results	Critique
Ugarelli et al., (2013)	Modeling of deterioration timeline of sewer systems	Multinomial logistic regression	Age, diameter, length, slope, material, and condition	Predicted 54% of outcomes for transition probabilities involving both condition classes 1 (excellent) and 2 (good), and it showed a 59% correct prediction for class 3 condition (fair).	Less than 100 sewer segments were considered. Larger data sample is necessary to obtain a better realistic model. Soil type not considered
Tade et al., (2019)	Practical benchmarking of statistical and machine learning models for predicting the condition of sewer pipes in Berlin, Germany	GompitZ, Markov-chains, and machine learning (Random Forest)	Age, material, effluent type, width, depth, District/location	Pipe condition rates	At pipe level Random Forest performs better than GompitZ The main weakness of the Random Forest model lies in its high False Positive rate Another weakness of the Random Forest model is the lack of physical information about pipe deterioration in the model's structure.
Sousa et al., (2014)	Adaptation of sewer networks using integrated rehabilitation management	Binary logistic regression	Shape, age, diameter, slope, and length, costs	Developed logistic regression model Relationships graphs of costs vs pipe diameter sensitivity vs specificity, and sensitivity vs probability cutoff.	The study considered climate change scenarios, used storm water and flooding. The models did not consider the consequences of failure. The data utilized is not large enough.
Shehab-Eldeen et al., (2001)	Consequence-of-Failure Model for Risk-Based Asset Management of Wastewater Pipes Using AHP	Analytic Hierarchy Process (AHP) Expert Opinion	Pipe age, pipe diameter, pie length, depth, access to pipe, distance to critical laterals, soil type, seismic zone, other infrastructure, water body and land use	Weighted risk scores and ranking	Not exhaustive enough for certain utilities to use. Requires more questionnaires and more subject matter experts. only 6 experts responded to the questionnaire. O&M hydraulic factors were not considered.

Table 0-13 Summary of Literature Review with Models

Author/References	Title	Methodology	Variables	Results	Critique
Laakso et al., (2018)	Risk assessment of sewer condition using artificial intelligence tools: application to the SANEST sewer system	ANN and support vector machines (SVM)	Material cohorts, diameter, depth, length, slope, and age	Tables of predicted and observed - operation & structural.	All models revealed more limited accuracy for estimating the sewers in worse condition.
Elmasry et al., (2017)	A decision support system for rehabilitation of sewer pipes	Database management system (DBMS) and decision support system (DSS)	Defect, structural, diameter of pipe, bends, hydraulic, original pipe, distance, duration, by-pass, future settlements, years in business and length, life, local suppliers, access, services, and cost	Calculated utility values.	The model overestimates condition ratings for good conditions and underestimates condition ratings for poor conditions.
Syachrani et al., (2013)	Modelling asset lifetimes and their role in asset management	Ordinal regression, Markov chain, Bayesian, monte Carlo, and simulation Artificial neural networks.	Pipe diameter, soil type, and pipe age	Relationships of LoF Vs CoF. Approaches of predicting asset failure. Practical Paper with recommendations.	Data collected was not large enough.
Ward and Savić (2012)	Sewer condition prediction and analysis of explanatory factors	Binary logistic regression, multinomial logistic regression, ordinal regression, random forest, and Boruta Algorithm	Installation depth, material type, age, pipe Diameter, length, and pipe slope	Random forest attained accuracy 62% when false negative rate (FNR) is 20%. binary logistic regression accuracy 56 % when FNR is 20%.	Did not consider hydraulic factors such as cracks, roots intrusion, and deformation and surface defects.

Table 0-14 Summary of Literature Review with Models

Author/References	Title	Methodology	Variables	Results	Critique
Rokstad and Ugarelli (2015)	Defect based deterioration model for sewer pipelines using Bayesian belief networks	Dynamic Bayesian network (DBN) and Monte Carlo simulation (MCS)	Age, size, material, diameter, length, street category, and effluent Inspection: Types of defects, structural condition rating, operational condition rating, overall condition rating	Structural, operational, and overall condition ratings were determined using the developed BBN model. Mean absolute error (MAE) and root mean square error (RMSE) were calculated between the predicted condition rating resulting from the BBN model and the actual ones.	The model overestimates the structural and overall condition ratings for good conditions and underestimates the structural and overall condition ratings for poor conditions
Yang and Su (2006)	Advanced criticality assessment method for sewer pipeline assets.	Real age prediction model Delphi workshop method, Probabilities of failure (PoF), Consequences of Failure (CoF)	Length, root, sludge, debris, diameter, slope, location, and Soil type.	Probability of failure (PoF) and Consequence of failure (CoFs) were generated. Criticality was measured as the combined effects of the event likelihood and the event consequence.	Data collected was not large enough. Location (land use) and capacity (pipe size) two components of consequence of failure were considered. Consequence of failure factors, such as environmental impacts, were incorporated. The study did not quantify the risk into monetized values.
Berardi et al., (2009)	A multi-objective optimization model for sewer rehabilitation considering critical risk of failure	Optimization model	Age, material, diameter, and criticality	Arrays of solutions are presented as Pareto optimal trade-off curves. Two non-monetary based objective functions were presented.	Optimizing the model globally is not guaranteed.

Table 0-15 Summary of Literature Review with Models

Author/References	Title	Methodology	Variables	Results	Critique
Tade et al., (2019)	Automation model of sewerage rehabilitation planning	Sewer inspection, assessment of structural conditions, computational of structural condition grades, determination of rehabilitation methods and substitution of materials	Sewer inspections, structural conditions of pipe sections, failure lengths, and rehabilitation methods and substitution materials	Adopt the appropriate rehabilitation methods and substitution materials for the failure pipes	Large amount of data will be required before such an automation model of planning optimal sewerage rehabilitation strategies can be broadly applied to other sewer systems.
Berardi et al., (2009)	An effective multi-objective approach to prioritization sewer pipe inspection	Markov chain model	Diameter, age, shape, material, length, gradient, cover depth, proximity to important locations, traffic load, and soil type	The statistical and artificial intelligence models predicted condition of sanitary sewer pipes with more than 80% accuracy	Limitations in terms of search effectiveness and efficiency.
Ugarelli et al., (2013)	Wastewater pipes in Oslo: from condition monitoring to rehabilitation planning	GompitZ tool model	Age, material, length	Obtained 4 scenarios with low and High consequences	Only emphasized data collection
Tade et al., (2019)	Modified sewer asset management to accommodate London's future sustainable development	Deterioration models (MDs)	Depth, location, length, size, slope, effluent type, and soil type.	Deterioration model for concrete sewer **Y = 9.169 x ln(x) +C**	This paper did not determine the risk values. The model only considered concrete sewer pipes cohorts.

Table 0-16 Summary of Literature Review with Models

Author/References	Title	Methodology	Variables	Results	Critique
Burn et al., (2010)	Innovative research program on the renewal of aging water infrastructure systems	GIS based genetic algorithm	Diameter, length, Accessibility	Large and small diameter sewers in two cities that were in excellent condition after being in use for 25, 23, 21, and 5 years, respectively	Data on current condition is not big enough.
Ugarelli et al., (2013)	Evaluating the role of deterioration models for condition assessment of sewers	GompitZ (non-homogenous Markov Chain) statistical model and Random Forest (RF) model	Pipe diameter, Age, Road Traffic, bedding soil	GompitZ outperformed RF on individual pipe predictions RF outperformed GompitZ network-level predictions with respect to predictive accuracy	Model predictions had lower accuracy than an uninformed estimate of the distribution of conditions (the selection variance).

2.6 Chapter Summary

This chapter reviewed the existing literature on sanitary sewer pipes asset condition. In this chapter, sewer pipes classification by material, diameter, and other attributes were discussed. The failure of sewer pipe and risk process evaluation were discussed. The risk process evaluation includes inspection data collection, condition assessment, risk, and condition rating. NASSCO, PACP quick rating protocol was discussed. Sewer pipe condition prediction methods steps acquiring a risk model were reviewed. The steps of risk assessment model included, data collection and preparation, segmentation, assessing risks, and managing risks. In addition, the risk matrix and logistic regression statistical and ANN artificial intelligence methods were discussed. The literature review indicated that MLR and ANN models can be developed to predict sanitary sewer pipes condition using inspection and condition assessment data.

Chapter 3

Logistic Regression and Neural Network Methods

3.1 Introduction

In this chapter, statistical logistic and multiple linear regressions are discussed. Artificial Intelligence ANN method is discussed. An explanation of ANN training, validation, and testing procedure and performance of the model is discussed. Lastly, summary of this chapter is described.

3.*2* Logistic Regression Method

Regression methods describe relationships between independent and dependent variables. Lomax and Hahs-Vaughn (2012), described independent variable as a predictor and dependent variable a criterion variable. The common regression methods are simple linear regression, multiple linear regression, and Logistic regression. These methods are represented in form of equations described as models. Equation 3-1, below, describes a simple linear regression.

$$Y_i = b_{YX}X_i + a_{YX} + e_i \qquad \text{Eq. 3-1}$$

Where:

Y is the criterion variable

X is the predictor variable

b_{yx} is the population slope for Y predicted by X

a_{yx} is the population intercept for Y predicted by X

e_i are sample residuals or errors of prediction

i represents an index for a case

Equation 3-2 represents multiple linear relationship of independent and dependent variables according to (Hawari et al. 2020) and (Mohammadi et al. 2019)

$$Y = a + \beta_1 x_1 + \beta_2 x_2 \ldots\ldots + \beta_n x_n + \varepsilon \qquad \text{Eq. 3-2}$$

Where:

Y = dependent variable

a = intercept

$\beta_1 \ldots . \beta_n$ = slope or coefficient

N = number of observations.

x = independent variables in the datasets

Simple and multiple regression models demonstrate relationship between one or more predictors or variables when the outcome is continuous. The relationship of independent variables and dependent variable is linear. Binary or dichotomous outcome variable in logistic regression distinguishes logistic model from linear regression model (Hosmer Jr et al., 2013).

A logistic model describes the relationship between an outcome (i.e., dependent or response) and a set of prediction (i.e., independent, or explanatory) variables, often referred to as covariates (Khashei and Bijari, 2010). Equation 3-3 represents logistic regression model according to (Hawari et al., 2020)

$$\log\left[\frac{\pi}{1-\pi}\right]Y = \left[\frac{p(y=1|x_1 \ldots x_n)}{1-p(y=1|x_1 \ldots x_n)}\right] = a + \beta_1 x_1 + \beta_2 x_2 \ldots\ldots + \beta_p x_p \qquad \text{Eq. 3-3}$$

Where:

Y = dependent variable

a = intercept parameter

β_p = regression coefficients associated with p independent variables.

Probability of (y =1) determined using exponential transformation.

$\pi = p(y = 1|x_1 \ldots . x_n)$

In this model new values of Y can be forecasted with new observed values of X.

Equation 3-3 shows general function of binary logistic regression. π represents Pr(Y=1) meaning probability associated with outcome of condition 1. Consequently, 1- π represents Pr(Y=0) meaning probability of outcome of condition 0. π/(1-π) means the odds of having (Y=1).

Multinomial logistic regression is an extension of binary logistic regression and can be used when dependent variable is categorical and has more than two levels (Hawari et al. 2020).

$$\ln\left[\frac{\Pr(Y=i|x_1 \ldots x_n)}{\Pr(Y=k|x_1 \ldots x_n)}\right] = \beta_0 + \beta_{i1}x_1 + \beta_{i2}x_2 \ldots\ldots + \beta_{in}x_n \qquad \text{Eq. 3-4}$$

where,

i = 1, 2, …, k – 1 correspond to categories of the dependent variable,

xs are independent variables,

n is the number of independent variables,

β_0 is the intercept for category i,

β_{is} are the regression coefficients of independent variables defined for each category i.

Assuming three sewer pipe conditions, equations 3.5 and 3.6 represents multinomial logistic regression for a pipe system with three condition levels 0, 1 and 2. Category zero (0) is used as the reference value. The model is developed with logit functions. To develop the model, p covariate and a constant term denoted by the vector x (Hosmer et al., 2013).

$$g_1(x) = \ln\left[\frac{\Pr(Y=1|x)}{\Pr(Y=0|X)}\right] = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 \ldots\ldots + \beta_{1p}x_p = x'\beta_1 \quad \text{Eq. 3-5}$$

$$g_2(x) = \ln\left[\frac{\Pr(y=2|x)}{\Pr(y=0|X)}\right] = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 \ldots\ldots + \beta_{2p}x_p = x'\beta_2 \quad \text{Eq. 3-6}$$

$$\Pr(Y=0|x) = \frac{1}{1+e^{g_1(x)}+e^{g_2(x)\prime}} \quad \text{Eq. 3-7}$$

$$\Pr(Y=1|x) = \frac{e^{g_1(x)}}{1+e^{g_1(x)}+e^{g_2(x)\prime}} \quad \text{Eq. 3-8}$$

$$\Pr(Y=2|x) = \frac{e^{g_2(x)}}{1+e^{g_1(x)}+e^{g_2(x)\prime}} \quad \text{Eq. 3-9}$$

Using the convention for the binary model, $\pi_j(x) = \Pr(Y = j|x)$ for j = 0, 1, 2.

3.2.1 Regression Assumptions Test

The commonly regression assumptions tested are independence, homogeneity, linearity, normality, multicollinearity, no outliers, and homoscedasticity. The tests are conducted by checking for symmetry in a histogram, frequency distribution, boxplot, or skewness and kurtosis statistics. of the assumptions are fixed by transformation. Transformation is used to correct for non-normality in regression analysis by transforming the dependent variable using log (to correct for positive skew) or the square root (to correct for positive skew) (Chughtai et al. 2008 and Ariaratnam et al., 2001). The most common form of transformation is log-transformation.

Skewness and Kurtosis test normality of data. Skewness measures the direction and degree of asymmetry of the data. Values between -3 and +3 are typical values of samples from a normal distribution (Young, 2017). The equation 3-10 presents skewness.

$$g = \frac{m_3}{m_2^{\frac{3}{2}}} \quad \text{Eq. 3-10}$$

Where m is central moment of the residuals.

Kurtosis is another moment estimator that measures the heaviness of the tails of a distribution. If the kurtosis statistic equals 3 and the skewness is o, then the distribution is normal (Young, 2017). Kurrtosis is presented in the equation 3-11.

$$g = \frac{m_3}{m_2^2} \qquad \text{Eq. 3-11}$$

3.2.2 Test of Significance of the Coefficients of Logistic Models.

The logistic regression model was checked for statistical validity utilizing change in Log Likelihood, Hosmer-Lemeshow Goodness-of-fit Test, Pseudo-variance explained coefficient of multiple determination (R^2), and Wald Test. Lomax and Hahs-Vaughn (2012) , described Log Likelihood as a ratio test that is like F test and that it tests the null hypothesis. The hypothesis test is that all the regression coefficients are equal to zero. Richard G. Lomax and Debbie L. Hahs-Vaughn, 2012, denoted the null hypothesis and alternative hypotheses as follows:

H_0: $\beta_1 = \beta_2 = \ldots = \beta_m = 0$

H_1: H_0 is false

The equation for the likelihood-ratio test statistic is denoted by equation 3-12 (Kulandaivel, 2004)

$$-2\log\left(\frac{L_0}{L_1}\right) = (-2\log L_0) - (-2\log L_1) \qquad \text{Eq. 3-12}$$

A chi-square test is produced with degrees of freedom equal to the difference in degrees of freedom of the models by multiplying the loglikelihood by -2 as shown in equation 3-13.

$$X^2 = -2(LL_{model} - LL_{baseline}) \qquad \text{Eq. 3-13}$$

There is a possibility of making type I and II errors. (Young, 2017), described Type I error as when a null hypothesis is rejected when it is true and Type II as when there is failure to reject null hypothesis when is false. Significance testing measures the strength of hypothesis (H_0) with probability (the p-value). In this study a significance is set to P-value ≤ 0.05 (95% confidence level).

Hypothesis test is important. The null hypothesis (H_0) assumes that coefficients β_0, $\beta_1 \ldots \beta_p$-1 are zero. The alternative hypothesis (Ha) assumes that not all the coefficients are equal to zero.

The likelihood function is used in the comparison of the observed and predicted values as shown in equation 3-14 (Hosmer Jr et al., 2013).

$$D = -2ln\left[\frac{(likelihood\ of\ the\ fitted\ model)}{(likelihood\ of\ the\ saturated\ model)}\right] \qquad \text{Eq. 3-14}$$

D is the likelihood ratio.

The likelihood ratio test is presented in equation 3-15.

$$D = -2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1 - y_i)\ln\left(\frac{1-\hat{\pi}_i}{1-y_i}\right)\right] \qquad \text{Eq. 3-15}$$

Where $\hat{\pi}_i = \hat{\pi}(x_i)$

The statistic D is called deviance and plays a role of residual sum-of-squares (SSE).

The error sum of square is

$$SSE = \sum_{i=1}^{n}(y_i - \grave{y})^2 \qquad \text{Eq. 3-16}$$

SSE is the sum of squared residuals.

This follows the definition of a saturated model that Where $\hat{\pi}_i = y_i$ and the likelihood is (Hosmer W. D et al., 2013).

$$l(\text{saturated model}) = \prod_{i=1}^{n} y_i^{yi} \times (1 - y_i)^{1-y_i} = 1.0 \qquad \text{Eq. 3-17}$$

This implies that the deviance is:

$$D = -2ln(likelihood\ of\ the\ fitted\ model) \qquad \text{Eq. 3-18}$$

Hosmer-Lemeshow Goodness-of-Fit-Test is a tool that is used to examine the overall model fit (Lomax and Hahs-Vaughn, 2012). This is a chi-square test that is not statistically significant. A model is accepted when there is non-statistical results of Hosmer-Lemeshow test.

Pseudo-variance explained is another overall model fit index for logistic regression. Cox and Snell, Nagelkerke; Hosmer and Lameshow; and traditional R^2 are the R^2 pseudo-variance explained values computed in logistic regression (Lomax and Hahs-Vaughn, 2012).

The Cox and Snell R^2 is computed in as in equation 3-19 (Lomax and Hahs-Vaughn, 2012).

$$R_{CS}^2 = 1 - \left(\frac{LL_{baseline}}{LL_{model}}\right)^{\frac{2}{n}} \qquad \text{Eq. 3-19}$$

To achieve a maximum value 1, Nageklkerke adjusts the Cox and Snell value as in Equation 3-20.

$$R_N^2 = \frac{R_{CS}^2}{1-(LL_{baseline})^{\frac{2}{n}}} \qquad \text{Eq. 3-20}$$

Hosmer-Lemeshow R^2 is computed as in equation 3-21 (Lomax and Hahs-Vaughn, 2012).

$$R_L^2 = \frac{-2LL_{model}}{-2LL_{baseline}} \qquad \text{Eq. 3-21}$$

Adamowski et al., (2012), described R^2 a measure of proportional variation in the structural and operational condition explained by a sewer's attributes. Young (2017), explained that R^2 is highly sensitive to sample size, its value is increased by adding more predictors, and magnitude of the slopes is not measured by R^2. Pearson Correlation was used to explain the strength of the relationship between variables. This is a way of identifying multi-collinearity. Pearson Correlation value indicates the strength of linear association between two independent variables. It describes the degree and direction of relationship of two variables. The coefficient of correlation is usually denoted as R and it ranges between $-1 < R < 1$. The plus and minus signs indicate the direction of the relation. The closer the value of coefficient to -1.00 or +1.00 the stronger the correlation between two variables (Hinton et al., 2014). The closer the value of R to 0, the weaker the linear relationship between the variables. Usually, if

$$0.0 \leq |R| \leq 0.3\text{: weak correlation} \qquad \text{Eq. 3-22}$$

$$0.3 \leq |R| \leq 0.7\text{: Moderate Correlation} \qquad \text{Eq. 3-23}$$

$$0.7 \leq |R| \leq 1\text{: Strong Correlation} \qquad \text{Eq. 3-24}$$

Wald Test is another test of significance of coefficients of variables in logistic regression. Wald test is the same as when using t-test in linear regression and is a ratio for maximum likelihood as shown in equation 3-25 (Young, 2017).

$$Z = \frac{\widehat{\beta_j}}{\widehat{s.e.\beta_j}} \qquad \text{Eq. 3-25}$$

where β_j is the coefficient of the predictor variable, and s.e. is the standard error of the coefficient. Wald is used to test the hypothesis H0: βj = 0. If βj = 0 for independent variables, the variables are not significant and should be removed from model. Alternatively, if Wald is not zero, the variables should be included in the model.

3.2.3 Classification Tables

In logistic regression, classification table is used to show the accuracy of the model prediction. This presented inform of the percent correct predictions. The classification table shows the actual or observed values versus the predicted values. Classification is the goal for the model (Hosmer et al., 2013).

3.2.4 Area Under the Receiver Operating Characteristic Curve (ROC)

The area under ROC has become the standard that evaluates the ability of the model to assign a higher probability to the outcome. ROC plots the probability of detecting true signal (Sensitivity) and false signal (1-specificity) (Hosmer et al., 2013). According to Hosmer et at, 2013, the area under the ROC curve is described as in the Table 3-1. Figure 3-1 shows the ROC.

Table 3-17 ROC Curve

ROC = 0.5	Suggests no discrimination
0.5 < ROC = 0.7	Poor discrimination
0.7 ≤ ROC < 0.8	Acceptable discrimination
0.8 ≤ ROC < 0.9	Excellent discrimination
ROC ≥ 0.9	Outstanding discrimination

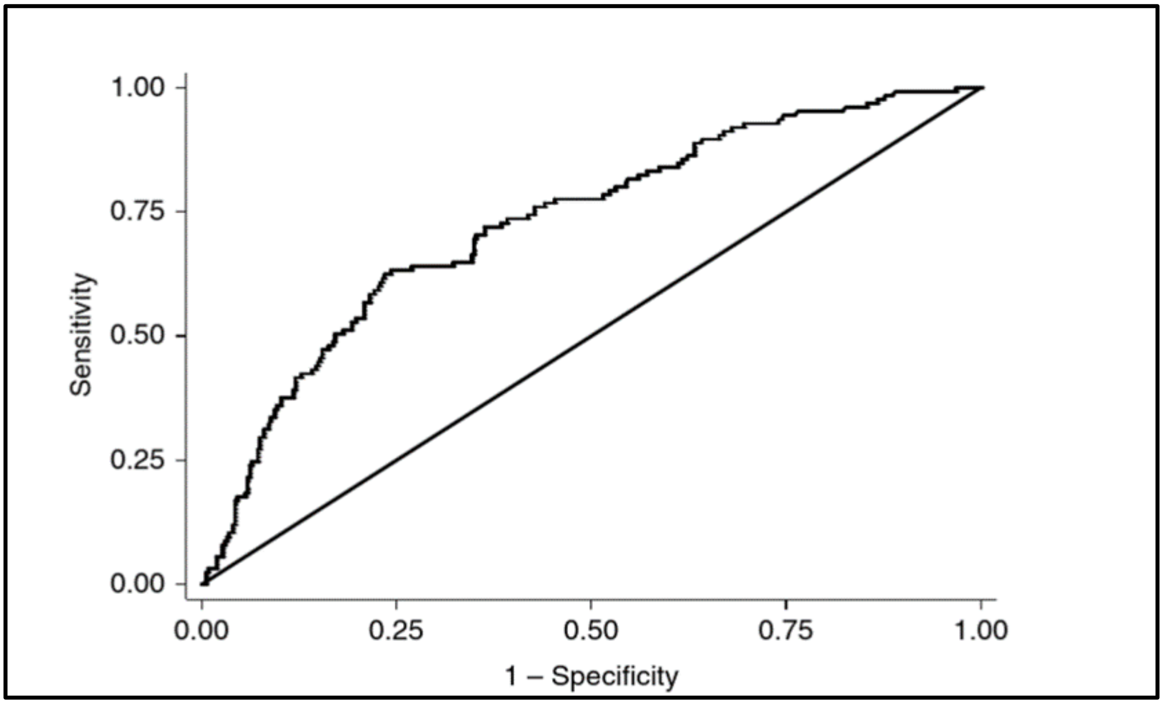

Figure 3-10 Sensitivity and 1-Specificity

(Hosmer et al., 2013)

3.3 Artificial Neural Network Method

ANN method is one of the machine learning intelligence techniques used to develop problem solving model. ANN is a computational model that is inspired by the biological nervous system (Chakraverty and Mall, 2017). The ANN is a structure with neurons or nodes interconnected in layers that are interconnected ANN method demonstrates a relationship between the input and output variables. ANN learns by processing information through neurons. The interneuron connections are assigned “weights”. The ANN structure is comprised of the input layer, hidden layer, and output layer. It is in the hidden layer where the processing takes place. This process is like a black box. ANN is used to analyze data with nonlinear relations. Adamowski et al., (2012), described ANN as data-driven process that has a flexible mathematical algorithm which can solve complex nonlinear relationships between input and output data sets.

3.3.1 Artificial Neural Network Architecture

The structure of ANN is comprised of three layers namely, input, hidden, and output layers. The input layer is where the data is entered in the model and computation is conducted. While data are processed in the hidden layer, results of ANN are produced in the output layer. Each layer consists of neurons as basic elements. A neural network can be described as a network of simple processing nodes or neurons, interconnected to each other in a specific order, performing simple numerical manipulations. There is number of neurons in the input, hidden and output layers. In this study, three-layer feed-forward neural networks with back propagation (BP) learning were constructed for computation of eleven physical and environmental input variables as shown in Figure 3-2.

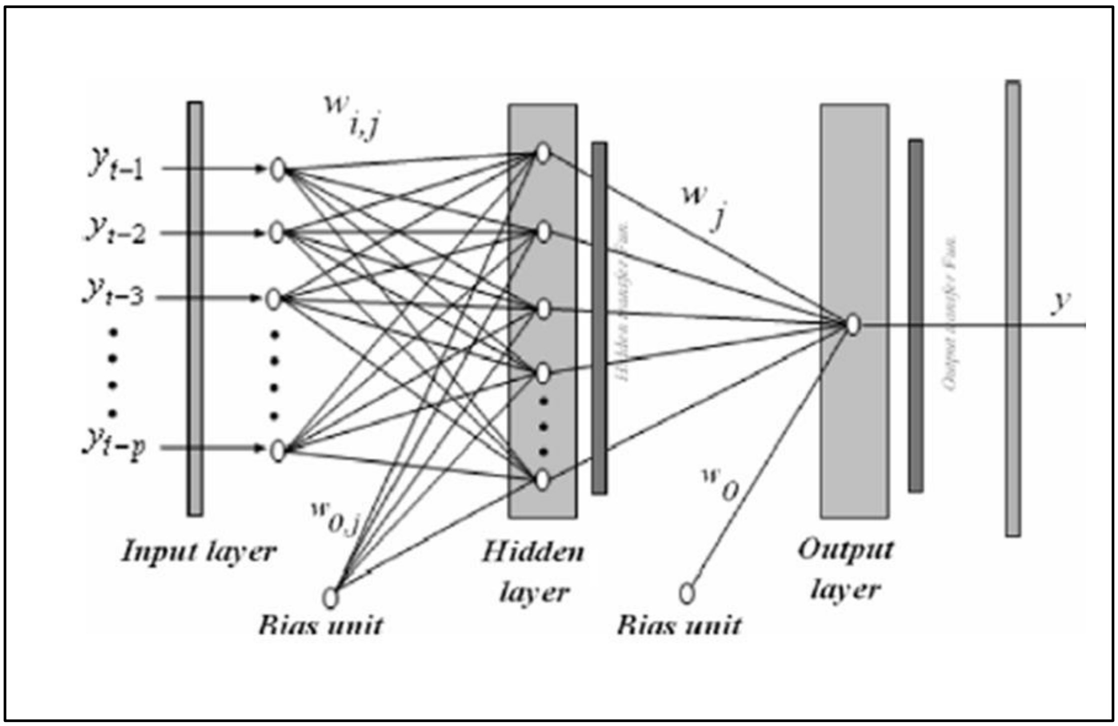

Figure 3-11 Neural Network Structure

(Chughtai et al., 2008)

Figure 3-2 shows the algorithm having nodes and network lines. The lines are assigned weights of the connecting nodes. Equation 3-26 shows a relationship between the input and output variables.

$$y_t = w_0 + \sum_{j=1}^{q} w_j \cdot g\left(w_{0j} + \sum_{i=1}^{p} w_{ij}\, y_{t-i}\right) + \varepsilon_t \qquad \text{Eq. 3-26}$$

Where, w_{ij}(i=0,1,2,…..p,j =1,2……q) and w_j(j = 0,1,2,…..,q) are model parameters often called connection weights; p is the number of input nodes; and q is the number of hidden nodes.

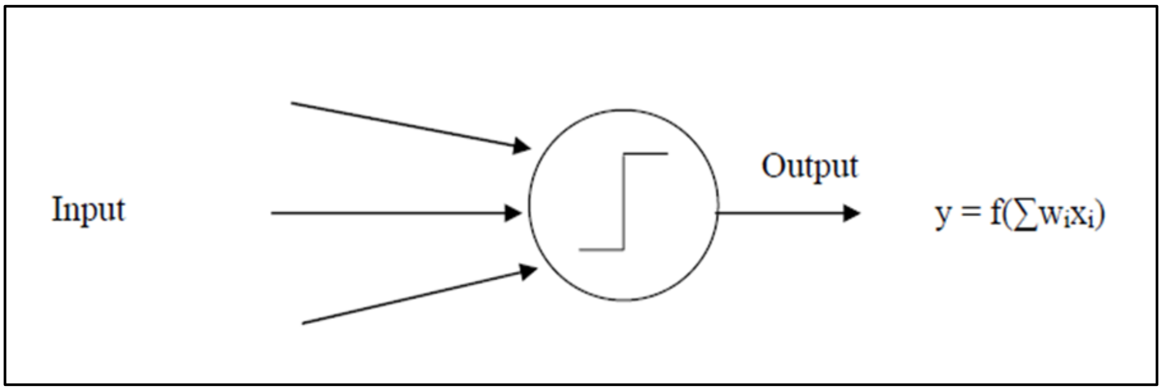

Figure 3-12 Schematic Diagram of a Single Artificial Neuron

(Adamowski et al., 2012 and Kulandaivel, 2004)

Figure 3-3 shows a diagram demonstrating input with a summation n feed into a neuron that computes the inputs produces a binary output y which is either +1 or -1. The bias weight, Ө, is introduced with a fixed input at +1. The bias weight allows greater flexibility of the learning process.

$$y = f\left(\sum_{i=1}^{n} W_i X_i\right) = \left|\begin{matrix} 1 \text{ when } \sum_{i=1}^{n} W_i X_i > 0 \\ -1 \text{ when } \sum_{i=1}^{n} W_i X_i \geq 0 \end{matrix}\right. \qquad \text{Eq. 3-27}$$

In equation 3-27, y is the neuron and f is a threshold function known as the neuron’s transfer function, which gives an output of +1 whenever Σwixi is greater than zero (the threshold value) or -1 whenever Σwixi is less than (or equal to) zero (Kulandaivel, 2004).

In this study the activation function used to act upon input to get output is bipolar sigmoid function. The bipolar sigmoid function Formula is in equation 3-28 and Figure 3-4 (Chakraverty and Susmita, 2017).

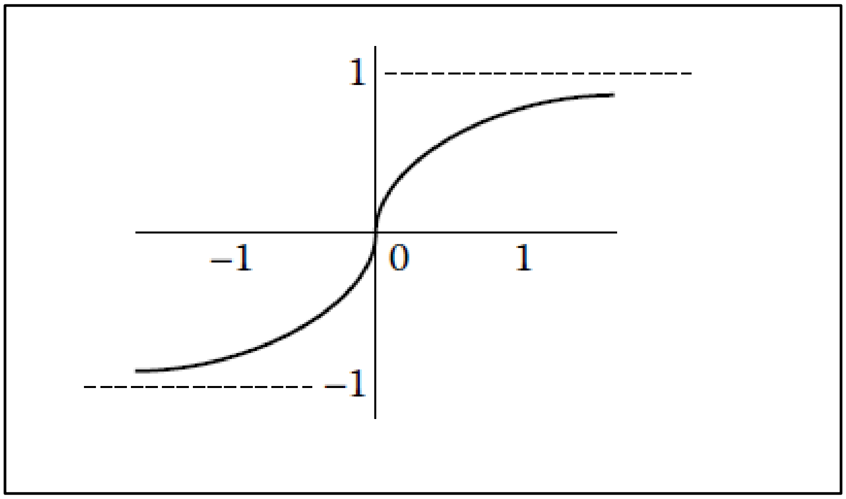

Figure 3-13 Plot on Bipolar Sigmoid Function

The output of the bipolar sigmoid function is between -1 and 1.

$$f(x) = \frac{1}{1+e^{-\lambda x}} - 1 = \frac{1-e^{-\lambda x}}{1+e^{-\lambda x}} \quad \text{Eq. 3-28}$$

3.3.2 Backpropagation Neural Network (BPNN)

BPNN comprises of Feed-Forward Neural network and Feedback Neural network. In Feed-Forward Neural network, the data is placed in the input layer is received by neurons as input and then feed their outputs to the next layer. The data comes from the input node to the output node in a forward way (Chakraverty and Mall, 2017). In feedback neural network the output of one-layer loops back to the previous layer. Feedback neural networks are used in optimization problems, where the network looks the best arrangement of interconnected factors (Chakraverty and Susmita, 2017). Figure 3-5 shows BPNN.

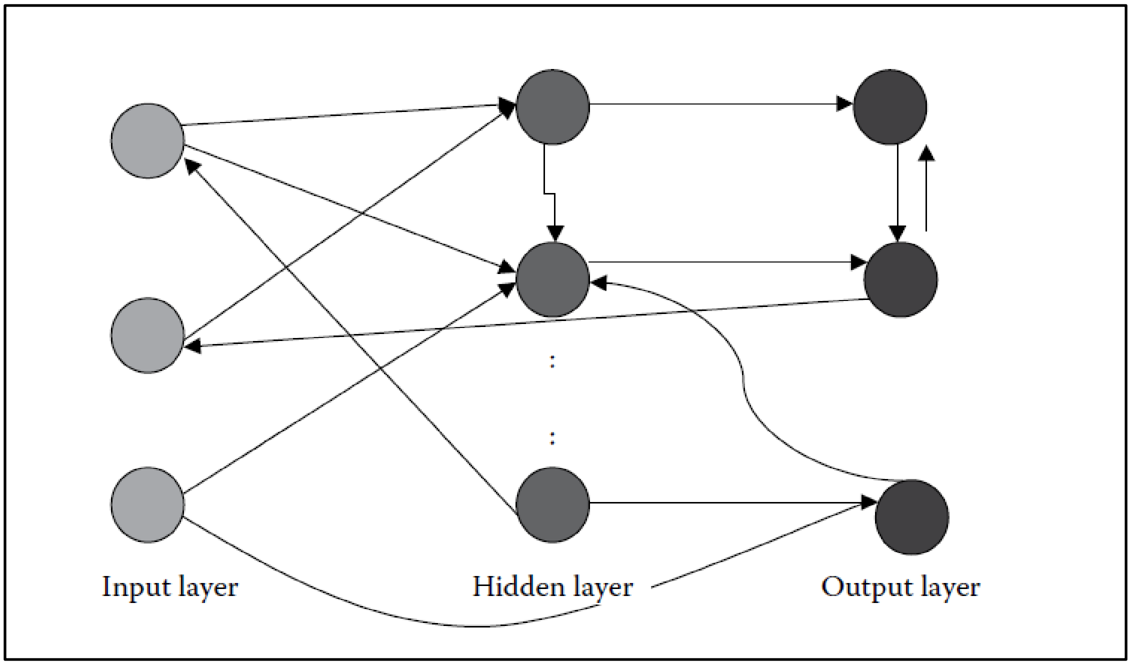

Figure 3-14 Diagram of Feedback Neural Network

(Chakraverty and Susmita, 2017)

According to Tavakoli, 2018, the BPNN network operates in two modes: mapping and training. The information flows forward in the mapping mode substitutes between forward and backward in training mode. The input layer sends the information of the independent variables to the hidden layer. The hidden later computes and assign weights to each variable. The hidden layer sends the results to the output layer. Each output-layer neuron completes a similar calculation and outputs the resulting value as an estimate of the dependent variable it represents (Tavakoli, 2018).

3.3.3 ANN Training

Training of ANNs is achieved by trial-and-error method with weights randomly taking numbers in training the model. The training stops when the neural network reaches lowest training and testing errors. Error calculation involves the derivative of output with respect to the input (Chakraverty and Mall, 2017). Considering a multilayer ANN with one input node x with h number of data, a hidden layer with m nodes, and one output unit. The network output N(x,p) is developed as (Chakraverty and Mall, 2017).

$$N(x, p) = \sum_{j=1}^{m} v_j \, s(Z_j) \qquad \text{Eq. 3-29}$$

The derivative of N(x,p) with respect to input x is

$$\frac{d^k N}{dx^k} = \sum_{i=1}^{m} v_j\, w_j^k s_j^k \qquad \text{Eq. 3-30}$$

The derivative N_a of the network with respect to other paraments is computed as in equation 3-31

$$\frac{\partial N_a}{\partial w_j} = x v_j P_j s_j^{(\Lambda+1)} + v_j \Lambda_j w_j^{\Lambda_{j=1}} \quad i = 1,2,\ldots,n \qquad \text{Eq. 3-31}$$

The gradient of error is found from the derivatives. The error function is minimized using the error back propagation learning method for unsupervised training (Chakraverty and Mall, 2017).

During training, the model can either be supervised or unsupervised learning. In supervised learning a comparison of computed output and corrected expected output is compared. The error is determined and then used to improve the performance of the model. In unsupervised learning, the network is provided with input data without desired outputs. The learning process is to adjust all the weights and let the output y approach the desired output so that the neuron performs the classification task correctly.

3.3.4 ANN Validation and Test data

When the model is being trained, validation and testing happens concurrently. The testing process utilized training and validation to assess the performance of the model. The performance of the ANN model was optimized by varying the number of neurons using a trial-and-error procedure. The number of neurons in the input and output layers were based on the input and output variables. Chughtai et al., 2008 stated that the performance of different models is assessed in terms of goodness of fit. Adamowski et al., (2012), described the coefficient of determination (R2) as measures the degree of correlation among the observed and predicted values. The R2 values are between 0 to 1. One (1) represents a perfect relationship between the data and the line drawn through them, and 0 represents no correlation between the data and the line. Adamowski et al. (2012) further stated that the root mean square error (RMSE) evaluates the variance of errors independently of the sample size. Adamowski et al., (2012), stated that a perfect fit between observed and forecasted values would have a RMSE of 0.

The metrics of precision, recall, and accuracy are also used in measuring performance of the model in category classification (Yin et al., 2020).The metrics of precision, recall, and accuracy are presented in Equations 3-32, 3-33, and 3-34, respectively (Yin et al., 2020).

$$\text{Precision} = \frac{TP}{TP+FP} \qquad \text{Eq. 3-32}$$

$$\text{Recall} = \frac{TP}{TP+FN} \qquad \text{Eq. 3-33}$$

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \qquad \text{Eq. 3-34}$$

Where the definitions of TP, FN, FP, TN are summarized in Table 3-2.

Table 3-18 Metrics of Measuring Performance

	True condition	
	Positive	Negative
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

Equations 3-35, 3-36, 3-37, 3-38, and 3-39 describes the metrics of measuring model performance.

$$True\ Positive\ Rate = TPR = Sensitivity = \frac{TP}{TP+FN} \qquad \text{Eq. 3-35}$$

$$True\ Negative\ Rate = TNR = Specificity = \frac{TN}{FP+TN} \qquad \text{Eq. 3-36}$$

$$True\ Positive\ Rate = TPR = Sensitivity = \frac{TP}{TP+FN} \qquad \text{Eq. 3-37}$$

$$False\ Positive\ Rate = FPR = 1 - Specificity \qquad \text{Eq. 3-38}$$

$$False\ Negative\ Rate = FNR = 1 - Sensitivity \qquad \text{Eq. 3-39}$$

Precision describes the correct rate for specific category, while recall is the rate of completeness of the classified sample for a specific category and accuracy description of percentage correctly classified sample for specific category (Yin et al., 2020).

3.3.5 Optimization of the ANN Structure

The minimum value of the mean squared error (MSE) of the training and validation sets is used to determine the optimization of the architecture of the ANN models and its parameters variation.

The mean squared error is calculated in equation 3-40 (Young, 2017).

$$\text{MSE} = \frac{\text{SSE}}{\text{df}_\text{E}} = \frac{\text{SSE}}{\text{n}-2} \qquad \text{Eq. 3-40}$$

Where SSE is the error sum squares, df_E = n-p., p is the dimension of β in the model, including intercept β_0.

The Lowest MSE for the training and the validation sets is used as the criteria for selecting the best neural networks architecture.

The root mean square error (RMSE) and the coefficient of determination (R^2) are used to determine the performance of the model. Chughtai et al., 2008, describe RMSE criteria as in equation 3-41.

$$\mathrm{RMSE} = \frac{\sqrt{\sum_{i=1}^{n}(C_i - E_i)^2}}{n} \qquad \text{Eq. 3-41}$$

Where n = number of observations, C_i = actual value, and E_i = estimated or predicted value.

3.4 Chapter Summary

This chapter discussed logistic regression and ANN methods. Regression assumptions test, test of significance coefficients, and the classification table of logistic regression models were described. ANN architecture, BPNN, ANN training, validation, testing, performance, and optimization of ANN structure were explained.

Chapter 4

Data Collection, Preparation, and Analysis

4.1 Introduction

In this chapter, data collection, preparation, and processing were discussed. Histograms showing the frequency of variables influencing the condition rating of sewer pipes were provided and discussed. The histograms were used to make comparisons of the factors influencing sewer pipes condition. In data collection, the data was grouped into independent and dependent variables. Descriptive statistics and correlation analysis of the data was presented in tables 4-2 and 4-3, respectively. Lastly, summary of this chapter is described.

4.2 Data Collection

This research was based on data collected from Dallas Water Utilities Wastewater collection system. The City of Dallas uses geographic information system (ArcGIS) as the primary source of information to manage and maintain their wastewater networks. The data was extracted from Geographical Information System (GIS) files for the City of Dallas Water Utilities (DWU) GIS web/data base and imported to excel worksheets.

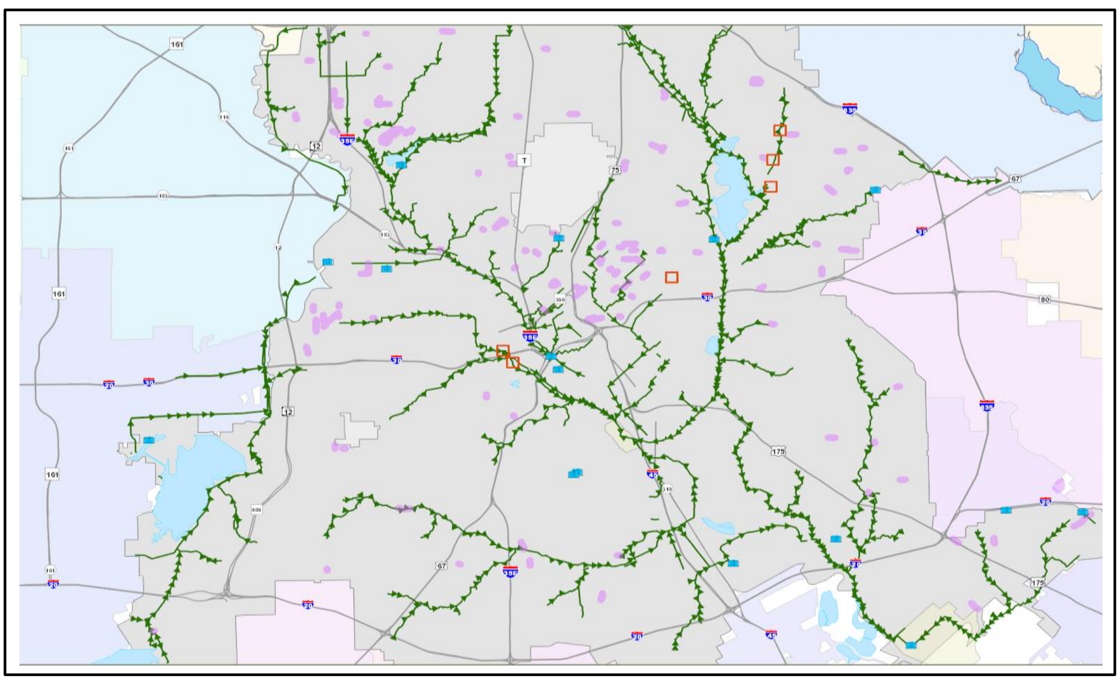

Figure 4-15 City of Dallas Sanitary Sewer Pipe Network

(Source: Dallas Water Utilities)

The database was initially filtered, and insufficient data was removed. The GIS data originated from condition assessment and CCTV inspection records. The condition of the sewer pipe was assessed using Pipeline Assessment and Certification Program (PACP) standard developed by National Association of Sewer Service Companies (NASSCO) (2018). The dataset in this study included pipe material, diameter, age, slope, depth, surface condition, soil type, corrosivity concrete, corrosivity steel, and pH variables.

Figure 4-16 Sanitary Sewer Pipe Segments

(Source: Dallas Water Utilities)

Figure 4-2 shows sewer pipes segments and their locations. Pipes sizes in inches, type of pipe materials, lot numbers and street names are shown.

4.3 Data Preparation and Processing

The data was grouped into input and output variables. The input variables were pipe material, diameter, age, slope, depth, surface condition, soil type, corrosivity concrete, corrosivity steel, and pH. The output variable was condition rating. The Dallas Wastewater Collection pipeline data acquired comprised of pipe segments/locations, length (manhole to manhole), pipe material, pipe diameter, pipe age (year of installation), depth (depth of backfill over the crown of pipe in ft), soil conditions, corrosivity (concrete and steel), slope, surface condition (highway, street, alley, and easement), and condition rating.

The following variables were transformed by recoding them into different variables.

Material: Materials were re-coded in three (3) types by assigning new values of 1 to 3.

1- CONC

2- PVC

3- VCP

There were few materials not in the category provided above. They were classified as missing data and were removed from the data sets.

Surface Condition: The surface condition where the pipes are located was grouped into four (4) and assigned new values 1 to 4.

1- Highway

2- Street

3- Alley

4- Easement

Soil type (Soil Texture): There were five (5) soil texture types. The soil texture types were assigned new values numbers as follows.

1- Sand

2- Loam

3- Clay

4- Rock

Corrosivity: Corrosivity for concrete (Corrosivity Concrete) and Steel (Corrosivity Steel) were assigned new values as follows.

Corrosivity Concrete: Nineteen (19) data sets were missing. They were removed.

1- Low

2- Moderate

3- High

Corrosivity Steel: The nineteen (19) missing data sets were removed.

2- Moderate

3- High

Condition Rating:

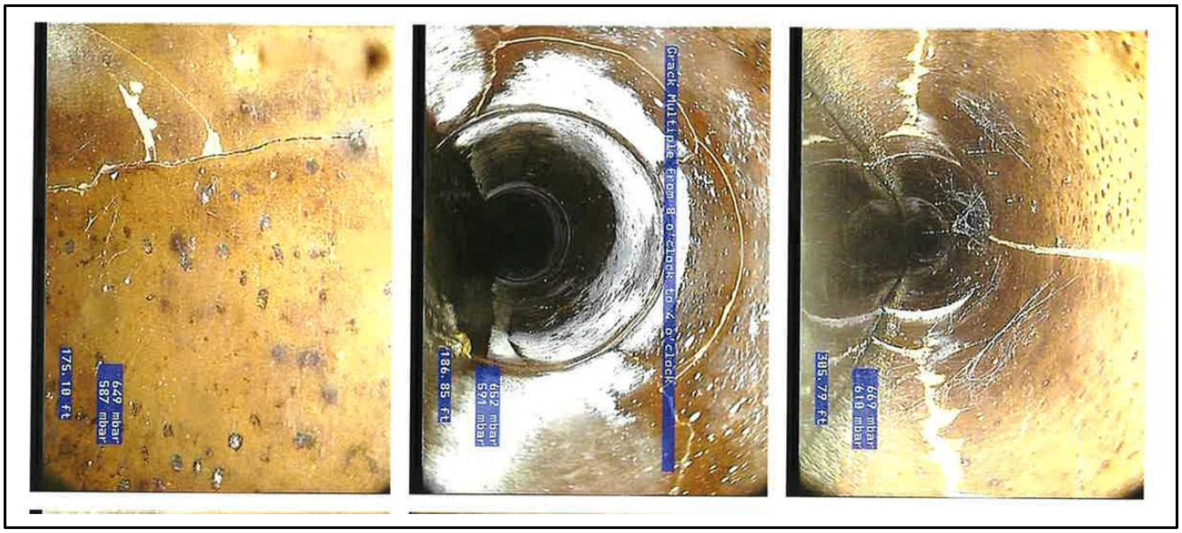

Figure 4-17 Sanitary Sewer Pipes Condition

(Source: Dallas Water Utilities)

The pipe condition score recorded from 1 to 5 were assigned new values.

1- Excellent, 2-Good, 3-Fair, 4- Poor, and 5-Extremely Poor

Twenty-four (24) data sets that were missing were removed.

Capacity: There were erroneous data sets with negative values of -67 and -4. The capacity or flow cannot be negative. The 2 data sets were removed.

Age: One (1) data set that had an age of 170 was removed.

Pipe material, diameter, depth, length, slope, age, surface condition, soil type, PH, and corrosivity, were grouped as physical and environmental independent variables. The condition rating score was used as the dependent variable. The variables were grouped into physical and environmental factors.

Table 4-19 Factors Influencing Sewer Pipe Condition Rating

Factors	Pipe Variables	Variables Description
Physical	Pipe material	Concrete (CONC.), Polyvinyl chloride (PVC), and Vitrified Clay (VCP). Continuous variable.
	Pipe Diameter (in.)	Nominal value of diameter of pipe. Categorical variable.
	Pipe length (feet)	Length of pipe between two manholes measured in feet. Continuous variable.
	Age (years)	Age of installed pipe is obtained by subtracting installation or rehabilitation year from inspection year. Continuous variable.
	Depth (feet)	Installation depth of pipe in ft
	Pipe slope (%)	Segment bed slope of pipe in percentage (upstream invert elevation – downstream invert elevation)/length. Continuous variable.
Environmental	Surface condition	Category of ground surface where the pipe is located (Highway, Street, Alley, and Easement)
	Soil type	Different texture of soils physical and chemical properties (Sand, Loam, Clay, and Rock).
	Corrosivity steel	Level of corrosion attack on steel (Moderate or High)
	Corrosivity concrete	Level of corrosion attack on steel (Low, Moderate, and High)
	pH	Soil pH

4.4 Regression Assumptions Tests

Regression analysis assumptions were checked using linear regression in IBM SPSS Statistics 27 data analysis software. The assumptions tested were independence, homogeneity, linearity, normality, multicollinearity, no outliers, and homoscedasticity. A test was conducted by checking for symmetry in a histogram, frequency distribution, boxplot, or skewness and kurtosis statistics.

4.5 Descriptive Statistics

The descriptive statistics in Table 4-2 shows 2,616 datasets for eleven (11) independent variables and one (1) dependent variable. The minimum and maximum age of the sewer pipes were one (1) year and 120 years, respectively. The minimum diameter of the sewer pipes was 6 in. whereas the maximum diameter was 90 in. The mean depth and length of the sewer pipes were 7.6 ft and 311 ft, respectively. The average kurtosis standard error was 0.096, while the average skewness standard error was 0.048

Table 4-20 Descriptive Statistics

Variables	N	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Diameter	2616	6	90	12.87	0.221	11.30	127.64	3.33	0.048	12.81	0.096
Age	2616	1	120	44.05	0.465	23.78	565.31	0.19	0.048	-1.10	0.096
Slope	2616	0	573	1.29	0.222	11.36	129.06	48.84	0.048	2455.65	0.096
Depth	2616	3	20	7.61	0.044	2.27	5.17	1.19	0.048	1.91	0.096
Length	2616	0	2054	310.97	4.861	248.62	61813.42	1.95	0.048	5.91	0.096
pH	2616	5.3	8.2	7.87	0.011	0.55	0.30	-2.30	0.048	5.68	0.096
Material	2614	1	3	2.07	0.014	0.72	0.52	-0.11	0.048	-1.09	0.096
Surface Condition	2616	1	4	2.59	0.016	0.84	0.70	0.78	0.048	-0.95	0.096
Soil Type	2616	1	4	2.64	0.015	0.77	0.60	-0.81	0.048	0.15	0.096
Corrosivity Concrete	2616	1	3	1.20	0.010	0.51	0.26	2.51	0.048	5.27	0.096
Corrosivity Steel	2616	1	2	1.94	0.005	0.24	0.06	-3.61	0.048	11.04	0.096
Condition Rating	2616	1	5	1.68	0.024	1.24	1.53	1.64	0.048	1.38	0.096

4.6 Condition Rating Score

Condition rating score was the dependent variable used in the data analysis. Each dataset was assigned a condition score generated from the condition assessment of the sanitary sewer pipes. The score followed the NASSCO PACP standard.

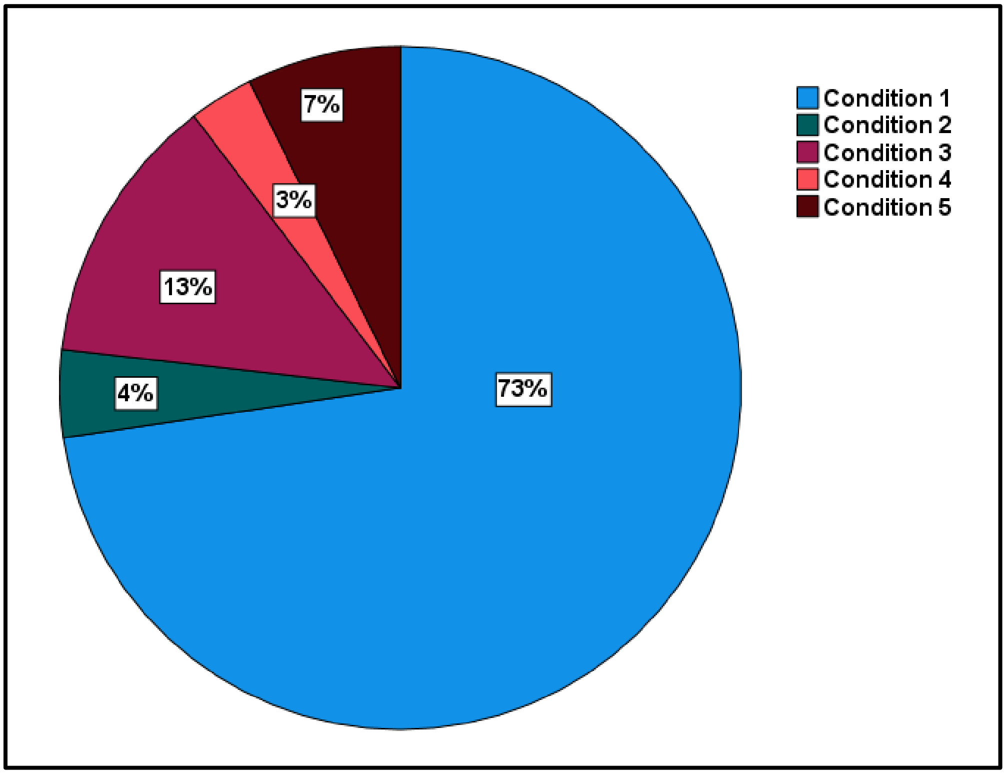

Figure 4-18 Distribution of Sewer Pipe Condition Rating

Figure 4-4 describes sewer pipe condition rating. Condition 1 represented sewer pipes in excellent condition. Condition 2 represents sewer pipes in good condition. Conditions 3,4, and 5 represents sewer pipes in Fair, Poor, and Extremely Poor conditions, respectively. The sewer pipes in excellent (1) and good (2) conditions have higher percent frequency compared to sewer pipes in fair (3), poor (4), and extremely poor (5) conditions. It was found that Seventy-two percent (73%) of the sewer pipes segments were in condition 1. Conditions 2, 3, and 4 of the sewer pipes segments were comprised of 4%, 13%, 3%, and 7%, respectively. This revealed that most of the sanitary sewer pipes were in good condition.

4.7 Physical Factors

The physical variables include pipe material, pipe diameter, pipe length, age, depth, and pipe slope.

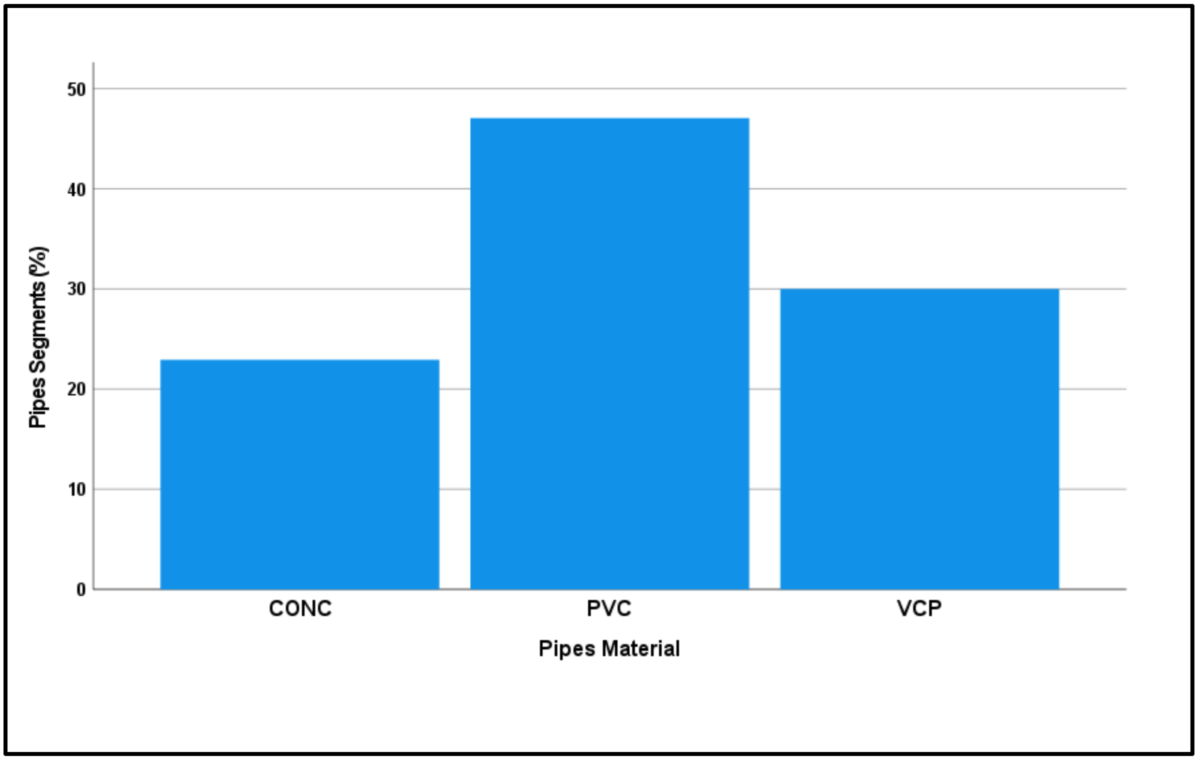

Figure 4-19 Pipe Material Distribution

Figure 4-5 shows pipe material distribution in the dataset. The pipe materials were evenly distributed. The distribution percent frequency demonstrated that most of the pipes comprised of PVC material. PVC pipe material had the highest frequency of 47%. VCP had the second-highest frequency followed by CONC pipe materials, respectively. The percent segments with VCP and CONC sewer pipes were 30% and 23%, respectively.

Figure 4-6 shows the distribution of pipe size in the datasets. Figure 4-6 shows that the minimum and maximum sewer pipes diameters were 6 in. and 72 in., respectively. Figure 4-6 illustrates that small diameter (<18 in.) sewer pipes were more in the dataset compared with larger diameter (>18 in.) sewer pipes. It was found that sewer pipes diameter of 8 inch had the largest frequency of 48%.

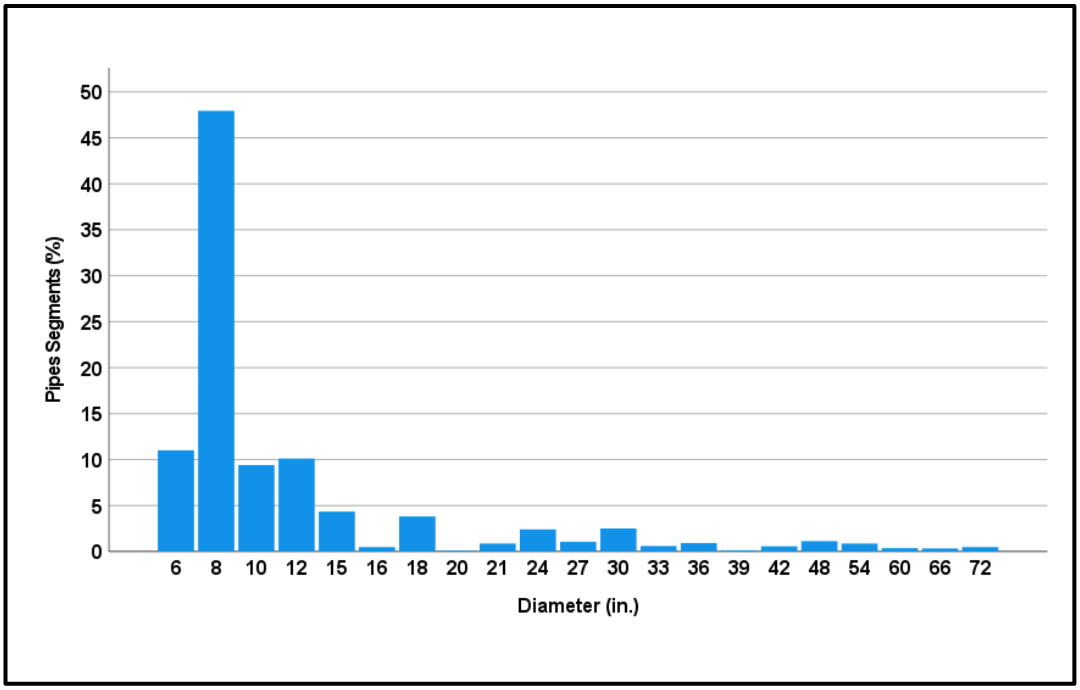

Figure 4-20 Pipe Size Distribution

Figure 4-7 shows pipe length distribution in the dataset. The mean and standard deviation of the sanitary sewer pipes length were 311 ft and 2450 ft, respectively. Figure 4-7 shows that most pipe lengths are between 100 ft to 500 ft. The longer sewer pipe means the number of pipe joints is reduced. Figure 4-7 demonstrates that pipe lengths greater than 1000 ft were less compared to shorter lengths less than 1000 ft.

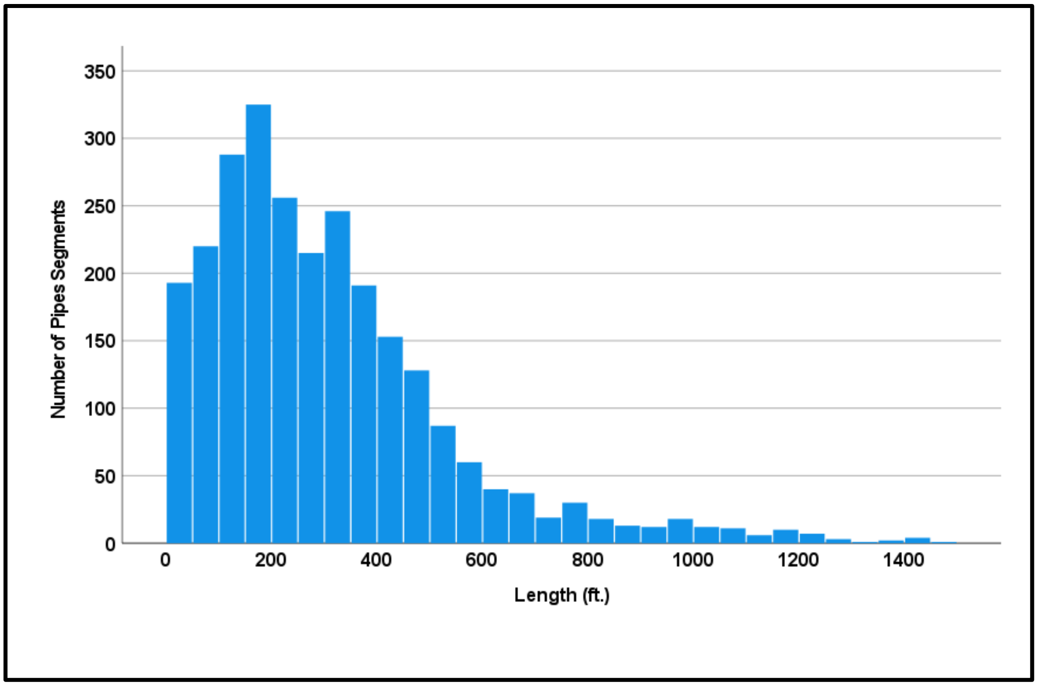

Figure 4-21 Pipe Length Distribution

Figure 4-8 shows the age of the distribution in the dataset. The mean and standard deviation of the sanitary sewer pipes age was 44 years and 24 years, respectively. The age of pipes with the highest frequency was 15-45 years and 55-75 years, respectively. The graph shows that more than 50% of the sewer pipes were more than 55 years old.

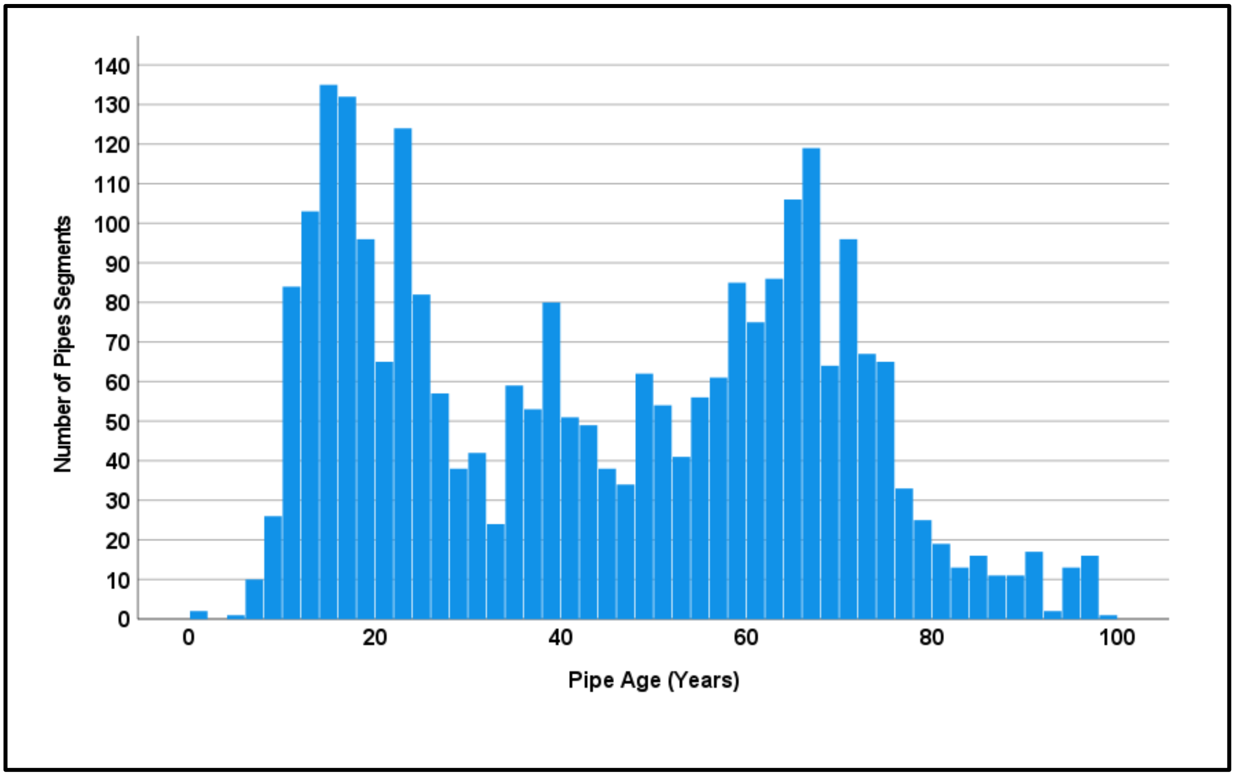

Figure 4-22 Pipe Age Distribution

Figure 4-9 shows the distribution of depth in the dataset. Figure 4-9 illustrates that most of the sewer pipes were buried from 5 ft to 10 ft. The Sewer pipe depth with the highest frequency was 7 ft followed by 5 ft, 8 ft, and 10 ft, respectively.

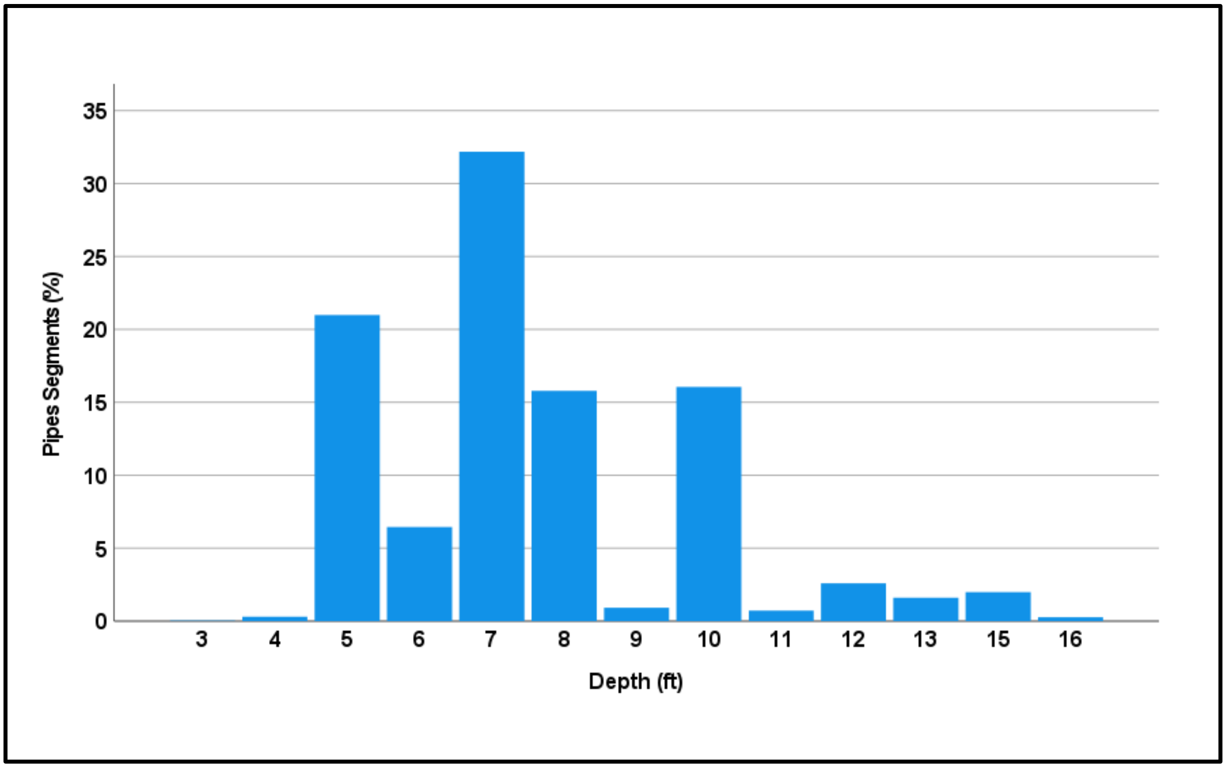

Figure 4-23 Pipe Depth Distribution

4.8 Environmental Factors

The environmental variables included surface condition, soil type, corrosivity concrete, corrosivity steel, and soil pH. Figure 4-10 shows the distribution of surface conditions in the location where the pipes were buried. Figure 4-13 reveals that sewer pipes located beneath streets were 61% of the total segments in the datasets. This was followed by sewer pipes located beneath easements (22%), alleys (16%), and Highways (1%) surface conditions, respectively, from the highest frequency to the least. The sewer pipes located beneath easements were more than the ones located in the alleys.

Figure 4-11 shows the frequency distribution of soil types in the location of the sewer pipes. There are four soil types: sand, loam, clay, and rock. Figure 4-11 demonstrates that 63% of the sewer pipe segments were installed in locations with clay soil. Loam soil had the next high frequency with 19% of the sewer pipes segments. Sandy soil type had 12% of the sewer pipes segments in the datasets Sewer pipes located in the rock soil type had the lowest frequency (7%).

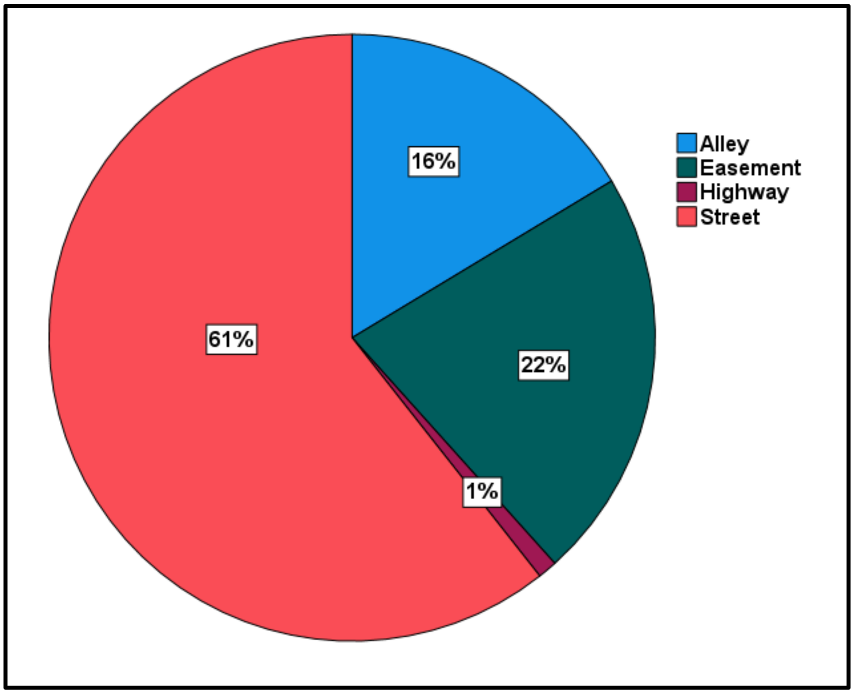

Figure 4-24 Surface Condition Distribution

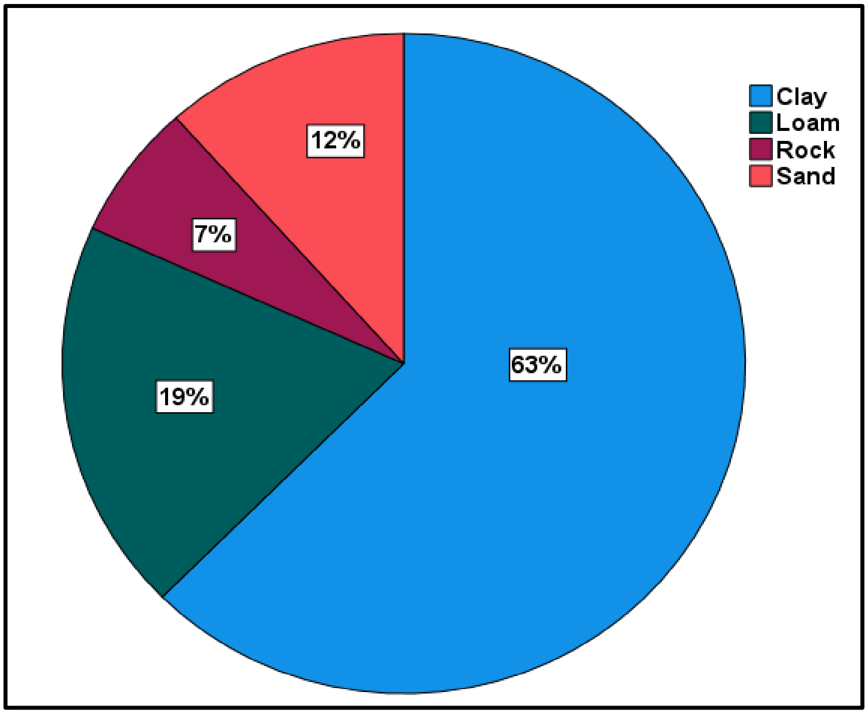

Figure 4-25 Soil Type Distribution

Figure 4-12 shows the distribution of soil corrosivity for concrete ranging from low to high. Figure 4-12 illustrates that most of the sewer pipes were in soils with low corrosivity for concrete. The soils with low corrosivity for concrete were 85% of the sewer pipe segments in the datasets. The sewer pipes located in soil with high corrosivity for concrete were 7% of the segments in the datasets. This is possible since most of the sewer pipes were in clay and loam soils. In addition, most of the pipe materials were PVC which was low in susceptibility to corrosion.

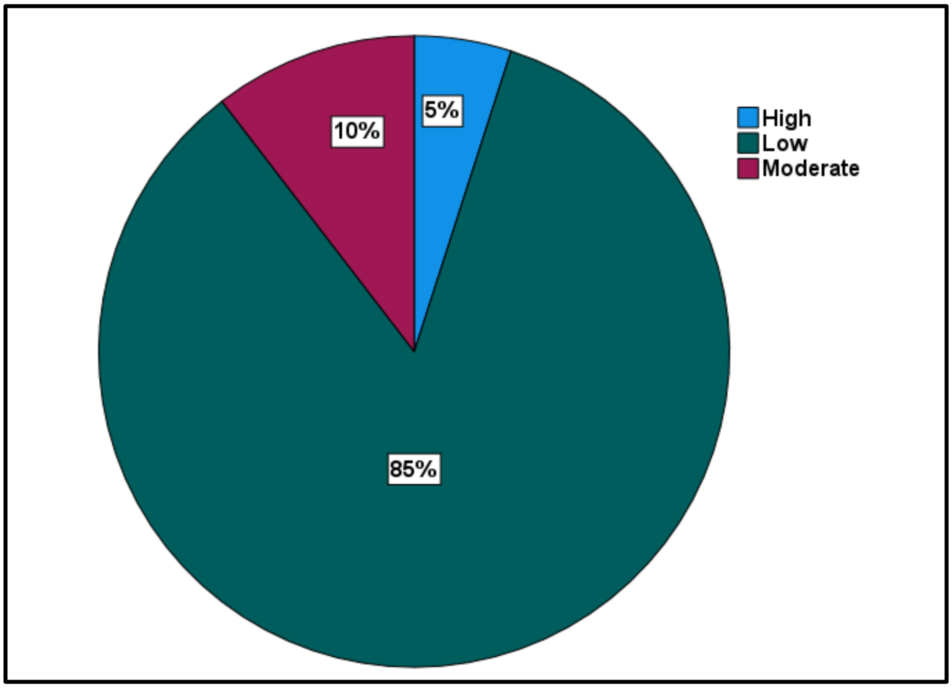

Figure 4-26 Level of Corrosivity for Concrete Distribution

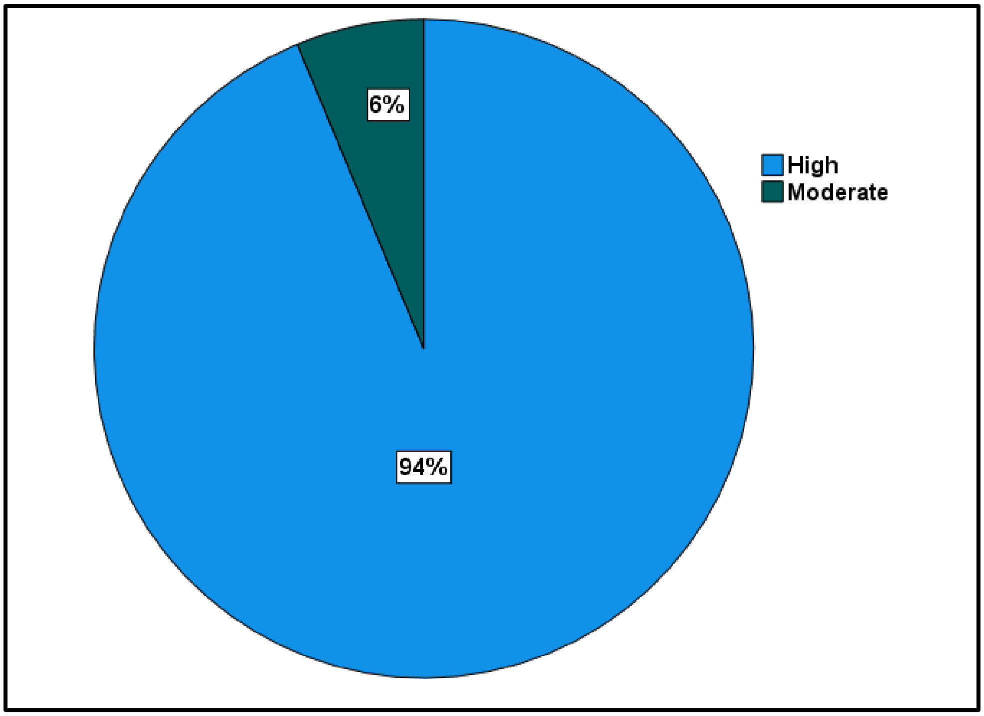

Figure 4-27 Level of Corrosivity Steel Distribution of

Figure 4-13 shows the distribution of sewer pipes with the level of moderate to high corrosivity steel. The sewer pipes in the soils with moderate corrosivity for steel were 6% of the sewer pipes segments in the datasets. Whereas the soils with high corrosivity for steel were 94% of the sewer pipes segments in the datasets.

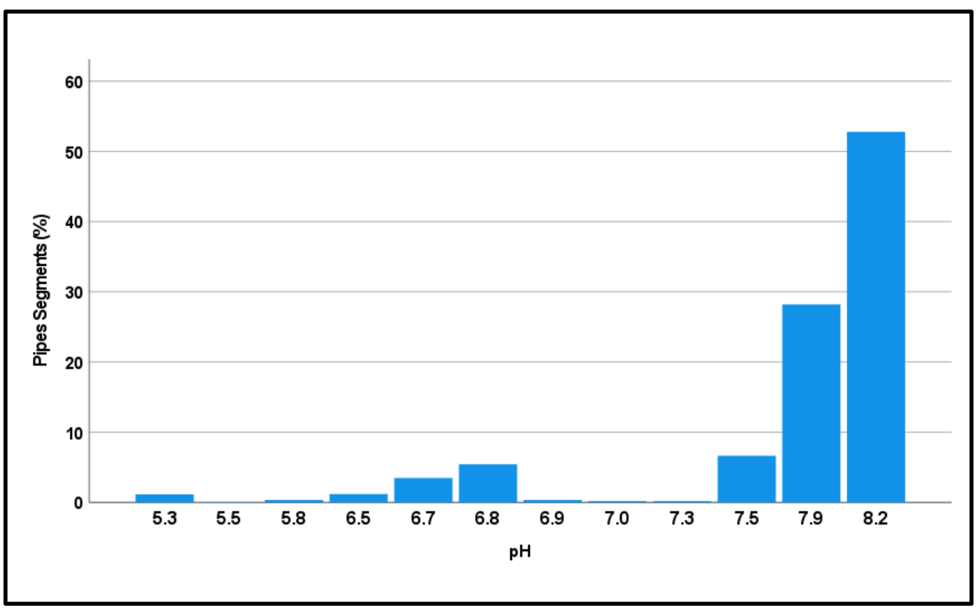

Figure 4-28 pH Distribution

Figure 4-14 shows the distribution of sewer pipes by pH. Figure 4-14 showed that most of the pipes were in soils with pH greater than 7.5. Few sewer pipes were in soils with pH of less than 7.5. Sewer pipes located in the soils with a pH of 8.2 were 53% of the sewer pipe segments provided in the datasets. Soils with pH 7.5 and 7.9 were 7% and 28% of the sewer pipes segments of the datasets, respectively.

4.9 Correlation Analysis

Table 4-3 demonstrates the relationship between the variables. Table 4-3 revealed all the variables correlated at a significance level of 0.01 and 0.05. The correlation efficiency was below 0.5 except for the correlation between pH and corrosivity steel (0.712), pH and Soil Type (0.673), pH corrosivity concrete (0.575), and soil type and corrosivity steel (0.542), respectively.

Table 4-21 Pearson Correlation Analysis

Variables	Diameter	Age	Slope	Depth	Length	pH	Material	Surface	Soil Type	Corrosivity Concrete	Corrosivity Steel	Condition Rating
Diameter	1	0.107**	-0.036	0.138**	0.339**	0.027	-0.365**	-0.017	-0.129**	-0.042*	-0.011	0.183**
Age		1	-0.041*	0.037	0.165**	0.038	0.070**	0.009	-0.029	-0.064**	-0.009	0.475**
Slope			1	0.015	-0.024	0.028	0.004	0.010	0.021	-0.020	0.010	-0.027
Depth				1	0.048*	0.017	-0.076**	0.080**	0.052**	0.010	0.014	-0.008
Length					1	-0.015	-0.133**	-0.028	-0.069**	-0.010	0.012	0.215**
pH						1	-0.096**	-0.016	0.673**	-0.575**	0.712**	0.031
Material							1	0.018	0.011	0.075**	-0.033	-0.025
Surface								1	-0.020	0.021	0.034	-0.017
Soil Type									1	-0.288**	0.542**	-0.020
Corrosivity Concrete										1	-0.117**	-0.045*
Corrosivity Steel											1	-0.010
Condition Rating												1

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

4.10 Chapter Summary

This chapter discussed data collection, data preparation and processing presented, descriptive statistics, preliminary data analysis, condition rating score, factors influencing pipe condition rating, and correlation analysis were presented. The physical and environmental factors influencing sewer pipes condition rating were discussed. The frequency distribution of independent variables and dependent variables was illustrated and described.

Chapter 5

Prediction Model Development

5.1 Introduction

In this chapter, MLR and ANN model development are illustrated. The MLR modis developed using IBM SPSS Statistics (version 27). ANN model is developed using Brainmaker California Scientific Software. Before the development of the models, eighty percent (80%) of the data is randomly selected and the remaining 20% is set aside to validate the models and or used as a case study to check the applicability of the models. Eighty-five percent (85%) and fifteen percent (15%) of the randomly selected 80 % data are used in developing and testing the ANNs model, respectively. Table 5-1 shows a sample of the 80% sewer pipes dataset.

5.2 Multinomial Logistic Regression Model

The objective of MLR analysis was to study the relationship between eleven (11) independent or predictor variables and one (1) dependent variable. Pipe material, diameter, age, slope, depth, surface condition, soil type, corrosivity concrete, corrosivity steel, and pH were independent variables used to generate prediction models. The condition rating score was the dependent variable.

5.2.1 Model Parameters Estimation

The data that was utilized to develop the prediction model was randomly divided into 80% and 20% for MLR model development and validation, respectively. MLR analysis was conducted and got one set of results which represents statistics for four (4) sewer pipes conditions (5-1). A Multinomial logistic regression model was developed with sewer pipe condition rating dependent variable with 5 categories, category 5 being reference category. One (1) model was built for five sewer pipes condition ratings 1, 2, 3, and 4. Condition 5 was used as the reference category.

Table 5-22 Sample of 80% of Sewer Pipes Dataset

ID	Diameter	Age	Pipe Material	Slope	Surface Condition	Depth	Length	pH	Soil Type	CorCon	CorSteel	Condition Rating
2472	12	43	PVC	0.24	Street	15	480	6.7	Sand	Low	Moderate	1
1814	10	50	VCP	0.1	Easement	15	421	6.7	Sand	Low	Moderate	1
843	6	97	VCP	0.8	Alley	15	264	6.7	Sand	Low	Moderate	1
2343	8	23	PVC	0.3	Street	15	236	6.7	Sand	Low	Moderate	1
2795	18	50	VCP	0.08	Alley	15	81	6.7	Sand	Low	Moderate	1
65	8	50	VCP	0.3	Street	11	536	6.7	Sand	Low	Moderate	1
623	12	71	CONC	0.6	Highway	10	472	6.7	Sand	Low	Moderate	1
624	24	64	CONC	0.12	Street	10	465	6.7	Sand	Low	Moderate	1
2366	12	51	VCP	0.3	Alley	10	401	6.7	Sand	Low	Moderate	1
3215	8	22	PVC	0.33	Street	10	384	6.7	Sand	Low	Moderate	1
3097	12	51	VCP	0.3	Street	10	325	6.7	Sand	Low	Moderate	1
1365	8	24	PVC	0.4	Alley	10	284	6.7	Sand	Low	Moderate	1
3327	48	29	PVC	0.14	Street	10	278	6.7	Sand	Low	Moderate	1
2146	12	39	PVC	2.1	Street	10	159	6.7	Sand	Low	Moderate	1
2295	15	66	VCP	0.32	Street	10	156	6.7	Sand	Low	Moderate	1
285	8	35	PVC	0.8	Easement	10	99	6.7	Sand	Low	Moderate	1
181	10	48	VCP	0.8	Alley	10	70	6.7	Sand	Low	Moderate	1
47	8	16	PVC	0.4	Street	10	24	6.7	Sand	Low	Moderate	1
2428	12	9	PVC	0.2	Street	8	480	6.7	Sand	Low	Moderate	1

Table 5-1 shows a sample of 80 % of the sewer pipes dataset. Likelihood ratio tests, the significance of variables was determined by Wald test and P-values set at 95% confidence levels. Table 5-2 shows likelihood ratio tests. Table 5-3 shows the model fit information.

Table 5-23 Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	3811.161	0.00	0	.
Diameter	3840.62	29.46	4	0.000
Age	3938.10	126.94	4	0.000
Slope	3817.32	6.16	4	0.188
Depth	3822.03	10.87	4	0.028
Length	3839.21	28.04	4	0.000
pH	3813.12	1.96	4	0.743
Pipe Material	3888.90	77.74	8	0.000
Surface Condition	3827.65	16.49	12	0.170
Soil Type	3832.76	21.60	12	0.042
Corrosivity Concrete	3828.06	16.90	8	0.031
Corrosivity Steel	3818.21	7.05	4	0.133

Table 5-24 Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	4839.163			
Final	3811.161	1028.002	68	0.000

From Table 5-3, the -2 Log-likelihood (-2LL) for the constant only model was the sum of Chi-square and (-2LL) for the full model. This was equal to 1028.002 plus 3811.161 which was equal to 4839.163. Since -2LL for the full model was less than (4839.163) and (1028.002) for the constant-only model, the model was a good fit. The developed model with a dependent variable with 5 categories is presented. The model is broken down into 4 tables of each category relative to condition 5 being reference category. Sewer pipe prediction parameter estimates for conditions 1, 2, 3, and 4 were presented in tables 5-4, 5-5, 5-6, and 5-7, respectively.

Table 5-25 Condition 1 Parameter Estimates

Variables	B	Std. Error	Wald	df	Sig.	Exp(B)
Intercept	4.650	3.790	1.505	1	0.220	
Diameter	**-0.022**	**0.007**	**10.390**	**1**	**0.001**	**0.978**
Age	**-0.056**	**0.006**	**81.139**	**1**	**0.000**	**0.945**
Slope	0.022	0.056	0.156	1	0.693	1.022
Depth	0.016	0.038	0.183	1	0.669	1.016
Length	**-0.001**	**0.000**	**20.852**	**1**	**0.000**	**0.999**
pH	0.279	0.630	0.196	1	0.658	1.322
Pipe Material = CONC	0.145	0.204	0.503	1	0.478	1.156
Pipe Material = PVC	0.646	0.372	3.020	1	0.082	1.908
Pipe Material = VCP	0b	.	.	0	.	.
Surface Condition = Alley	0.128	0.240	0.286	1	0.593	1.137
Surface Condition = Easement	0.257	0.219	1.376	1	0.241	1.293
Surface Condition = Highway	-0.071	0.803	0.008	1	0.930	0.931
Surface Condition = Street	0b	.	.	0	.	.
Soil Type = Clay	0.399	0.858	0.216	1	0.642	1.491
Soil Type = Loam	0.506	0.900	0.316	1	0.574	1.658
Soil Type=Rock	0.460	0.942	0.238	1	0.626	1.584
Soil Type = Sand	0b	.	.	0	.	.
Corrosivity Concrete = High	-1.287	0.828	2.414	1	0.120	0.276
Corrosivity Concrete = Low	-1.332	0.863	2.385	1	0.123	0.264
Corrosivity Concrete = Moderate	0b	.	.	0	.	.
Corrosivity Steel = High	-0.410	0.988	0.172	1	0.678	0.664
Corrosivity Steel = Moderate	0b	.	.	0	.	.

Table 5-26 Condition 2 Parameter Estimates

Variables	B	Std. Error	Wald	df	Sig.	Exp(B)
Intercept	5.024	5.799	0.751	1	0.386	
Diameter	**-0.080**	**0.022**	**12.975**	**1**	**0.000**	**0.923**
Age	**-0.032**	**0.009**	**11.571**	**1**	**0.001**	**0.968**
Slope	-0.037	0.093	0.158	1	0.691	0.964
Depth	0.018	0.057	0.094	1	0.759	1.018
Length	-0.001	0.000	1.456	1	0.228	0.999
pH	-0.514	0.976	0.277	1	0.598	0.598
Pipe Material = CONC	**-0.752**	**0.335**	**5.030**	**1**	**0.025**	**0.472**
Pipe Material = PVC	-0.968	0.534	3.288	1	0.070	0.380
Pipe Material = VCP (Reference)	0b	.	.	0	.	.
Surface Condition = Alley	0.164	0.353	0.216	1	0.642	1.179
Surface Condition = Easement	0.039	0.322	0.015	1	0.903	1.040
Surface Condition = Highway	0.148	1.261	0.014	1	0.906	1.160
Surface Condition = Street (Reference)	0b	.	.	0	.	.
Soil Type = Clay	-0.704	1.192	0.348	1	0.555	0.495
Soil Type = Loam	0.175	1.235	0.020	1	0.887	1.192
Soil Type = Rock	-1.021	1.343	0.578	1	0.447	0.360
Soil Type = Sand (Reference)	0b	.	.	0	.	.
Corrosivity Concrete = High	0.785	1.198	0.429	1	0.512	2.192
Corrosivity Concrete = Low	0.222	1.340	0.027	1	0.869	1.248
Corrosivity Concrete = Moderate (Reference)	0b	.	.	0	.	.
Corrosivity Steel = High	2.176	1.532	2.018	1	0.155	8.814
Corrosivity Steel = Moderate (Reference)	0b	.	.	0	.	.

Table 5-27 Condition 3 Parameter Estimate

Variables	B	Std. Error	Wald	df	Sig.	Exp(B)
Intercept	0.817	4.342	0.035	1	0.851	
Diameter	-0.009	0.007	1.650	1	0.199	0.991
Age	**-0.026**	**0.007**	**15.222**	**1**	**0.000**	**0.974**
Slope	-0.081	0.081	1.002	1	0.317	0.922
Depth	-0.078	0.043	3.269	1	0.071	0.925
Length	0.000	0.000	2.515	1	0.113	1.000
pH	0.424	0.720	0.347	1	0.556	1.528
Pipe Material = CONC	**0.737**	**0.224**	**10.828**	**1**	**0.001**	**2.089**
Pipe Materia = PVC	-0.659	0.440	2.244	1	0.134	0.517
Pipe Material = VCP (Reference)	0b	.	.	0	.	.
Surface Condition = Alley	0.416	0.259	2.567	1	0.109	1.515
Surface Condition = Easement	0.418	0.239	3.047	1	0.081	1.519
Surface Condition = Highway	-18.955	6041.973	0.000	1	0.997	0.000
Surface Condition = Street (Reference)	0b	.	.	0	.	.
Soil Type = Clay	0.142	0.978	0.021	1	0.884	1.153
Soil Type = Loam	0.628	1.026	0.374	1	0.541	1.873
Soil Type = Rock	-0.146	1.088	0.018	1	0.893	0.864
Soil Type = Sand (Reference)	0b	.	.	0	.	.
Corrosivity Concrete = High	-0.025	0.901	0.001	1	0.977	0.975
Corrosivity Concrete = Low	-0.705	0.948	0.552	1	0.457	0.494
Corrosivity Concrete = Moderate (Reference)	0b	.	.	0	.	.
Corrosivity Steel = High	-1.005	1.112	0.817	1	0.366	0.366
Corrosivity Steel = Moderate (Reference)	0b	.	.	0	.	.

Table 5-28 Condition 4 Parameter Estimates

Variables	B	Std. Error	Wald	df	Sig.	Exp(B)
Intercept	3.952	5.541	0.509	1	0.476	
Diameter	**-0.050**	**0.019**	**7.020**	**1**	**0.008**	**0.951**
Age	-0.002	0.010	0.060	1	0.807	0.998
Slope	-0.159	0.147	1.160	1	0.282	0.853
Depth	-0.078	0.070	1.247	1	0.264	0.925
Length	0.000	0.000	0.001	1	0.980	1.000
pH	-0.409	0.929	0.194	1	0.660	0.664
Pipe Material = CONC	-0.211	0.332	0.403	1	0.526	0.810
Pipe Material = PVC	-0.714	0.661	1.165	1	0.280	0.490
Pipe Material = VCP (Reference)	0b	.	.	0	.	.
Surface Condition = Alley	0.337	0.365	0.851	1	0.356	1.401
Surface Condition = Easement	-0.070	0.372	0.036	1	0.850	0.932
Surface Condition = Highway	-18.332	0.000	.	1	.	0.000
Surface Condition = Street (Reference)	0b	.	.	0	.	.
Soil Type=Clay	**2.21**	**1.12**	**3.89**	**1**	**0.05**	**9.12**
Soil Type = Loam	**2.27**	**1.15**	**3.93**	**1**	**0.05**	**9.68**
Soil Type = Rock	2.01	1.30	2.39	1	0.12	7.46
Soil Type = Sand (Reference)	0b	.	.	0	.	.
Corrosivity Concrete = High	**-2.340**	**1.114**	**4.411**	**1**	**0.036**	**0.096**
Corrosivity Concrete = Low	**-2.515**	**1.163**	**4.673**	**1**	**0.031**	**0.081**
Corrosivity Concrete = Moderate (Reference)	0b	.	.	0	.	.
Corrosivity Steel = High	0.328	1.425	0.053	1	0.818	1.389
Corrosivity Steel = Moderate (Reference)	0b	.	.	0	.	.

From Tables 5-4, 5-5, 5-6, and 5-7, in the column with the values of B means for one-unit change natural log of condition rating, results in unit change natural log of variables. The negative value of B means that an increase in the value of the independent variable results in a decrease in the predicted probability of the dependent variable. A positive value of B means an increase of the independent variable leads to an increase in the predicted sewer pipe condition (dependent variable). Wald value revealed the significance of the independent variables. The odds ratios Exp(B) shows the number of times greater (or percent) an independent variable was for everyone unit of the dependent variable (condition rating).

Table 5-4 represents sewer pipe condition 1, the highest Wald value is for age (81.139) followed by length (20.852) and diameter (10.390). This demonstrated that length, age, and diameter had more significance in the model. Diameter, age, and length have P Values of 0.001, 0.000, and 0.000, respectively. This means diameter, age, and length variables had a high influence on the sewer pipes condition 1. For this model, the Exp(B) values for the significant variables were diameter (97.9%), age (94.5%), and length (99.9%).

Table 5-5 represents sewer pipe condition 2. From Table 5-5, diameter had the highest Wald value (12.975) followed by age (11.571) and pipe material (CONC) (5.030). This revealed that diameter, age, and pipe material had more significance in the model predicting sewer pipe condition 2. Diameter, age, and pipe material (CONC) have P Values of 0.000, 0.001, and 0.025, respectively. This means diameter, age, and pipe material (CONC) variables had a high influence on the sewer pipes condition 2. For this model, the Exp(B) values for the significant variables were diameter (92.3%), age (96.8%), and pipe material (CONC) (47.2%).

Table 5-6 shows sewer pipe condition 3. In table 5-6, age had the highest Wald value (15.222) followed by pipe material (CONC) (10.828). It was revealed that age and pipe material had more significance in the model predicting sewer pipe condition 3. Diameter, age, and pipe material (CONC) have P Values of 0.000, 0.001, and 0.025, respectively. This means an age and pipe material (CONC) variables had a high influence on the sewer pipes condition 3. For this model, the Exp(B) values for the significant variables were age (97.4%), and pipe material (CONC) (208.9%).

Table 5-7 represents sewer pipe condition 4. Diameter had the highest Wald value (7.020) followed by corrosivity concrete (Low) (4.673), corrosivity concrete (High)(4.411), soil type (Loam) (3.93), and soil type (Clay) (3.89). It was demonstrated that age and pipe material had more significance in the model predicting sewer pipe condition 3. Diameter, soil type (Clay), soil type (Loam), corrosivity concrete (High), and corrosivity concrete (Low) have P Values of 0.008, 0.05, 0.05, 0.036, and 0.031, respectively. This means diameter, soil type (Clay), soil type (Loam), corrosivity concrete (High), and corrosivity concrete (Low) variables had a high influence on the sewer pipes condition 4. For this model, the Exp(B) values for the significant variables were diameter (95.1%), soil type (Clay) (912%), soil type (Loam) (968%), corrosivity concrete (High) (9.6%), and corrosivity concrete (Low) (8.1%).

5.2.2 Validation of MLR Model

The model parameters estimation tables in section 5.2.1 were used to derive four MLR equations. The one set model equation was broken down to 4 equations, one for each category relative to the reference category for sewer pipe condition 5. The four equations were used to predict sewer pipe conditions 1, 2, 3, and 4 relative to sewer pipes condition 5 that was used as a reference category. The variables coefficients (β) were used to develop the 4 multinomial logistic regression equations relative to condition 5 (C=5) reference category.

Equation 5-1 presents a MLR model developed to predict sewer pipe condition 1.

$$\mathrm{g_1(x)} = \ln\left[\frac{\mathrm{Pr(C=1)}}{\mathrm{Pr(C=5)}}\right] = \mathbf{0.978} * \mathbf{Diameter} + \mathbf{0.945} * \mathbf{Age} + 1.023 * \mathrm{Slope} + 1.018 * \mathrm{Depth} + \mathbf{0.999} * \mathbf{Length} + 1.321 * \mathrm{pH} + 1.146 * \mathrm{MaterialCONC} + 1.899 * \mathrm{MaterialPVC} + 0.721 * \mathrm{SurfaceAlley} + 0.771 * \mathrm{SurfaceEasement} + 0.879 * \mathrm{SurfaceHighway} + 0.619 * \mathrm{SoilTypeClay} + 1.037 * \mathrm{SoilTypeLoam} + 0.942 * \mathrm{SoilTypeRock} + 0.962 * \mathrm{CorrosivityConcreteHigh} + 3.653 * \mathrm{CorrosivityConcreteLow} + 1.533 * \mathrm{CorrosivitySteelHigh} \quad \text{Eq. 5-1}$$

Where:

Pr (C = 1) is the probability of sanitary sewer pipe condition dependent variable being condition 1 relative to condition 5.

Pr (C = 5) is the probability of reference category condition 5

Diameter, Age, Slope, Depth, Length, Material, Surface, Soil Type, and Corrosivity are independent variables that influence sanitary sewer pipe condition

Equation 5-2 presents a MLR model developed to predict sewer pipe condition 2.

$$g_2(x) = ln \left[\frac{Pr(C=2)}{Pr(C=5)}\right] = \mathbf{0.923} * \textbf{Diameter} + \mathbf{0.968} * \textbf{Age} + 0.964 * \text{Slope} + 1.018 * \text{Depth} + 0.999 * \text{Length} + 0.598 * \text{pH} + \mathbf{0.473} * \textbf{MaterialCONC} + 0.379 * \text{MaterialPVC} + 1.116 * \text{SurfaceAlley} + 0.961 * \text{SurfaceEasement} + 1.133 * \text{SurfaceHighway} + 2.758 * \text{SoilTypeClay} + 3.289 * \text{SoilTypeLoam} + 1.374 * \text{SoilTypeRock} + 0.572 * \text{CorrosivityConcreteHigh} + 0.459 * \text{CorrosivityConcreteLow} + 0.114 * \text{CorrosivitySteelHigh} \quad \text{Eq. 5-2}$$

Where:

Pr (C = 2) is the probability of sanitary sewer pipe condition dependent variable being condition 2 relative to condition 5.

Pr (C = 5) is the probability of reference category condition 5

Diameter, Age, Slope, Depth, Length, Material, Surface, Soil Type, and Corrosivity are independent variables that influence sanitary sewer pipe condition.

Equation 5-3 represents a MLR model developed to predict sewer pipe condition 3.

$$g_3(x) = \ln \left[\frac{Pr(C=2)}{Pr(C=5)}\right] = 0.991 * \text{Diameter} + \mathbf{0.974} * \textbf{Age} + 0.922 * \text{Slope} + 0.924 * \text{Depth} + 1.00 * \text{Length} + 1.532 * \text{pH} + \mathbf{2.090} * \textbf{MaterialCONC} + 0.518 * \text{MaterialPVC} + 0.660 * \text{SurfaceEasement} + 0.998 * \text{SurfaceHighway} + 1.163 * \text{SoilTypeClay} + 2.168 * \text{SoilTypeLoam} + 1.335 * \text{SoilTypeRock} + 0.507 * \text{CorrosivityConcreteHigh} + 1.029 * \text{CorrosivityConcreteLow} + 2.731 * \text{CorrosivitySteelHigh} \quad \text{Eq. 5-3}$$

Where:

Pr (C =3) is the probability of sanitary sewer pipe condition dependent variable being condition 3 in relative to condition 5.

Pr (C = 5) is the probability of reference category condition 5

.

Diameter, Age, Slope, Depth, Length, Material, Surface, Soil Type, and Corrosivity are independent variables that influence sanitary sewer pipe condition.

Equation 5-4 represents a MLR model developed to predict sewer pipe condition 4.

$$g_4(x) = \ln \left[\frac{Pr(C=4)}{Pr(C=5)}\right] = \mathbf{0.951} * \textbf{Diameter} + 0.998 * \text{Age} + 0.853 * \text{Slope} + 0.925 * \text{Depth} + 1.00 * \text{Length} + 0.663 * \text{pH} + 0.812 * \text{MaterialCONC} + 0.489 * \text{MaterialPVC} + 1.073 * \text{SurfaceEasement} + 1.503 * \text{SurfaceHighway} + 0.134 * \text{SoilTypeClay} + 1.298 * \text{SoilTypeLoam} + 1.223 * \text{SoilTypeRock} + 0.843 * \text{CorrosivityConcreteHigh} + \mathbf{10.377} * \textbf{CorrosivityConcreteLow} + 0.719 * \text{CorrosivitySteelHigh} \quad \text{Eq. 5-4}$$

Where:

Pr (C =4) is the probability of sanitary sewer pipe condition dependent variable being condition 4 in relative to condition 5.

Pr (C = 5) is the probability of reference category condition 5

Diameter, Age, Slope, Depth, Length, Material, Surface, Soil Type, and Corrosivity are independent variables that influence sanitary sewer pipe condition.

Equations 5-5, 5-6, 5-7, 5-8, and 5-9 shows probabilities of sewer pipes conditions 1, 2,3, 4, and 5 occurring.

$$\Pr(\mathrm{C} = 1|\mathrm{x}) = \frac{e^{g_1(x)}}{1+e^{g_1(x)}+e^{g_2(x)}+e^{g_3(x)}+e^{g_4(x)}} \quad \text{Eq. 5-5}$$

$$\Pr(\mathrm{C} = 2|\mathrm{x}) = \frac{e^{g_2(x)}}{1+e^{g_1(x)}+e^{g_2(x)}+e^{g_3(x)}+e^{g_4(x)}} \quad \text{Eq. 5-6}$$

$$\Pr(\mathrm{C} = 3|\mathrm{x}) = \frac{e^{g_3(x)}}{1+e^{g_1(x)}+e^{g_2(x)}+e^{g_3(x)}+e^{g_4(x)}} \quad \text{Eq. 5-7}$$

$$\Pr(\mathrm{C} = 4|\mathrm{x}) = \frac{e^{g_4(x)}}{1+e^{g_1(x)}+e^{g_2(x)}+e^{g_3(x)}+e^{g_4(x)}} \quad \text{Eq. 5-8}$$

$$\Pr(\mathrm{C} = 5|\mathrm{x}) = \frac{1}{1+e^{g_1(x)}+e^{g_2(x)}+e^{g_3(x)}+e^{g_4(x)}} \quad \text{Eq. 5-9}$$

5.3 ANN Model

Development of neural network model included preparing data as inputs and output, training and testing the model. The data sets were randomly divided as follows: Training (85%), and Testing (15%). During training, the network was fed with inputs. The network then generated output. The network checked the results with correct answers and made corrections to internal connections while minimizing the errors. During testing, inputs were paired with outputs were provided. Testing is the same as training. Validation of the model was conducted after the model was developed. In model validation, only inputs were used to predict the sewer pipes condition.

5.3.1 ANN Data Processing Software Selection

The data was stored in Microsoft Excel. It was processed and grouped into input and output variables. The created input data comprised of independent variables. These variables included pipe material, diameter, age, slope, depth, surface condition, soil type, corrosivity concrete, corrosivity steel, and pH. The created output data comprised of the dependent variable. The dependent variable was sewer pipes condition rating.

The BrainMaker, a commercially available Simulator distributed by California Scientific Software was used to develop the neural network model. BrainMaker used data stored in neural networks files. The neural network files were Definition (.def), Fact (.fct), and Testing (.tst). The definition file is a text file that was created to be read by BrainMaker. The definition file is given an extension.def. The definition file explains all the information that the BrainMaker needs to know such as the number of neurons, type of data, and output. The fact file provides data into BrainMaker. There are fact files for training, testing, and validation (running). The fact file for training, testing, and running have extensions ". fct", ". tst", and ". in", respectively. During training, BrainMaker used data stored in the training fact file and the definition file.

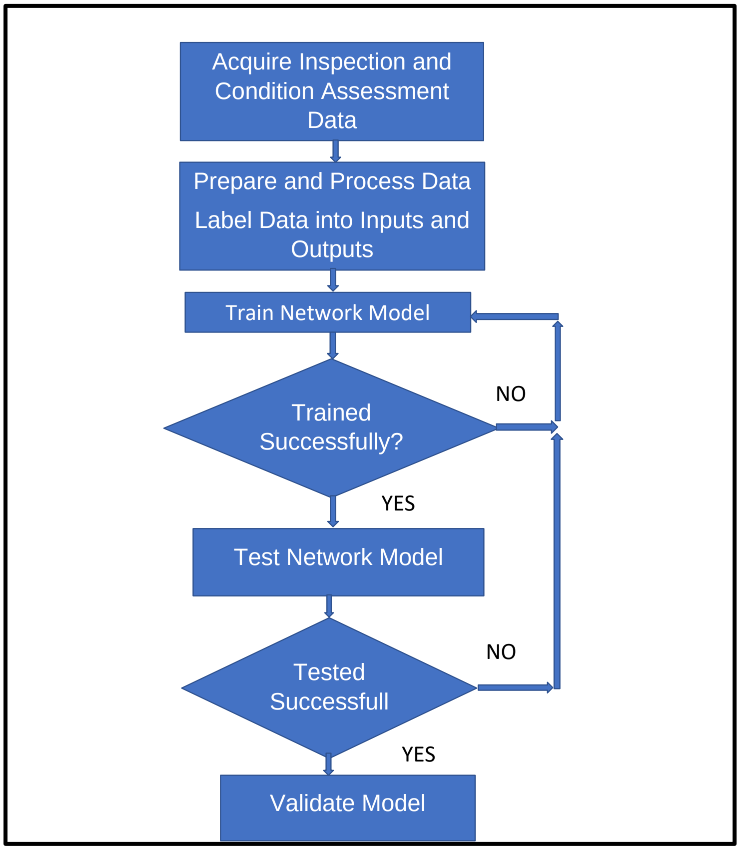

Figure 5-29 ANN Model Development Procedure

Source: Kulandaivel, (2004)

Figure 5-1 shows the process of developing the ANN model. The following are the steps that were used to develop the neural network model.

- Acquire inspection and condition assessment data.

- Prepare and process data. Categorical data were split into categories. The data was labeled into inputs and outputs.
- Train network model. Twelve different neural network architectures were trained and obtained the optimal architecture with the lowest errors.
- Tested network model. The architectures were tested using 15% of the data.
- Validated model. New data was used as a case study to validate the use of the model.

Datasets randomly divided: Training (70%), and Testing (30%) for IBM SPSS Neural network Software and Training (85%), and Testing (15%) for Brain Maker Neural Network Software, California Scientific Software.

Table 5-29 Training and Testing Data

Data	Data Evaluation	N	Percent
Sample	Training	1850	70%
	Testing	764	30%
Valid	--	2614	100%
Excluded	--	2	--
Total	--	2616	--

Table 5-8 shows datasets divided into Training (70%) and Testing (30%) using IBM SPSS Neural network Software.

The backpropagation algorithm was used in training the neural network model. Training involved presenting inputs to the network. The network uses the input variables to establish a relation between the inputs and outputs that are placed in NetMaker. The dataset is split randomly into training (85%) and testing (15%) facts using the NetMaker preferences. NetMaker puts percentage of the data to be trained and tested. The data is processed in Netmaker and saved in the BrainMaker file. Figure 5-2 shows prepared data in NetMaker that is ready for modeling by the BrainMaker software.

NetMaker - Sewer2.Dat

File Column Row Label Number Symbol Operate

	Annote	Input	Input	Input	Input	Input	Input	Pattern	Input	Input	Input	Input	Input
	ID	Diameter	Age	Slope	Depth	Length	pH	Condition	Material	Surface	SoilType	CorrConc	CorrSteel
1	318	10	59	0.1	6	575.35	7.5	1	Mater3	Surfa4	SoilT3	CorrC3	CorrS2
2	3255	8	22	0.5	7	154.1	7.9	1	Mater2	Surfa4	SoilT3	CorrC1	CorrS2
3	930	8	14	0.61	15	581.22	7.9	1	Mater2	Surfa2	SoilT3	CorrC1	CorrS2
4	1988	12	65	0.5	5	365.97	7.9	3	Mater3	Surfa3	SoilT3	CorrC1	CorrS2
5	2171	8	18	2.3	8	504.53	8.2	1	Mater2	Surfa2	SoilT4	CorrC1	CorrS2
6	2596	8	65	3.6	8	78.94	7.9	1	Mater1	Surfa4	SoilT3	CorrC1	CorrS2
7	1034	10	14	0.25	7	238.38	6.5	1	Mater2	Surfa4	SoilT1	CorrC2	CorrS1
8	1760	8	20	0.6	7	118.03	7.9	2	Mater1	Surfa2	SoilT3	CorrC1	CorrS2
9	1615	8	24	7	6	98.79	8.2	1	Mater2	Surfa2	SoilT3	CorrC1	CorrS2
10	3343	8	13	573	10	375.7	8.2	1	Mater2	Surfa3	SoilT3	CorrC1	CorrS2
11	1597	24	8	0.3	8	551.62	7.9	4	Mater1	Surfa2	SoilT3	CorrC1	CorrS2
12	2511	10	27	1.8	5	129.95	8.2	1	Mater2	Surfa4	SoilT4	CorrC1	CorrS2
13	385	10	60	0.3	7	311.56	8.2	1	Mater1	Surfa2	SoilT3	CorrC1	CorrS2
14	448	15	42	0.54	10	296.51	8.2	1	Mater3	Surfa2	SoilT2	CorrC1	CorrS2
15	3359	54	65	0.2	10	555.85	8.2	3	Mater1	Surfa2	SoilT2	CorrC1	CorrS2
16	258	12	57	0.01	5	372.85	8.2	1	Mater3	Surfa4	SoilT3	CorrC1	CorrS2
17	3188	15	39	0.1	10	752.5	7.9	1	Mater2	Surfa2	SoilT3	CorrC1	CorrS2
18	297	42	57	0.14	8	930.82	8.2	5	Mater1	Surfa4	SoilT2	CorrC1	CorrS2
19	83	8	39	0.4	5	673.57	6.8	1	Mater2	Surfa4	SoilT1	CorrC2	CorrS2
20	1207	8	22	3.8	7	108.3	8.2	1	Mater2	Surfa3	SoilT3	CorrC1	CorrS2

Figure 5-30 View of NetMaker Data Processing

NetMaker created three files, namely, definition (.def), training (.fct), and testing (.tst). The training and testing files were saved in the BrainMaker software.

5.3.3 ANN Architecture

The neural network architecture comprises there (3) layers, namely, input, hidden, and output layers. The hidden layer is known as hidden neurons. The number of `neurons should be sufficient to provide optimal performance in the modeling prediction process. Too few or too many neurons will not enable the network to acquire knowledge that can be generalized for future predictions. There are 3 ways of determining the ideal number of hidden neurons. Equations 5-1 and 5-2 show 2 ways of calculating the number of hidden neurons.

$$Number\ of\ hidden\ neurons = \frac{of\ Data\ sets - Outputs}{C(\#Input + \#Output + 1)} \qquad \text{Eq. 5-1}$$

Where, C = 2-5

$Number\ of\ Neurons = (\frac{\#\ Inputs + \#\ Outputs)}{2})$ Eq. 5-2

Equation 5-2 was suggested by BrainMaker Manual. In this study, Equation 5-2 was used to calculate the number of neurons.

= (22+1)/2 =12 Neurons

In this study, starting at one (1) neuron experiment was conducted from 1 neuron to 15 neurons. The neural network with the least testing error was selected.

Table 5-30 Training and Testing Errors

Model #	Architecture	RMS Training	RMS Testing
1	22-4-1	0.3089	0.2745
2	22-5-1	0.3165	0.2857
3	**22-6-1**	**0.2860**	**0.2620**
4	22-7-1	0.3001	0.2647
5	22-8-1	0.3048	0.2720
6	22-9-1	0.3009	0.2662
7	22-10-1	0.3010	0.2751
8	22-11-1	0.3021	0.2716
9	22-12-1	0.3001	0.2730
10	22-13-1	0.3001	0.2695
11	22-14-1	0.3002	0.2712
12	22-15-1	0.3001	0.2700

Model # 3 was found to be optimal with the least training and testing errors. Model # 3 was chosen for model development.

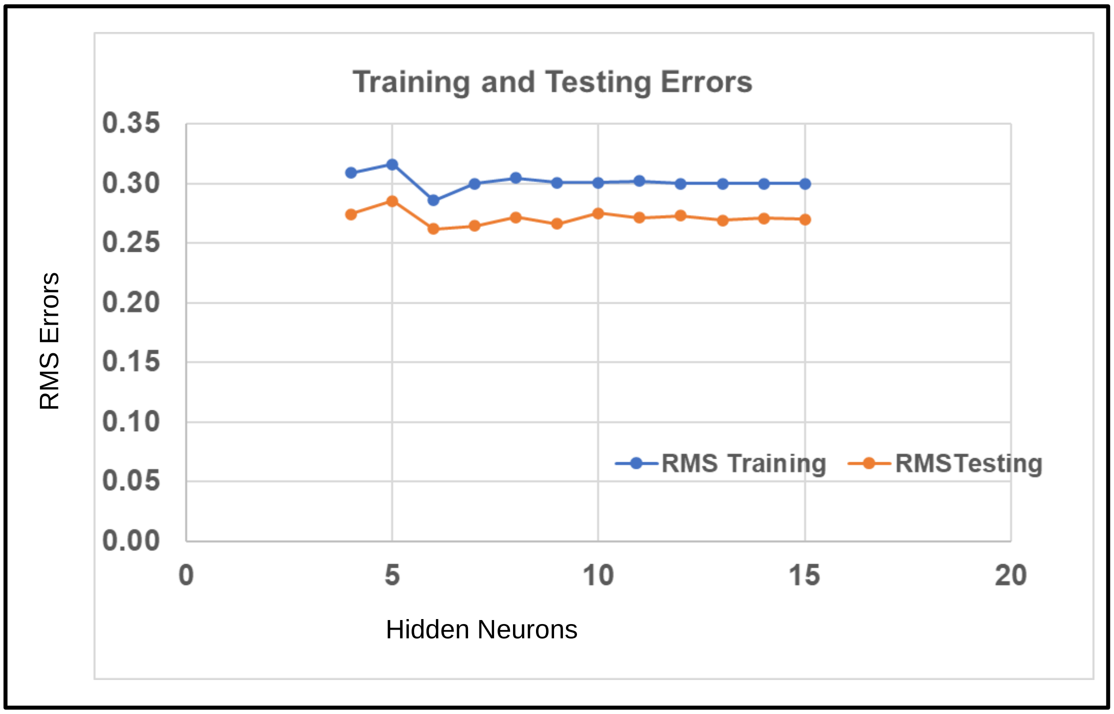

Figure 5-31 Training and Testing Errors for different Architectures

Figure 5-3 shows a graph that represents training and testing errors. In the graph, the optimal neurons were selected. It was shown that the optimal neurons were 6. The training and testing errors show a consistent trend.

The neural networks structure of the model was presented in Figure 5-4. The structure was comprised of the input layer, hidden layer, and output layer. The input layer was comprised of independent variables. The input variables used were diameter, age, slope., depth, length, pH, material, surface condition, soil type, corrosivity concrete, and corrosivity steel. The hidden layer was comprised of the 6 neurons. The output layer consisted of sewer pipe condition s 1, 2, 3,4, and 5. Like it is shown for diameter parameter there is a network of a relation between the input, hidden, and output layers. Similar relationships apply for all other parameters. This illustrates how ANN is working.

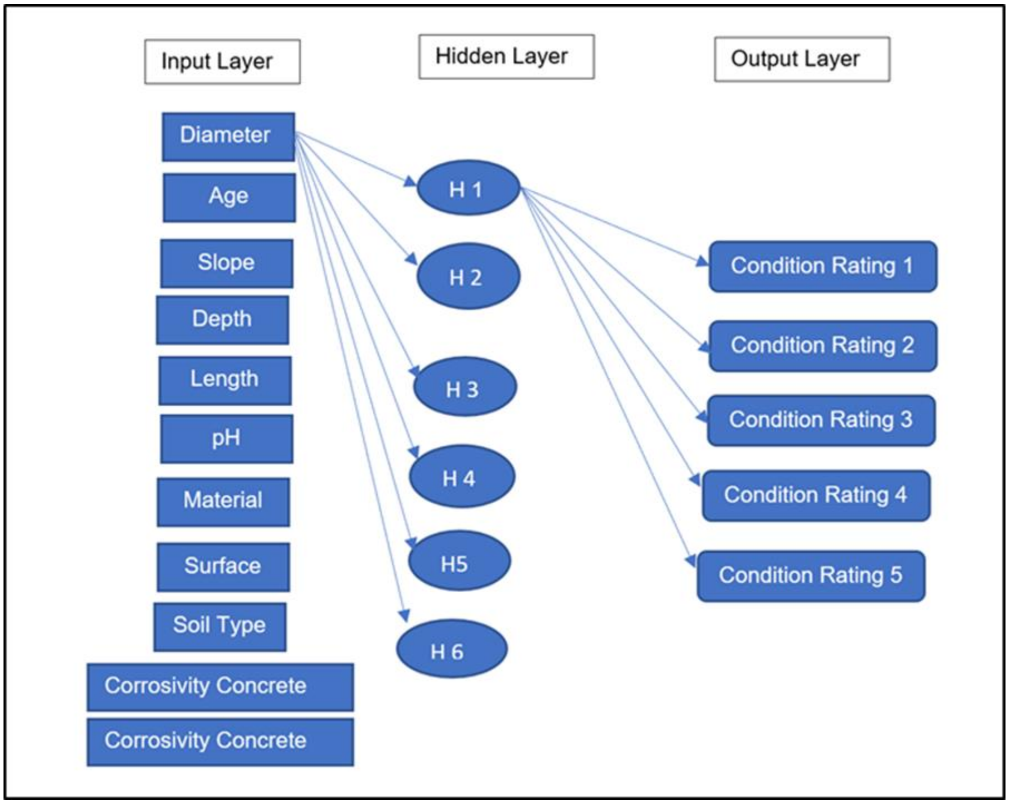

Figure 5-32 ANN Structure

Figure 5-4 shows a network with 3 layers.

5.3.4 ANN Model Development

In BrainMaker, the network size of the optimal neurons was set at 6. Various training and testing tolerances were tested starting at 0.1 and 0.1, respectively. Training and testing tolerance of 0.3 and 0.3, respectively, was selected to be optimal having the lowest training and testing errors. The training was achieved by the trial-and-error method with weights randomly taking numbers in training the model. The training was stopped when the neural network reached the lowest training and testing errors. The training algorithm helps distribute the error to arrive at the minimum error. The information moves forward in the network to predict the output. While minimizing the error achieved through several iterations, the backpropagation algorithm redistributes the error and adjusts the weights. A complete cycle of training is called ‘epoch’.

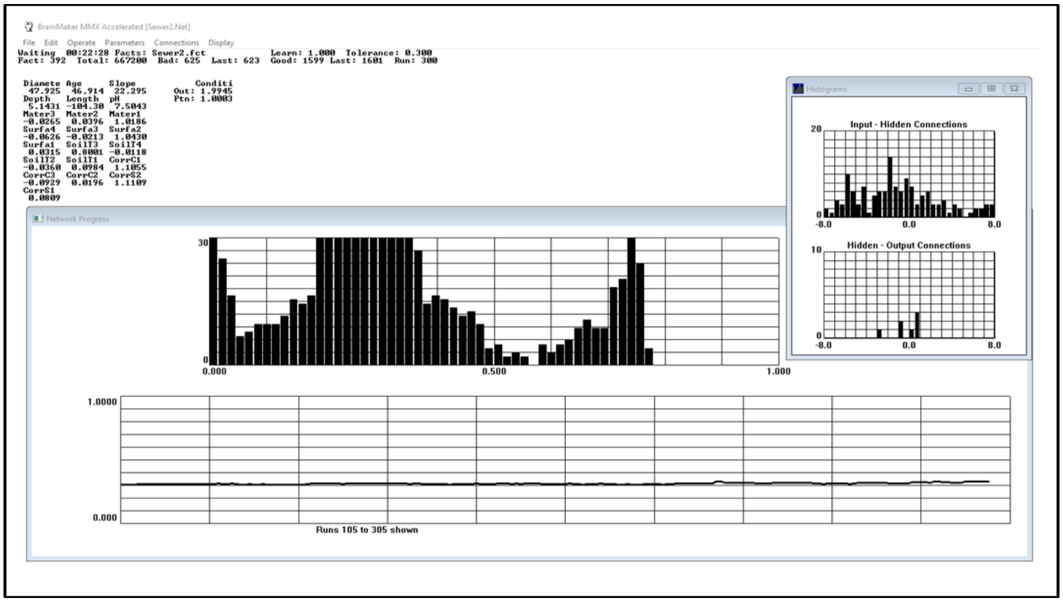

Figure 5-33 Neural Network Training Progression Snapshot

Figures 5-6 and 5-7 show results of average and RMS errors and testing results of model #3.

NetMaker - Sewer2.Sts

File Column Row Label Number Symbol Operate

	NotUsed	NotUsed	NotUsed	NotUsed	NotUsed	NotUsed	NotUsed	NotUsed	NotUsed	NotUsed	NotUsed
	Run	TotFacts	Good	Bad	BadOutputs	TotalBad	Learn	Tolerance	AvgError	RMSError	hh:mm:ss
135	135	300240	1665	559	559	82191	1	0.3	0.2519	0.3048	00:11:19
136	136	302464	1644	580	580	82771	1	0.3	0.2597	0.3063	00:11:20
137	137	304688	1652	572	572	83343	1	0.3	0.2632	0.3077	00:11:22
138	138	306912	1645	579	579	83922	1	0.3	0.2612	0.3067	00:11:23
139	139	309136	1656	568	568	84490	1	0.3	0.2625	0.3065	00:11:24
140	140	311360	1683	541	541	85031	1	0.3	0.2597	0.3068	00:11:26
141	141	313584	1684	540	540	85571	1	0.3	0.2604	0.308	00:11:27
142	142	315808	1678	546	546	86117	1	0.3	0.2585	0.3059	00:11:28
143	143	318032	1686	538	538	86655	1	0.3	0.2657	0.3111	00:11:29
144	144	320256	1645	579	579	87234	1	0.3	0.2723	0.3141	00:11:31
145	145	322480	1630	594	594	87828	1	0.3	0.2801	0.317	00:11:32
146	146	324704	1631	593	593	88421	1	0.3	0.2763	0.314	00:11:34
147	147	326928	1623	601	601	89022	1	0.3	0.2776	0.314	00:11:35
148	148	329152	1607	617	617	89639	1	0.3	0.2782	0.314	00:11:36
149	149	331376	1617	607	607	90246	1	0.3	0.2808	0.3147	00:11:38
150	150	333600	1604	620	620	90866	1	0.3	0.2776	0.3141	00:11:39

Figure 5-34 Average and RMS Errors

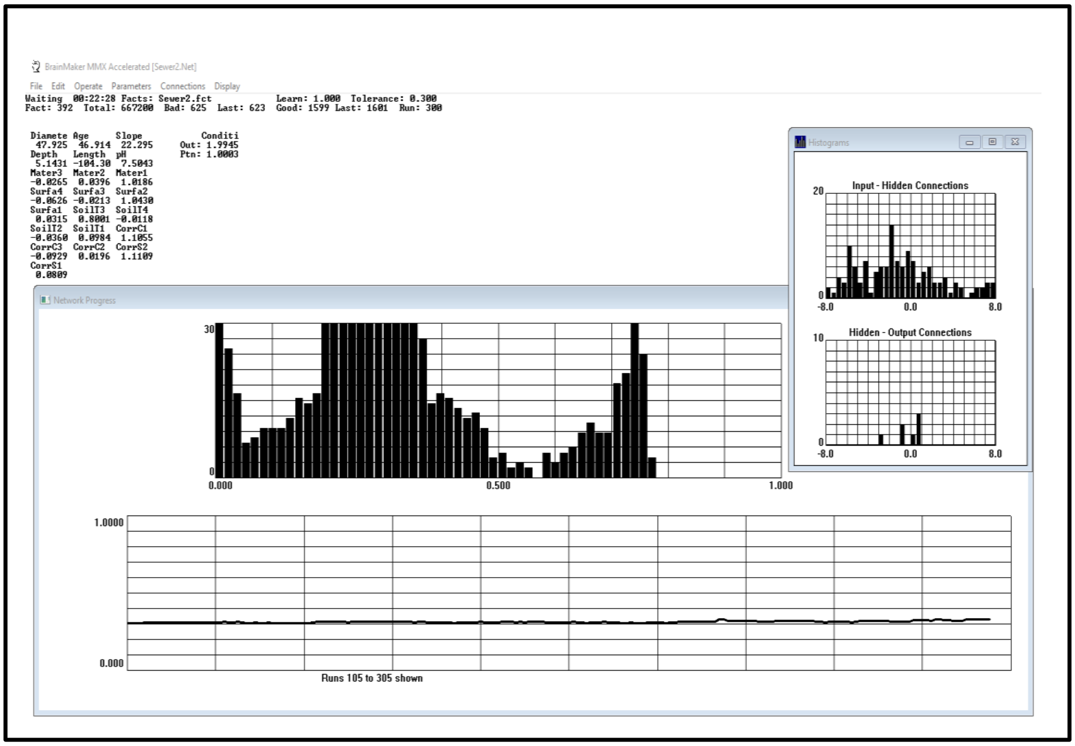

Figure 5-35 Testing Results of Model #3

Table 5-31 Summary of Training and Testing Results of Model #3

Total Facts	Good	Bad	Tolerance	Average Error	RMS Error
Training Configuration					
2224	1599(72%)	625 (28%)	0.3	0.2519	0.3048
Testing Configuration					
392	334(85%)	58 (15%)	0.3	0.227	0.2823

Table 5-10 shows that the model learned 72% of the facts and predicted 85% of the testing factors.

Figure 5-8 shows a tested neural networks model. Figure 5-8 shows that the model underpredicted sewer pipe conditions 4 and 5. In the future research the model will need to be further fine-tuned for sewer pipe conditions 4 and 5.

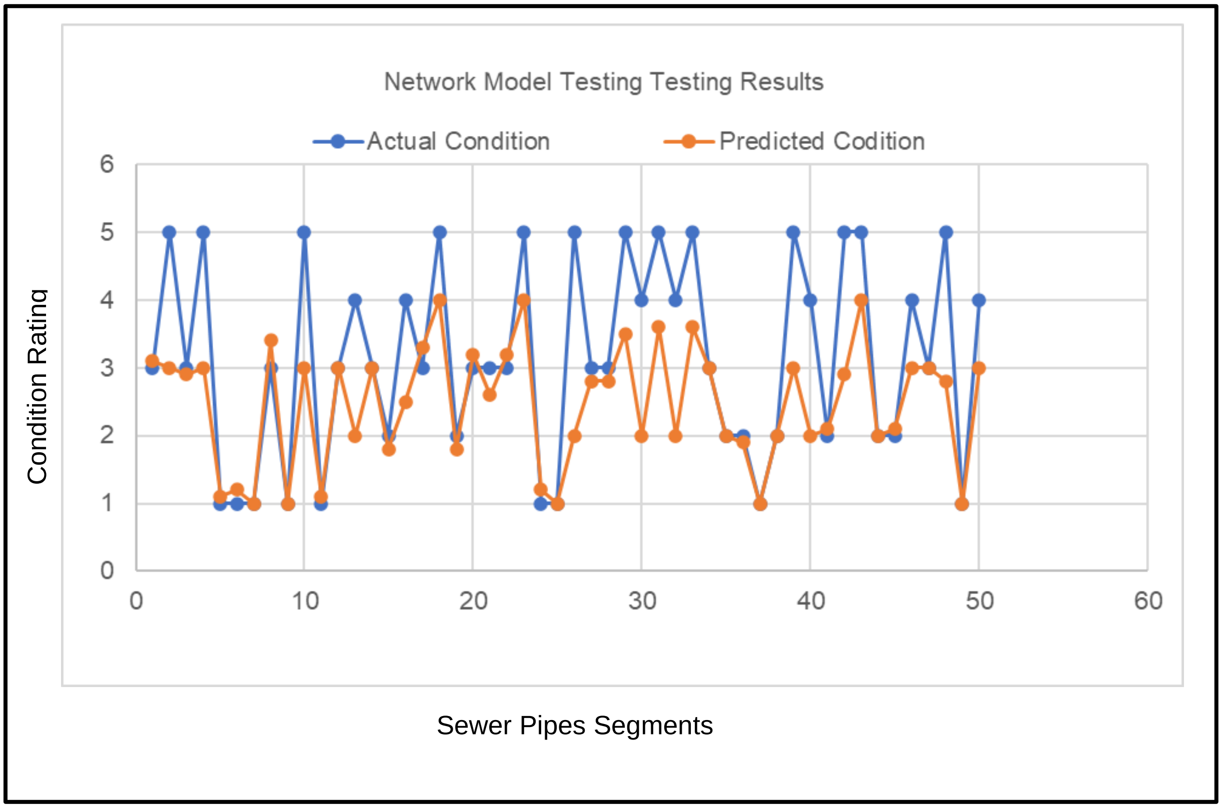

Figure 5-36 Plot showing Testing Results of Actual and Predicted Sewer Condition

5.4 Chapter Summary

In this chapter MLR model was developed using 80 % of randomly selected datasets. The MLR model were validated by using 20 % of the remaining randomly selected datasets. The ANN model was developed through training, validation, and testing.

Chapter 6

Results and Discussions

6.1 Introduction

In this chapter results and discussions of MLR and ANN models are presented. The accuracy of the models was discussed using a classification table, sensitivity, and 1-specificity curves, and ROC curve, Model influence variables were presented. The significance coefficients of the models are used to point out the variables that influence sewer pipe conditions. The confidence level used in the data analysis was 95%. The validation of MLR was discussed. The justification of the results was discussed.

6.2 Performance of the Models

6.2.1 MLR Model Classification Table

Table 6-1 shows that overall, 75% of the sewer pipe conditions were correctly predicted by the multinomial logistic regression model. Prediction of sewer pipe condition 1 was 97% correctly with 3 percent incorrectly predicted. This demonstrated that the model had a high accuracy in predicting condition 1. According to the classification table conditions 2, 3, 4, and 5 were 0%, 28%, 4%, and 14% correctly estimated. The prediction was consistent with the available datasets that were analyzed. From Chapter 4, section 4.6 condition rating score, sewer condition 1 datasets were 73% of the sewer pipe segments.

Table 6-32 Classification Table

Observed	Predicted											
	1		2		3		4		5		Percent	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
1	1846	--	--	--	--	44	--	--	--	11	97%	3%
2	--	104	--	--	--	2	--	--	--	2	0%	100%
3	--	233	--	--	93	0	--	--	--	10	28%	72%
4	--	66	--	--	--	4	3	--	--	7	4%	96%
5	--	138	--	--	--	27	--	--	26	--	14%	86%
Overall Percentage											75%	25%

Sewer pipes datasets condition 2, 3, 4, and 5 were 4%, 12%, 3%, and 7%, respectively of the sewer pipes segments. This explains why the percent prediction correct rate was low in conditions 2, 3, 4, and 5 compared to condition 1.

6.2.2. ANN model Performance

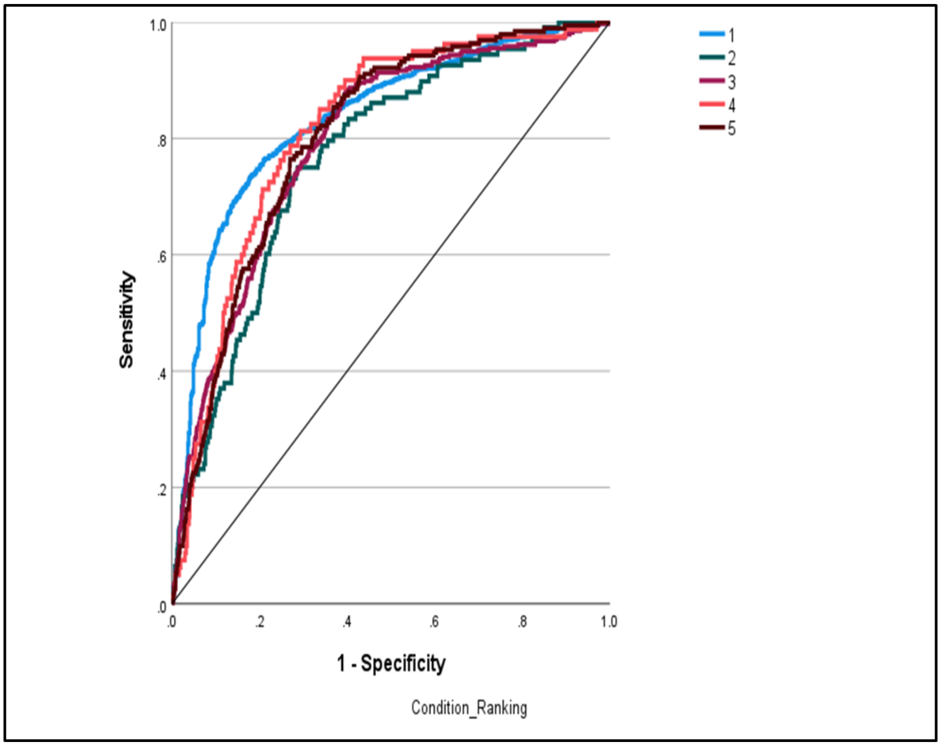

Figure 6-37 Plot of Model Sensitivity and 1-Specificity

Figure 6-1 is derived from the metrics of measuring model performance, equations 3-35, 3-36, 3-37, and 3-38.

Table 6-33 ROC Curve

Condition	Area Under Curve
1	0.833
2	0.768
3	0.794
4	0.815
5	0.802

The area under the ROC curve provided in Table 6 -1 demonstrated the performance of the ANN model. According to Hosmer et al., 2013, when the area is close to one (1), it refers to the perfect model. When the area is greater than 0.7, it implies an acceptable model. From Table 6-2, it was shows that the model is acceptable to be used in the prediction of sewer pipes conditions.

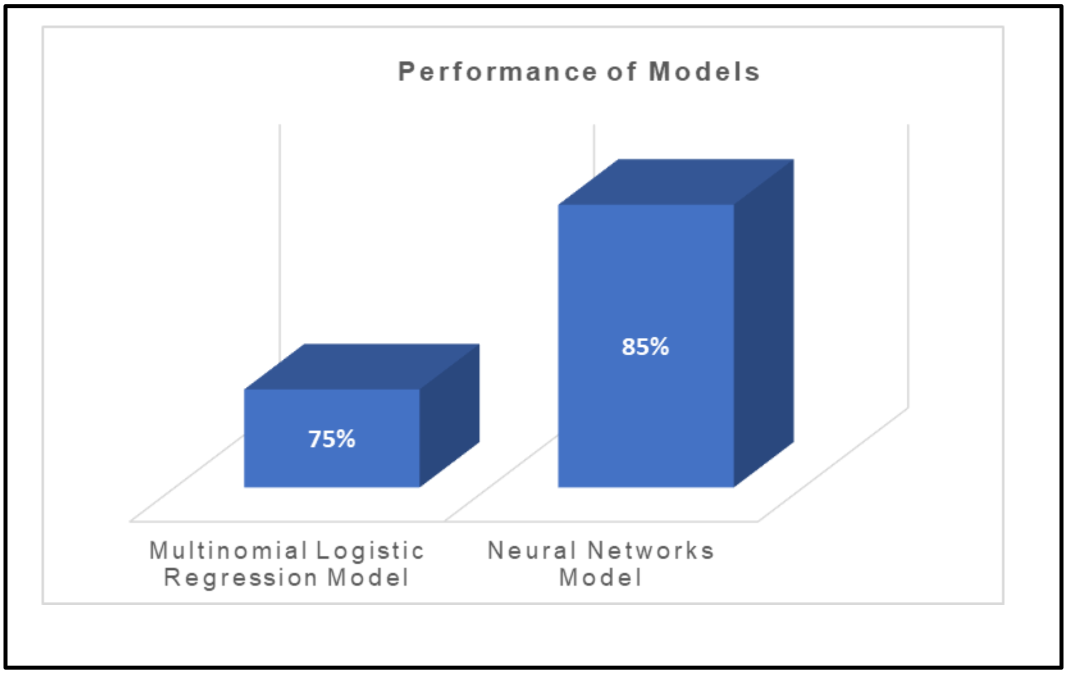

Figure 6-38 Comparison of Model Performance

The accuracy of MLR and ANN model was compared. The ANN model was found to be better in predicting the sewer pipe condition compared to the MLR model. Figure 6-2 shows that the prediction accuracy of the logistic regression model was 75% and that of the ANN model was 85%.

6.3 Discussions

This study was set to develop MLR and ANN sanitary sewer condition assessment prediction models. Observed datasets independent and dependent variables were utilized to develop the models. The independent variables that influenced the condition rating of the sewer pipes were presented in order of significance in Figure 6 -3. Age, depth, slope, diameter, and depth which are physical factors were found to be the most important predictors compared to environmental factors.

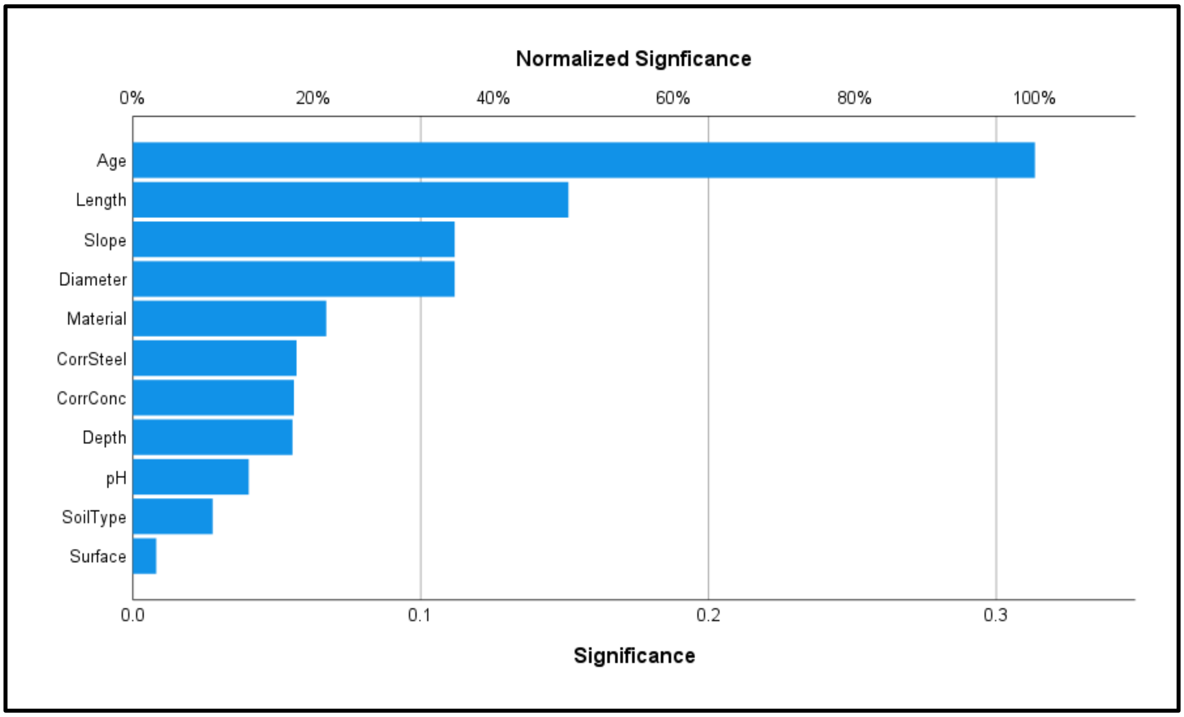

Figure 6-39 Influence of Variables

The significant variables that influence sewer pipes condition rating were diameter, age, length, pipe material (CONC), soil type (Loam), soil type (Clay), corrosivity concrete (High), and corrosivity concrete (Low). Influencing and non-influencing Variables were determined by significance Value ($P<0.05$) based on a 95% Confidence Level.

Pipe diameter was found to be significant in sewer conditions 1, 2, and 4 with a significant level of 95% the significant value $P<0.05$. In condition 1, 2, and 4, the value P was 0.001, 0.000, and 0.008, respectively. Age was one of the significant variables. Age was found to be significant in sewer pipe conditions 1, 2, 3, and 4 with a significant value of $P<0.05$. The significant P value in conditions 1, 2, and 3 were 0.000, 0.001, and 0.000, respectively. Pipe length was a significant factor in sewer pipe condition 1. The significant value of the pipe length was 0.000. Pipe material was a significant variable in the prediction model in sewer pipe conditions 2 and 3. The significant value of P was 0.025 and 0.001, respectively. The PVC material had most of the pipes that were in good condition.

Loam and clay soil types were found to be significant predictors. The significant value P was <0.05 in condition 4. The significant values were 0.05 and 0.05, respectively. Corrosivity Concrete (High) and corrosivity Concrete (Low) variables are significant in the sewer pipes condition. The significant value P was <0.05 was found to be in condition 4. The significant values were 0.036 and 0.031, respectively. Depth, slope, corrosivity steel, and pH variables were found not to be significant in predicting sewer pipes condition.

6.4 Justification of Results

The model results were consistent with similar studies conducted by other authors. Significant P-Value and ODDs ratio (Exp(B)) generated in this study were compared with the results of other authors. In this dissertation, the diameter was found to be significant in conditions 1, 2, and 4 with P values of 0.001, 0.000, and 0.008, respectively. The Exp(B) was 0.978, 0.923, and 0.951 for sewer pipes condition 1, 2 and 4, respectively. The Wald for sewer conditions 1, 2, and 4 were found to be 10.390, 12.975, 7.020, respectively. The diameter of the pipe being significant was consistent with Tscheikner-Gratl et al., (2014), Fuamba and Lubini (2011), and Laakaso et al., (2018).

Age was another important factor that influences sewer pipes condition. In this dissertation, age was found to be significant in predicting sewer pipe conditions. The P-value for age factor in conditions 1, 2, and 3 were 0.000, 0.001, and 0.000, respectively. Wald was 81.139, 11.57, and 15.222 for conditions 1, 2, and 3. The Exp(B) was 0.845, 0.968, and 0.974 for conditions 1 and 2, respectively. Tscheikner-Gratl et al. (2014), Fuamba and Lubini (2011), and Laakaso et al., (2018) found age to be a significant factor.

It was revealed that length was a significant factor in influencing the sewer pipes condition 1. The P-value was 0.000, Wald and Exp (B) were 20.852 and 0.999, respectively. This was confirmed by a study conducted by Tscheikner-Gratl et al., (2014), Davies et al. (2001), Fuamba and Lubini (2011), Laakaso et al., (2018). Pipe material (Conc) was a significant factor in influencing sewer pipe conditions. In conditions 2, P-value, Wald, and Exp(B) were 0.025, 5.030, and 0.472, respectively. In condition 3, P-value, Wald, and Exp(B) were 0.001, 10.828, and 2.089, respectively. Pipe material was found to be insignificant in conditions 1, and 4. This is consistent with Fuamba and Lubini (2011).

Soil type (Clay), soil type (Loam), and soil type rock were illustrated to be significant in sewer conditions 2,3, and 4. In condition 2 for clay soil P-value, Wald, and Exp(B) were 0.05, 3.89, and 9.12. In condition 3 for clay soil P-value, Wald, and Exp(B) were 0.021, 0.021, and 1.153. Similarly for loam soil, the P-value, Wald, and Exp(B) were 0.05, 3.89, and 9.12, respectively and for loam soil, the P-value, Wald, Exp(B) for soil type (Rock) was 0.018, 0.0.18, and 0.864, respectively. Laakaso et al., (2018) found that soil type is a significant variable in the prediction of sewer pipe condition. In condition 4, soil type clay and loam have Wald of 3.89 and 3.93, P-value 0.05 and 0.05, and Exp(B) of 9.12 and 9.68, respectively.

Corrosivity concrete was found to be significant in influencing pipe conditions. The P-value, Wald, and Exp(B) were 0.036, 4.411, and 0.096, respectively. Laakaso et al., (2018) found corrosivity to be of very high significance. The model prediction accuracy in this dissertation was compared with other authors. The results are presented in Table 6-3.

Table 6-34 Prediction Accuracy

Model	Author	Prediction Accuracy
Multinomial Logistic Regression	Salman and Salem (2012)	52%
	Malek (2019)	65%
	Laakaso et al., (2018)	62%
	Sousa et al., (2014)	65%
	Chughtai and Zayed (2008)	72%
Neural Network	Sousa et al., (2014)	72%-82%

6.5 Chapter Summary

In this chapter results of MLR and ANN models were discussed. Accuracy of the models was discussed using classification table, sensitivity and 1-specificity curves, and ROC curve, and Model influence variables were presented. The justification of the results was discussed.

Chapter 7

Conclusions and Recommendations for Future Research

7.1 Conclusions

In this dissertation, MLR and ANN models were developed to predict sanitary sewer pipe conditions. The models were developed, validated, and tested in prediction sewer pipe condition scores to prioritize pipes to be rehabilitated and or replaced and further condition assessment. The developed models added knowledge in the tools used to predict sanitary sewer pipes condition. Predicting and knowing the sanitary sewer pipes condition rating score would be beneficial to policymakers and sanitary sewer utilities managers in prioritizing rehabilitation and/or replacement of sanitary sewer pipes. The MLR was built with 80% of the randomly selected dataset. The randomly remaining 20% of the data was utilized in the validation of the model. The ANNs model was trained, validated, and tested. The feed-forward network with a backpropagation learning algorithm was employed. Based on the model results, the significant physical factors influencing sanitary pipes condition rating included diameter, age, pipe material, and length. Soil type was the environmental factor that influenced sanitary sewer pipes failure.

The accuracy of the performance of the MLR and ANN was found to 75 % and 85%, respectively. About this research's main objective, it was determined that the use ANN model provided an accurate prediction of sanitary sewer pipes condition by testing the results of the condition rating values.

The significance of the independent variables was found to be in the following order. Age (100%), Diameter (80%), Slope (62%), Length (62%), Flow (60%), pH (40%), Corrosivity Steel (38%), Soil Type (36%), Depth (35%), Pipe Material (25%), Surface Condition (22%), and Corrosivity (20%). With 95% significance level the following variables were found to be significant. Diameter (Pvalue=0.001), age (Pvalue=0.000), length (Pvalue=0.000), pipe material(CONC) (Pvalue=0.001)), soil type (Loam) (Pvalue=0.05), soil type (Clay) (Pvalue=0.001), corrosivity concrete (Pvalue=0.001), and corrosivity concrete (Low) (Pvalue=0.001).

7.2 Practical Applications

Prediction models may be instrumental to sanitary sewer utilities managers in the decision-making process in rehabilitation and replacement of sewer pipes. Sanitary sewer condition assessment and data collection through CCTV inspection can be costly. Due to inaccessibility and inadequate funding, only about one third of sanitary sewer systems are inspected every 5 years. Prediction models can assist in expediting the evaluation of the condition rating of sewer pipes using independent variables. City Engineers can use existing data and use one of the models to predict the condition of sewer pipes underground. Using the existing data, the MLR and ANN models can predict the sewer pipes conditions 1 through 5.

Steps of the practical application of the MLR Model.

1. Obtain following variables for a specific sewer pipe segment: Pipe material, diameter, age, slope, depth, surface condition, soil type, corrosivity concrete, corrosivity steel, and pH.
2. Substitute the values of the variables of the sewer pipe segment in equations 5-1, 5-2, 5-3, and 5-4.
3. Apply the results from Equations 5-1, 5-2,5-3, and 5-4 in equations 5-5, 5-6, 5-7, 5-8, and 5-9 to determine the probability of that pipe segment being in one of the 5 conditions (1, 2, 3, 4, and 5). The probability will be between 0 and 1. A close value to 1 will mean the sewer pipe is in the given condition represented by the equation. The value being close to zero (0) it means that the sewer pipe is not in condition represented by the equation.
4. Compare the values calculated in equations 5-5, 5-6, 5-7, 5-8, and 5-9. The equation which gives the highest value, will imply the condition rating of the sewer pipe segment. The highest value indicates the condition rating of the pipe.
5. Decide to replace or not to replace the sewer pipe or recommend condition assessment upon determining the condition of the sewer pipe.

Steps of the Practical Application of ANN model

1. Obtain the following data for sewer pipe segments with known condition as a pattern: Pipe material, diameter, age, slope, depth, surface condition, soil type, corrosivity concrete, corrosivity steel, and pH.

2. Place the data in an excel spreadsheet.
3. Label variables as input, training pattern, annotation or not used in Netmaker.
4. Training the network using 85% of data with known sewer pipe condition rating.
5. Test the network using 15% of data with known sewer pipe condition rating.
6. Get data of a sewer pipe segment without known condition rating and run it to generate the condition of a sewer pipe.
7. Determine the condition rating of the pipe segment from the generated from the run.
8. Pipe segments with condition ratings 4 and 5 are flagged and selected to be replaced.
9. Decide to either replace or not replace the sewer pipe or recommend condition assessment.

7.3 Recommendations for Future Research

This dissertation raised several important points to be considered for future research as listed below:

- There is a need of utilizing more datasets to increase the accuracy of the prediction models. Data collected for analysis should include more uniformly distributed number of observations in every condition.
- More training and testing of the ANN model are needed to validate its prediction strength for conditions 4 and 5.
- The data for this dissertation was collected from the City of Dallas. Other cities should be included in future studies and results compared with results of this research.
- Excluded pipe material types could be included for further model development.
- These models can be improved by utilizing wastewater type and volumetric flow rate variables in predicting and comparing the MLR and ANN models.
- Consequences of Failure should be considered in developing these prediction models.

Appendix A

Abbreviations

AC - asbestos-cement

ASCE – American Society of Civil Engineers

ANN - Artificial Neural Network

BPNN - Backpropagation Neural Network

CBD - central business district

CCTV - Closed-circuit television

COF – Consequence of Failure

CP - concrete pipe

CUIRE - Center for Underground Infrastructure Research and Education

DBMS - Database management system

DBN - Dynamic Bayesian network

DFW - Dallas Fort Worth

DI - ductile iron

DSS - Decision support system

DWU - Dallas Water Utilities

FNR - False Negative Rate

FRP -Fiberglass Reinforced Plastic

GIS - Geographical Information System

HDPE - High Density Polyethylene

-2LL - -2 Log likelihood

LF - Linear Feet

LOF – Likelihood of Failure

in - inch

MCS - Monte Carlo simulation

MLR – Multinomial Logistic Regression

MSE - mean squared error

NASSCO - National Association of Sewer Service Companies

O&M - Operation and maintenance

OMR – Operational Maintenance Rating

OSR - Operational Structural Rating

PACP - Pipeline Assessment and Certificate Program

PCCP - prestressed concrete cylinder pipe

PCP - polymer concrete pipe

PE – polyethylene

PoF - Probability of failure

PVC - Polyvinyl Chloride Pipe

R - Coefficient of correlation

R^2 – Coefficient of Determination

RCP – Reinforced Concrete

RF - Random Forest

RI – Rating Index

RMSE - root mean squared error

ROC - Receiver Operating Characteristic Curve

RRSE – Root Relative Square Error

SG - Structural grade

SVMs - Support Vector Machines

USEPA - United States Environmental Protection Agency

TBL - Triple Bottom Line

UTA - The University of Texas at Arlington

VCP - Vitrified clay pipe

WTPs - Pretreatment and Wastewater Treatment Plants

Appendix B

Tests of Assumptions

B-1 Check Normality of variables

Normality of dependent variable pipe condition rating and independent variables was checked to show whether there is the skewedness or normality. The frequency histograms were derived and presented in Figures B-1 to B-11.

The basic statistics of the independent and dependent variables are presented in Table B-1. Based on existing measured values of different variables and their correlative analysis, total 11 factors (variables) including diameter, age, and pH influence the model development. The data sets were statistically compared for any significant difference among them using F-test (p = 0.05) for each of the measured variables.

B-2 Standardize data

Table B-35 Standardized Variables Case processing

Variables	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Zscore(Pipe Material)	1096	96.10%	44	3.90%	1140	100.00%
Zscore(Surface Condition)	1096	96.10%	44	3.90%	1140	100.00%
Zscore(Soil Type)	1096	96.10%	44	3.90%	1140	100.00%
Zscore(Condition Rating)	1096	96.10%	44	3.90%	1140	100.00%
Zscore(Corrosivity Concrete)	1096	96.10%	44	3.90%	1140	100.00%
Zscore(Corrosivity Steel)	1096	96.10%	44	3.90%	1140	100.00%
Zscore(Diameter)	1096	96.10%	44	3.90%	1140	100.00%
Zscore(Age)	1096	96.10%	44	3.90%	1140	100.00%
Zscore(Slope)	1096	96.10%	44	3.90%	1140	100.00%
Zscore(Depth)	1096	96.10%	44	3.90%	1140	100.00%
Zscore(Length)	1096	96.10%	44	3.90%	1140	100.00%
Zscore(pH)	1096	96.10%	44	3.90%	1140	100.00%

Table B-1 shows standardized variables. The data is standardized with only 3.9% of the cases missing.

Table B-36 Tests of Normality of Standardized Data

Variables	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Zscore (Pipe Material)	0.273	1096	0.000	0.875	1096	0.000
Zscore(Surface Condition)	0.372	1096	0.000	0.723	1096	0.000
Zscore(Soil Type)	0.361	1096	0.000	0.766	1096	0.000
Zscore(Condition Rating)	0.471	1096	0.000	0.533	1096	0.000
Zscore(Corrosivity Concrete)	0.54	1096	0.000	0.237	1096	0.000
Zscore(Corrosivity Steel)	0.536	1096	0.000	0.299	1096	0.000
Zscore(Diameter)	0.246	1096	0.000	0.724	1096	0.000
Zscore(Age)	0.112	1096	0.000	0.945	1096	0.000
Zscore(Slope)	0.313	1096	0.000	0.321	1096	0.000
Zscore(Depth)	0.221	1096	0.000	0.87	1096	0.000
Zscore(Length)	0.127	1096	0.000	0.836	1096	0.000
Zscore(pH)	0.351	1096	0.000	0.635	1096	0.000

a. Lilliefors Significance Correction

Table B-2 shows that all variables are significant.

B-3 Correlations

Table B-37 Standardized Variables Descriptive Statistics

Variables	Mean	Std. Deviation	N
Zscore(Pipe Material)	0	1	1112
Zscore(Surface Condition)	0	1	1139
Zscore(Soil Type)	0	1	1135
Zscore(Condition Rating)	0	1	1136
Zscore(Corrosivity Concrete)	0	1	1137
Zscore(Corrosivity Steel)	0	1	1137
Zscore(Diameter)	0	1	1140
Zscore(Age)	0	1	1140
Zscore(Slope)	0	1	1137
Zscore(Depth)	0	1	1137
Zscore(Length)	0	1	1140
Zscore(pH)	0	1	1135

Table B-3 shows all variables were standardized with a mean of 0 and Standard Deviation of 1.

B-4 Validation Data Descriptive Statistics

Table 4-38 Validation Data Descriptive Statistics

Variables	N	Minimum	Maximum	Mean	Std. Deviation
Pipe Material	259.00	1.00	5.00	2.36	0.99
Surface Condition	263.00	1.00	4.00	2.62	0.88
Soil Type	258.00	1.00	4.00	2.53	0.73
Condition Rating	262.00	0.00	1.00	0.75	0.44
Corrosivity Concrete	260.00	0.00	1.00	0.05	0.21
Corrosivity Steel	260.00	0.00	1.00	0.92	0.27
Diameter	263.00	6.00	78.00	18.92	14.53
Age	263.00	1.00	101.00	42.83	22.70
Slope	263.00	0.00	4.80	0.76	0.91
Depth	263.00	4.00	20.00	7.67	2.47
Length	263.00	17.00	2054.00	329.24	261.94
pH	258.00	5.30	8.20	7.86	0.56

Table B-4 shows that the descriptive data for validation.

Table B-39 Standardized Validation Case Processing Summary

Variables	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Zscore(Pipe Material)	253	96.2%	10	3.8%	263	1
Zscore(Surface Condition)	253	96.2%	10	3.8%	263	1
Zscore(Soil Type)	253	96.2%	10	3.8%	263	1
Zscore(Condition Rating)	253	96.2%	10	3.8%	263	1
Zscore(Corrosivity Concrete)	253	96.2%	10	3.8%	263	1
Zscore(Corrosivity Steel)	253	96.2%	10	3.8%	263	1
Zscore(Diameter)	253	96.2%	10	3.8%	263	1
Zscore(Age)	253	96.2%	10	3.8%	263	1
Zscore(Slope)	253	96.2%	10	3.8%	263	1
Zscore(Depth)	253	96.2%	10	3.8%	263	1
Zscore(Length)	253	96.2%	10	3.8%	263	1
Zscore(pH)	253	96.2%	10	3.8%	263	1

Table B-5 shows only 3.8% of the data was missing.

Table B-40 Tests of Normality of the Validation Datasets

Variables	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Zscore(Pipe Material)	0.26	253.00	0.00	0.87	253.00	0.00
Zscore(Surface Condition)	0.35	253.00	0.00	0.76	253.00	0.00
Zscore(Soil Type)	0.34	253.00	0.00	0.79	253.00	0.00
Zscore(Condition Rating)	0.47	253.00	0.00	0.54	253.00	0.00
Zscore(Corrosivity Concrete)	0.54	253.00	0.00	0.21	253.00	0.00
Zscore(Corrosivity Steel)	0.54	253.00	0.00	0.30	253.00	0.00
Zscore(Diameter)	0.27	253.00	0.00	0.70	253.00	0.00
Zscore(Age)	0.12	253.00	0.00	0.95	253.00	0.00
Zscore(Slope)	0.24	253.00	0.00	0.72	253.00	0.00
Zscore(Depth)	0.20	253.00	0.00	0.88	253.00	0.00
Zscore(Length)	0.16	253.00	0.00	0.81	253.00	0.00
Zscore(pH)	0.34	253.00	0.00	0.63	253.00	0.00

Table B-6 shows that all variables are significant.

B-5 Check Linearity

Figures B-1 to B-12 show the scatterplots comparing the observed and forecasted.

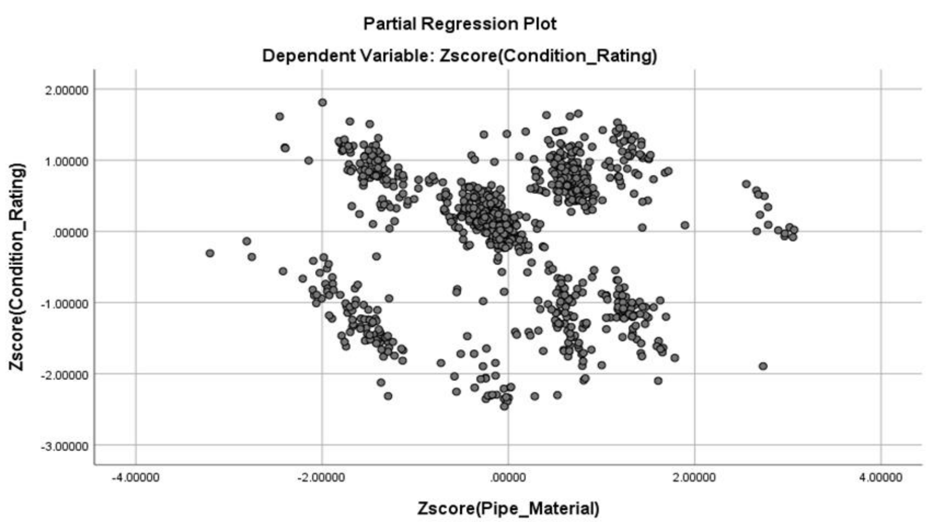

Figure B-40 Checking linearity between pipe condition rating and pipe material datasets

Figure B-1 shows scatter graph of pipe material standardized datasets. The plot shows that material variable scattered. The plot shows that material variable is independent. The negative residuals do not cluster together, and the positive residuals do not cluster together.

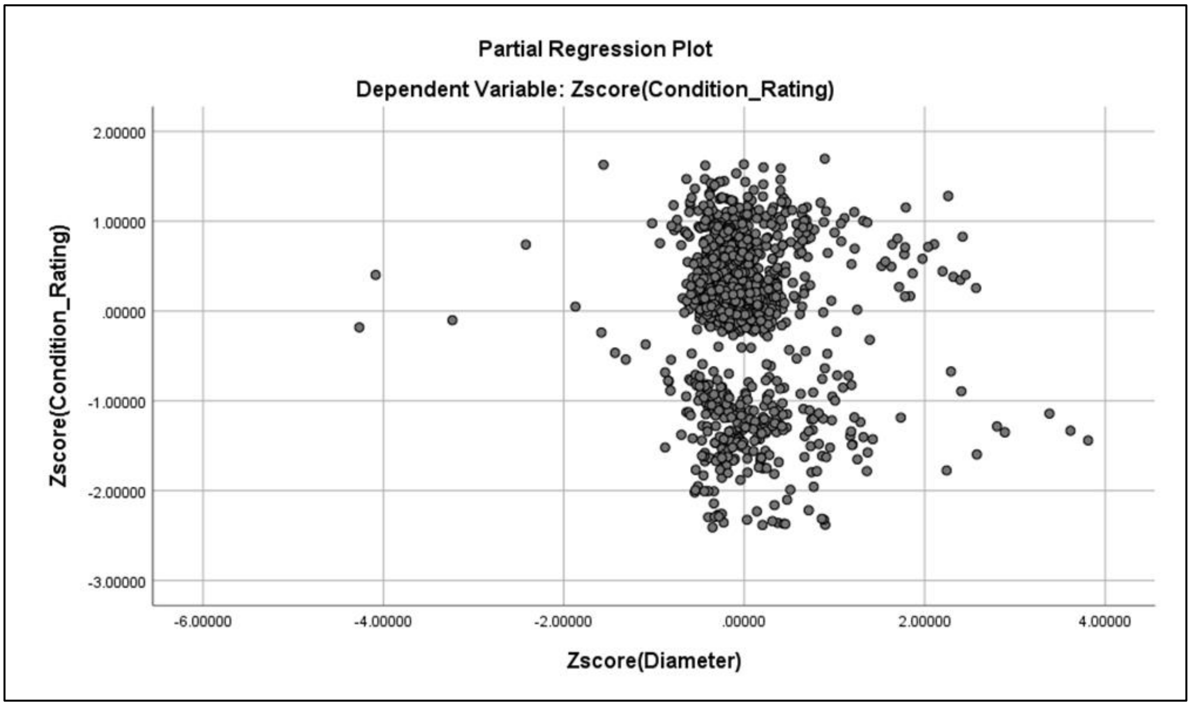

Figure B-41 Checking linearity between pipe condition rating and pipe diameter datasets.

Figure B-2 shows a scatter graph of pipe diameter variable standardized datasets. The plot shows that the diameter variable is scattered. The plot shows that the diameter variable is independent. The negative residuals do not cluster together, and the positive residuals do not cluster together.

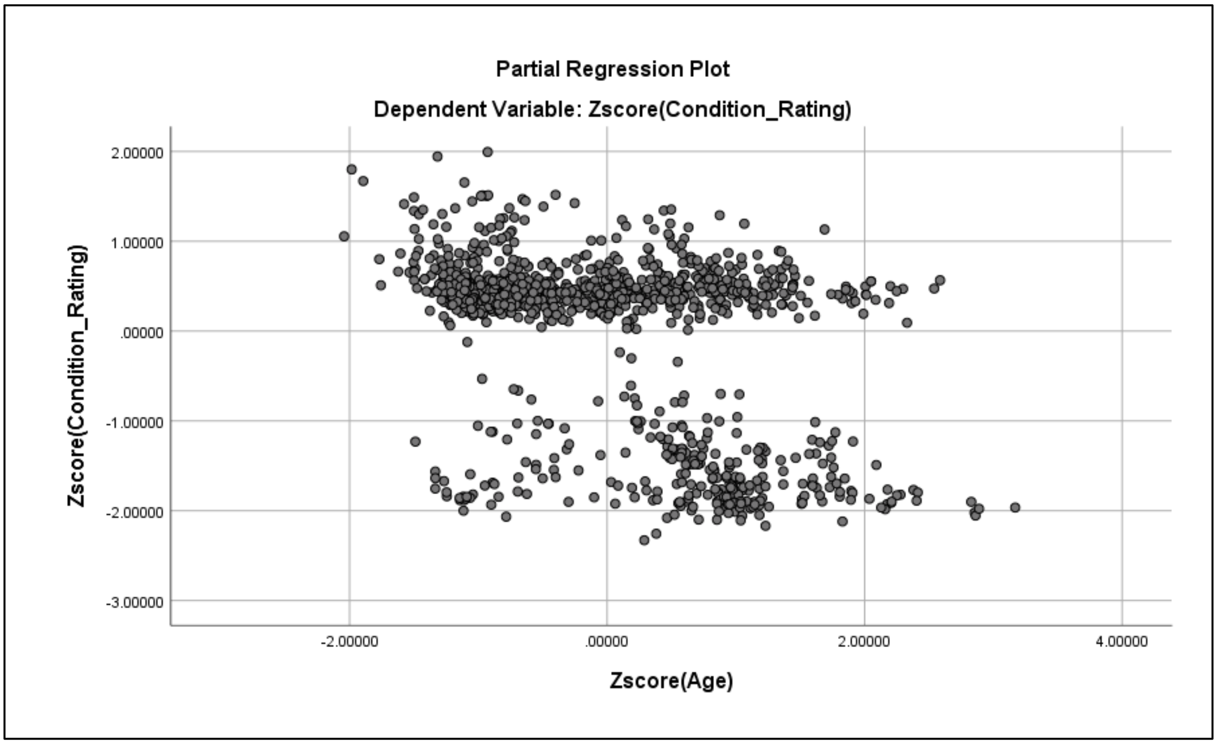

Figure B-42 Checking linearity between pipe condition rating and age datasets.

Figure B-3 shows a scatter graph of pipe age variable standardized datasets. The plot shows that diameter variable datasets are scattered. The negative residuals do not cluster together, and the positive residuals do not cluster together.

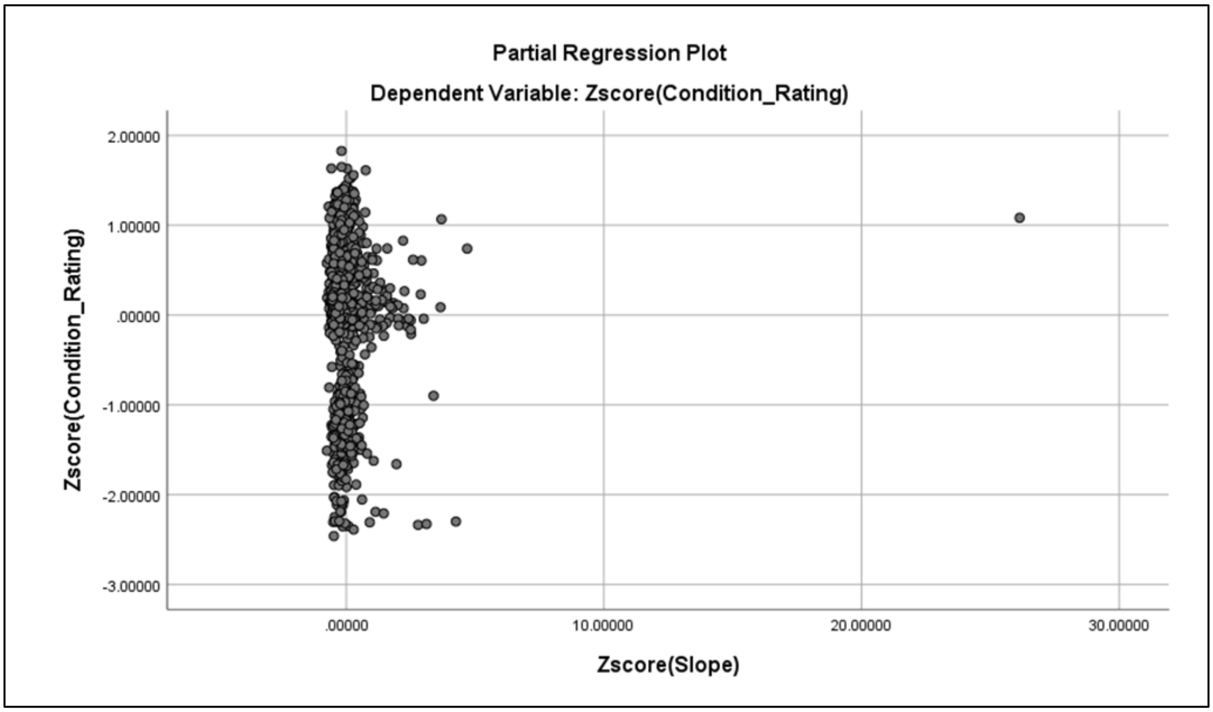

Figure B-43 Checking linearity between pipe condition rating and slope datasets.

Figure B-4 shows that the slope variable datasets are scattered. The plot shows that the slope variable is independent. The negative residuals do not cluster together, and the positive residuals do not cluster together.

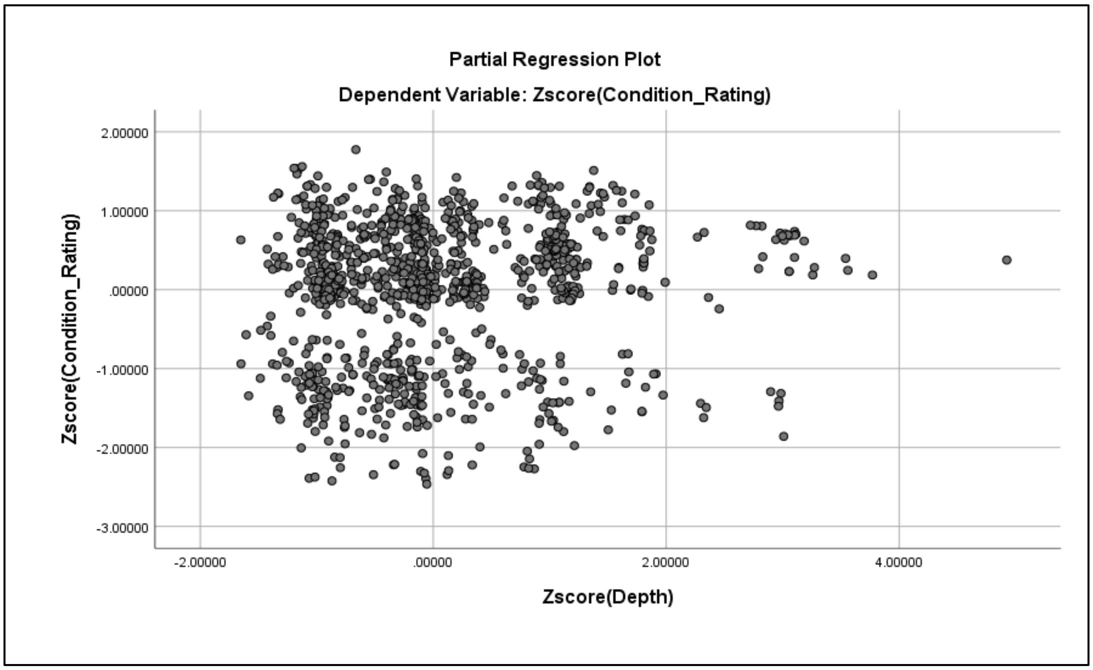

Figure B-44 Checking linearity between pipe condition rating and depth datasets.

Figure B-5 shows that the depth variable datasets are scattered. The plot shows that the depth variable is independent. The negative residuals do not cluster together, and the positive residuals do not cluster together.

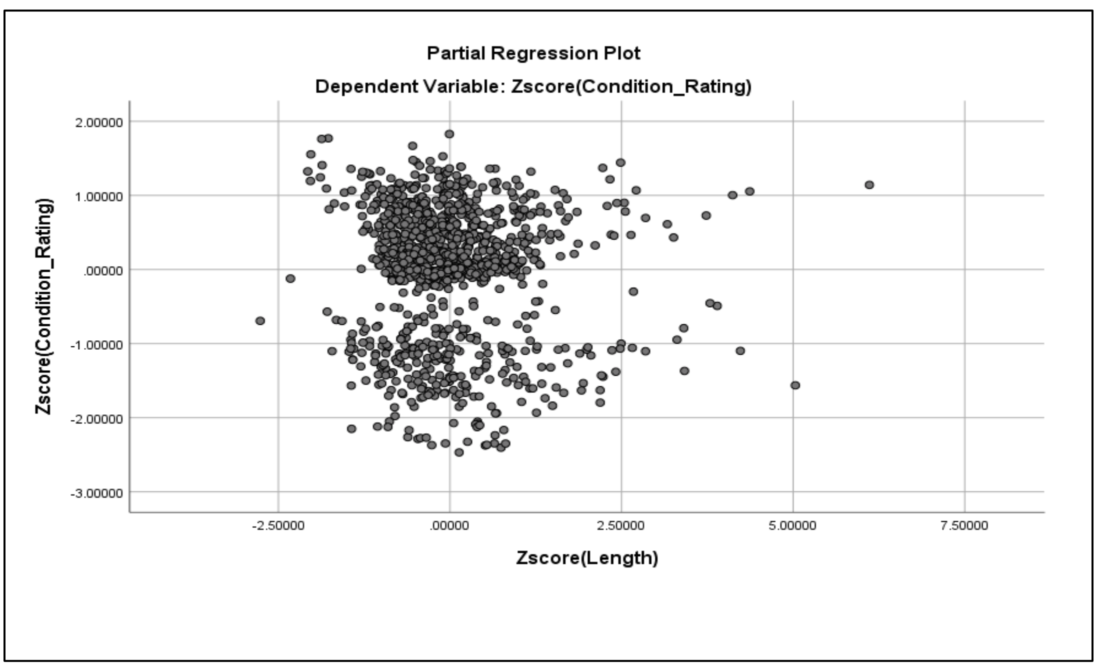

Figure B-45 Checking linearity between pipe condition rating and length datasets.

Figure B-6 shows that the length variable datasets are scattered. The plot shows that the length variable is independent. The negative residuals do not cluster together, and the positive residuals do not cluster together.

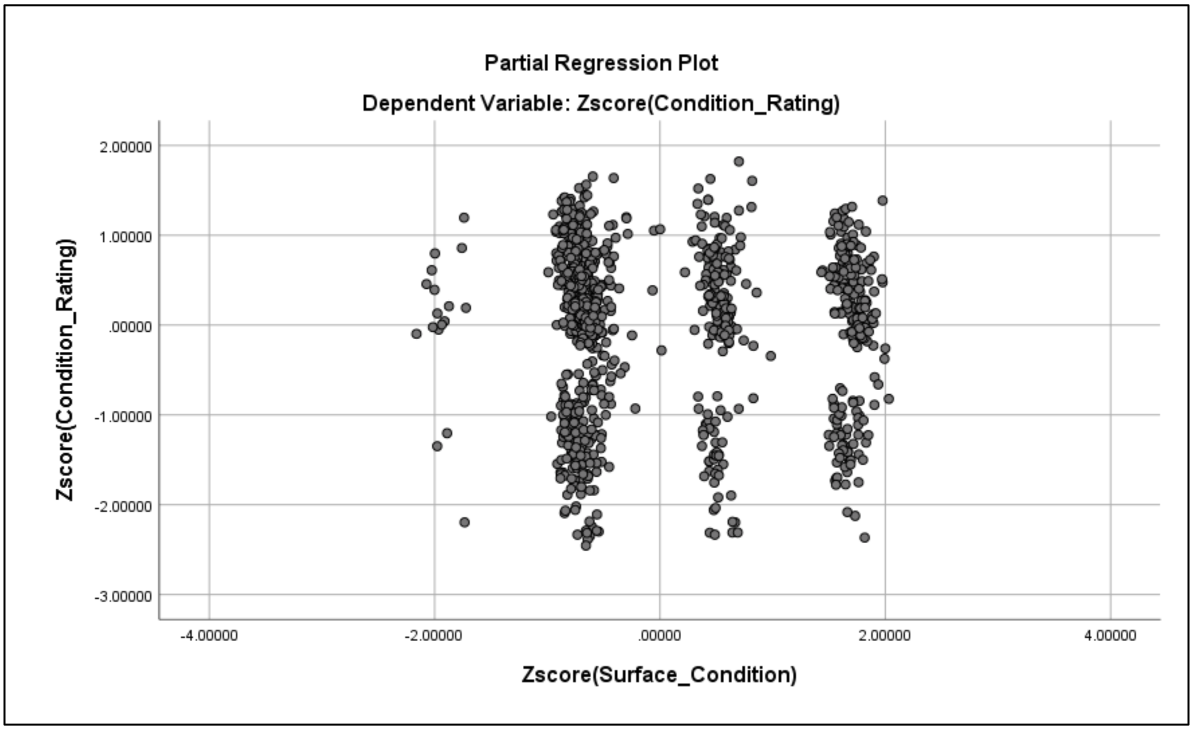

Figure B-46 Checking linearity between pipe condition rating and surface condition datasets.

Figure B-7 shows that the surface condition variable datasets are scattered. The plot shows that the surface condition variable is independent. The negative residuals do not cluster together, and the positive residuals do not cluster together.

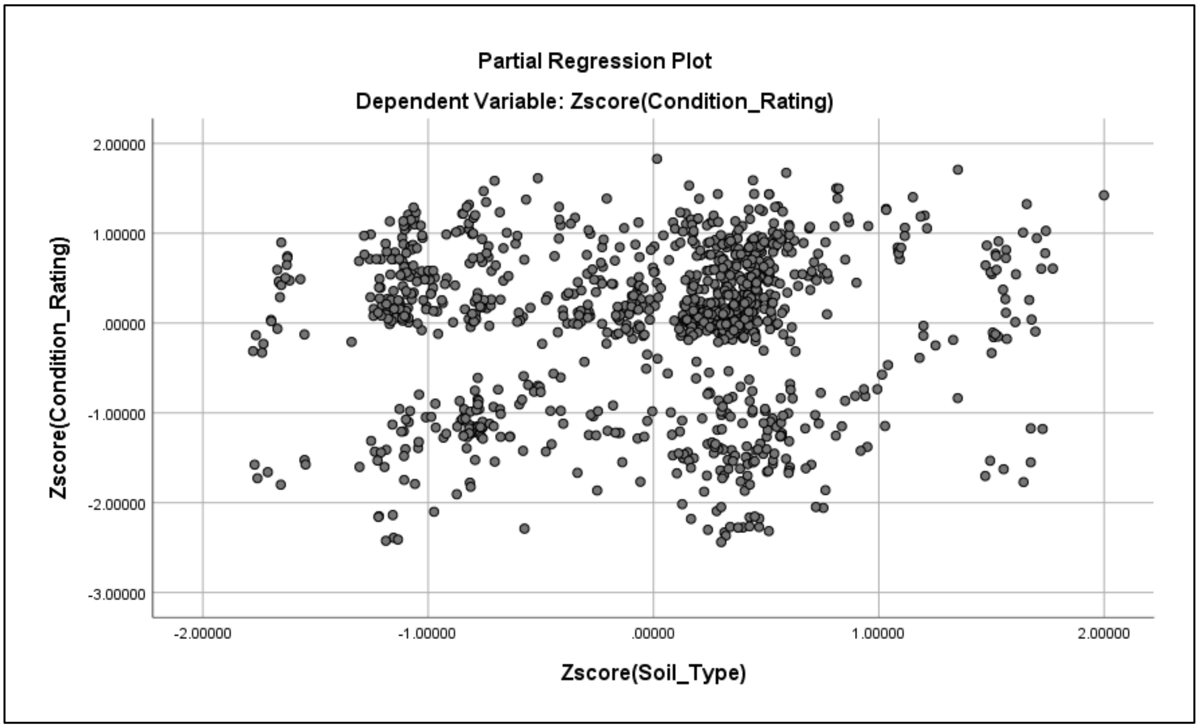

Figure B-47 Checking linearity between pipe condition rating and soil type datasets.

Figure B-8 shows that the soil type variable datasets are scattered. The plot shows that the soil type variable is independent. The negative residuals do not cluster together, and the positive residuals do not cluster together.

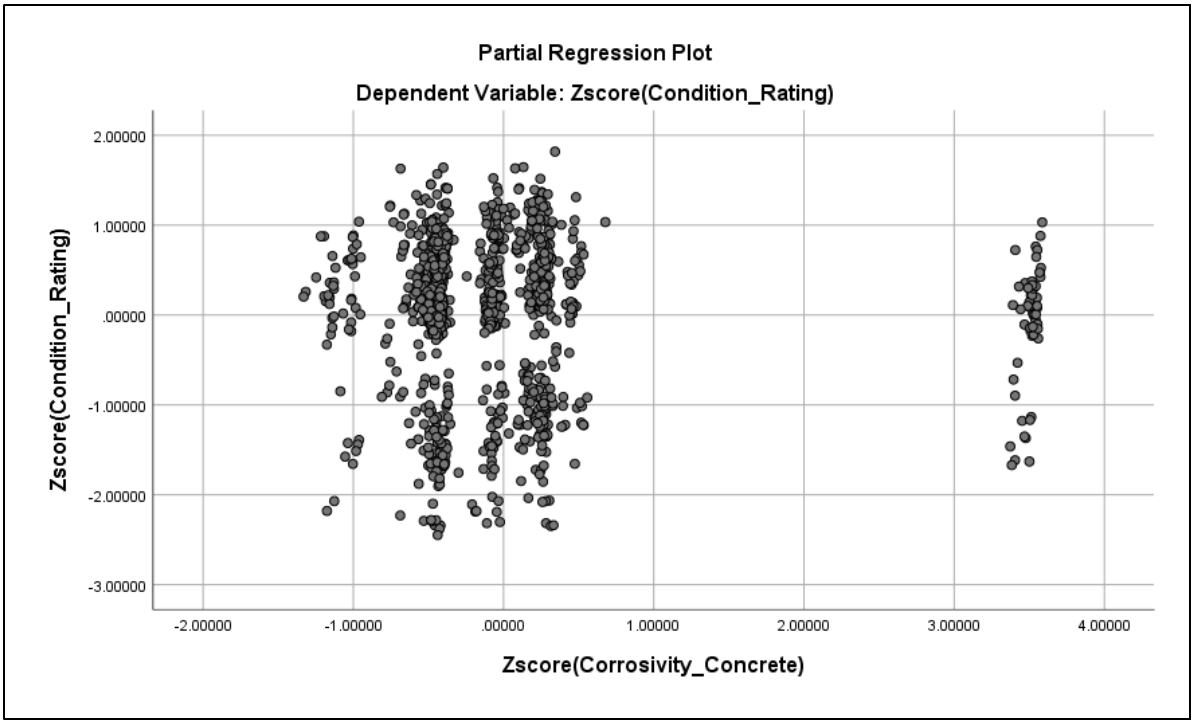

Figure B-48 Checking Linearity between Pipe Condition Rating and Corrosivity Concrete Datasets

Figure B-9 shows that the corrosivity concrete variable datasets are scattered. The plot shows that the corrosivity concrete variable is independent. The negative residuals do not cluster together, and the positive residuals do not cluster together.

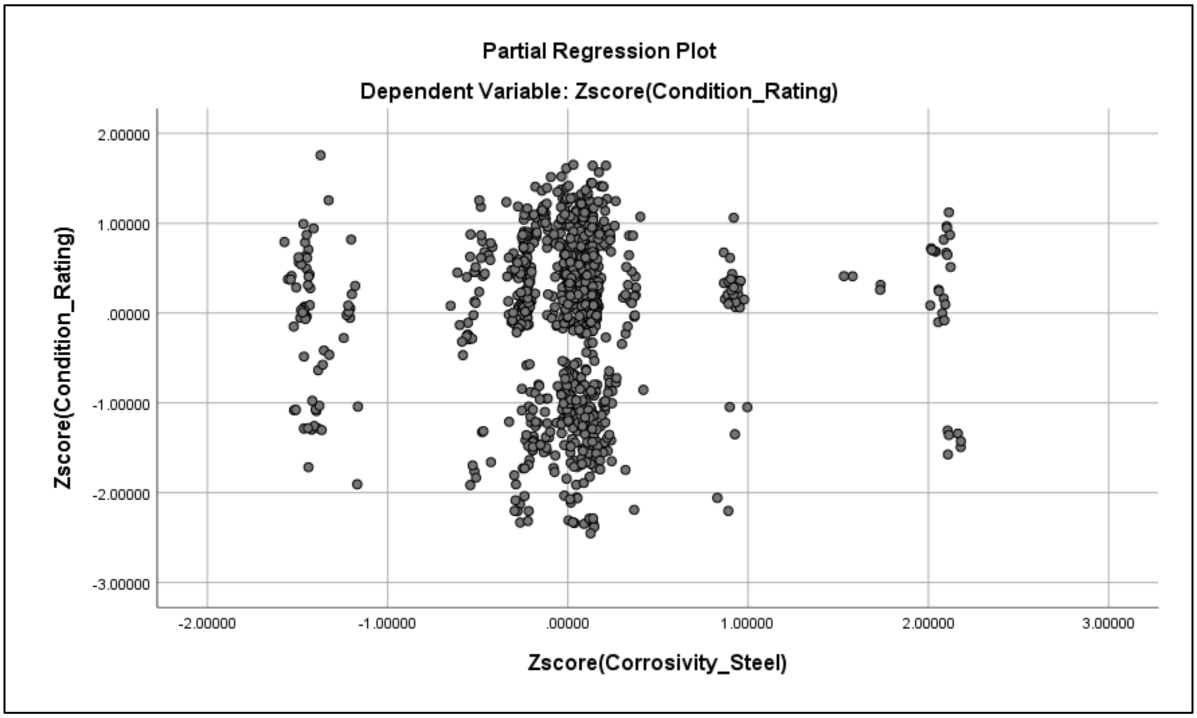

Figure B-49 Checking linearity between pipe condition rating and corrosivity steel datasets.

Figure B-10 shows that the corrosivity steel variable datasets are scattered. The negative residuals do not cluster together, and the positive residuals do not cluster together. The plot shows that soil corrosivity for the steel variable is independent. The negative residuals do not cluster together, and the positive residuals do not cluster together.

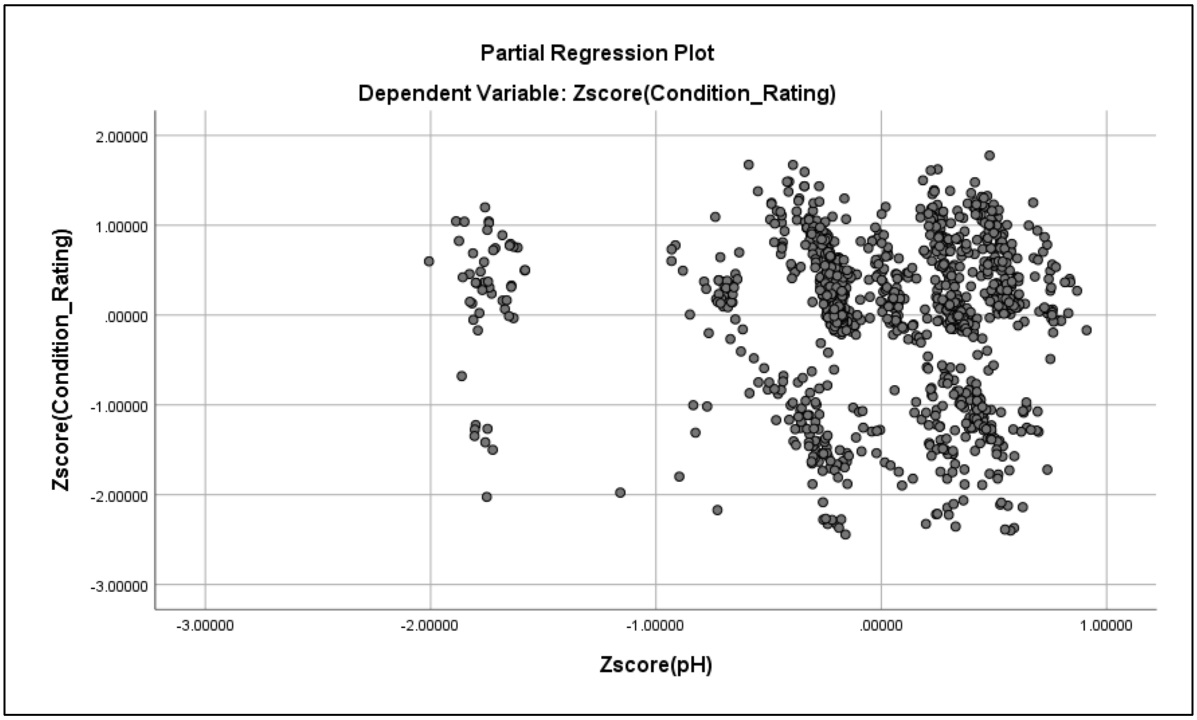

Figure B-50 Checking linearity between pipe condition rating and pH datasets.

Figure B-11 shows that the pH variable datasets are scattered. The negative residuals do not cluster together, and the positive residuals do not cluster together. The plot shows that the pH variable is independent. The negative residuals do not cluster together, and the positive residuals do not cluster together.

Appendix C

Sample Sanitary Sewer Pipes Data

DWUKEY	Material	Surface Condition	Soil Texture	CORCON	CORSTEE	PACP Score	Flow	Pipe Material	Surface Condition	Soil Type	Condition Rating	Corrosivity Concrete	Corrosivity Steel	Diameter	Age	Slope	Depth	Length	pH
1154492	PVC	Highway	Clay	Low	High	1 - Excellent	0.79	2	1	3	1	0	1	10	28	0.30	9	87.20	7.90
197151	CONC	Street	Loam	Low	High	1 - Excellent	1.41	1	2	2	1	0	1	12	71	0.30	6	235.92	8.20
191381	CONC	Street	Clay	Low	High	1 - Excellent	43.79	1	2	3	1	0	1	33	84	0.22	10	71.92	7.90
138828	CONC	Alley	Clay	High	High	3 - Fair	3.94	1	3	3	0	1	1	15	59	0.87	11	305.81	7.50
1370645	RC	Highway	Loam	Low	High	5 - Extremely Poor	47.05	4	1	2	0	0	1	54	43	0.07	8	572.12	8.20
905305	VCT	Easement	Rock	Low	High	2 - Good	5.44	3	4	4	1	0	1	18	61	0.60	15	205.62	8.20
1091486	PVC	Highway	Clay	Low	High	1 - Excellent	2.07	2	1	3	1	0	1	10	20	2.16	8	131.25	8.20
1005827	VCT	Street	Clay	Low	High	2 - Good	1.36	3	2	3	1	0	1	12	100	0.30	8	578.98	7.90
131811	PVC	Street	Clay	Low	High	1 - Excellent	1.25	2	2	3	1	0	1	15	39	0.10	10	886.35	7.90
1482573	PVC	Street	Clay	Low	High	1 - Excellent	2.35	2	2	3	1	0	1	18	12	0.12	7	599.09	7.90
1291736	PVC	Street	Clay	Low	High	1 - Excellent	1.67	2	2	3	1	0	1	8	15	4.50	8	43.45	8.20
195447	RC	Easement	Loam	Low	High	3 - Fair	10.87	4	4	2	0	0	1	33	91	0.72	5	39.56	8.20
1409376	RC	Street	Clay	High	High	3 - Fair	71.37	4	2	3	0	1	1	30	83	1.20	10	68.81	7.50
203021	CONC	Street	Loam	Low	High	1 - Excellent	17.49	1	2	2	1	0	1	30	62	0.07	11	954.07	8.20
1246929	PVC	Street	Clay	Low	High	1 - Excellent	0.48	2	2	3	1	0	1	8	18	0.40	5	103.23	7.90
924610	VCT	Easement	Clay	Low	High	2 - Good	1.47	3	4	3	1	0	1	12	53	0.30	10	187.39	7.90
176461	PVC	Street	Clay	Low	High	1 - Excellent	1.64	2	2	3	1	0	1	12	22	0.56	7	485.49	7.90
195527	CONC	Street	Loam	Low	High	5 - Extremely Poor	17.14	1	2	2	0	0	1	24	68	0.30	5	157.83	8.20
209751	VCT	Street	Loam	Moderate	High	1 - Excellent	1.94	3	2	2	1	0	1	15	50	0.20	11	1037.16	7.50
928037	VCT	Street	Sand	Low	Moderate	1 - Excellent	0.79	3	2	1	1	0	0	8	50	0.01	5	576.66	6.70
921859	VCT	Street	Sand	Low	Moderate	2 - Good	0.64	3	2	1	1	0	0	8	49	0.94	8	173.54	6.70
872259	PVC	Street	Rock	Low	High	1 - Excellent	4.65	2	2	4	1	0	1	12	15	4.00	5	183.34	8.20
866360	CONC	Alley	Clay	Low	High	1 - Excellent	0.89	1	3	3	1	0	1	8	85	0.01	5	221.62	8.20
187853	RC	Easement	Clay	High	High	3 - Fair	31.08	4	4	3	0	1	1	60	40	0.03	10	997.39	7.50
1457769	PVC	Street	Clay	High	High	1 - Excellent	0.81	2	2	3	1	1	1	8	14	0.56	8	209.06	7.50
1284262	PVC	Street	Clay	Low	High	1 - Excellent	1.23	2	2	3	1	0	1	8	16	2.45	5	296.75	8.20
203248	RC	Easement	Loam	Low	High	3 - Fair	21.27	4	4	2	0	0	1	27	67	0.58	7	503.25	8.20
1130716	PVC	Street	Sand	Moderate	Moderate	1 - Excellent	0.70	2	2	1	1	0	0	12	16	0.20	8	83.51	6.50
153486	PVC	Alley	Clay	Low	High	1 - Excellent	1.00	2	3	3	1	0	1	10	16	0.80	8	180.96	7.90
1125956	PVC	Street	Clay	Low	High	1 - Excellent	1.04	2	2	3	1	0	1	12	24	0.20	8	583.69	7.90
1369859	PVC	Alley	Sand	Moderate	Moderate	5 - Extremely Poor	0.96	2	3	1	0	0	0	10	40	0.01	6	476.89	5.80
1631253	VCT	Street	Loam	Low	High	3 - Fair	7.11	3	2	2	0	0	1	18	68	1.10	8	165.70	8.20
1314780	PVC	Street	Clay	Low	High	1 - Excellent	0.65	2	2	3	1	0	1	12	15	0.09	8	37.59	7.90
1100657	PVC	Street	Clay	Low	High	1 - Excellent	1.13	2	2	3	1	0	1	12	17	0.32	7	270.18	8.20
1101157	RC	Street	Clay	Low	High	3 - Fair	40.19	4	2	3	0	0	1	66	18	0.03	8	495.76	7.90
1466860	PVC	Street	Sand	Moderate	Moderate	1 - Excellent	184.12	2	2	1	1	0	0	36	15	0.38	6	58.62	5.30
139506	CONC	Street	Clay	Low	High	1 - Excellent	1.04	1	2	3	1	0	1	10	74	0.40	8	25.73	7.90
117810	PVC	Street	Clay	Low	High	1 - Excellent	5.62	2	2	3	1	0	1	18	41	0.63	7	125.33	7.90
190358	PVC	Street	Sand	Moderate	High	1 - Excellent	1.61	2	2	1	1	0	1	12	25	0.20	8	583.25	6.80

Figure C-51 Sample of Sanitary Sewer Pipes Data

References

Adamowski, J., Fung Chan, H., Prasher, S. O., Ozga-Zielinski, B., and Sliusarieva, A. (2012). “Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada.” Water Resources Research, 48(1).

Alegre, H. (2010). “Is strategic asset management applicable to small and medium utilities?” Water Science and Technology.

American Society of Civil Engineers (ASCE). (2017). 2021 Report Card for America’s Infrastructure. Washington, DC.

Berardi, L., Giustolisi, O., Savic, D. A., and Kapelan, Z. (2009). “An effective multi-objective approach to prioritisation of sewer pipe inspection.” Water Science and Technology.

Bruaset, S., Rygg, H., and Sægrov, S. (2018). “Reviewing the long-term sustainability of urban water system rehabilitation strategies with an alternative approach.” Sustainability (Switzerland), MDPI AG, 10(6).

Burn, S., Marlow, D., and Tran, D. (2010). “Modelling asset lifetimes and their role in asset management.” Journal of Water Supply: Research and Technology - AQUA, 59(6–7), 362–377.

Caradot, N., Riechel, M., Fesneau, M., Hernandez, N., Torres, A., Sonnenberg, H., Eckert, E., Lengemann, N., Waschnewski, J., and Rouault, P. (2018). “Practical benchmarking of statistical and machine learning models for predicting the condition of sewer pipes in Berlin, Germany.” Journal of Hydroinformatics, IWA Publishing, 20(5), 1131–1147.

Chakraverty, S., and Mall, S. (2017). Artificial Neural Networks for Engineers and Scientists.

Chughtai, F., M., and Zayed, T. (2008). “Infrastructure Condition Prediction Models for Sustainable Sewer Pipelines.” Journal of Performance of Constructed Facilities, 5(22), 333–341.

Elmasry, M., Hawari, A., and Zayed, T. (2017). “Defect based deterioration model for sewer pipelines using Bayesian belief networks.” Canadian Journal of Civil Engineering, Canadian Science Publishing, 44(9), 675–690.

Hawari, A., Alkadour, F., Elmasry, M., and Zayed, T. (2020). “A state-of-the-art review on condition assessment models developed for sewer pipelines.” Engineering Applications of Artificial Intelligence, 93.

Hinton, P.R., McMurray, I., and Brownlow, C. (2014). SPSS Explained, Second Edition. Routledge Taylor & Francis Group, New York, New York.

Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). Applied Logistic Regression. John Wiley & Sons. Inc., Columbus, Ohio.

Hou, P., Yi, X., and Dong, H. (2020). "A spatial statistic-based risk assessment approach to prioritize the pipeline inspection of the pipeline network." Energies, MDPI AG, 13(3).

Kaushal, V. (2019). Comparison of environmental and social costs of trenchless cured-in-place pipe renewal method with open-cut pipeline replacement for sanitary sewers. Dissertation, The University of Texas at Arlington.

Khashei, M., and Bijari, M. (2010). "An artificial neural network (p, d, q) model for timeseries forecasting." Expert Systems with Applications, 37(1).

Khazraeializadeh, S. (2012). A Comparative Analysis on Sewer Structural Condition Grading Systems Using Four Sewer Condition Assessment Protocols.

Kulandaivel, G. (2004). "Sewer Pipeline Condition Prediction Using Neural Network Models."

Laakso, T., Kokkonen, T., Mellin, I., and Vahala, R. (2018). "Sewer condition prediction and analysis of explanatory factors." Water (Switzerland), MDPI AG, 10(9).

Lomax, R. G., and Hahs-Vaughn, D. L. (2012). Statistical Concepts.

Lubini, A. T., and Fuamba, M. (2011). "Modeling of the deterioration timeline of sewer systems." Canadian Journal of Civil Engineering, 38(12), 1381–1390.

Malek Mohammadi, M., Najafi, M., Kermanshachi, S., Kaushal, V., and Serajiantehrani, R. (2020). "Factors Influencing the Condition of Sewer Pipes: State-of-the-Art Review." Journal of Pipeline Systems Engineering and Practice, 11(4).

Mohammadi, M. M., Najafi, M., Kaushal, V., Serajiantehrani, R., Salehabadi, N., and Ashoori, T. (2019). "Sewer pipes condition prediction models: A state-of-the-art review." Infrastructures, MDPI Multidisciplinary Digital Publishing Institute.

Muhlbauer, K. (2004). Pipeline Risk Management Manual Ideas, Techniques, and Resources Third Edition. Oxford.

Najafi, M. (2016). Pipeline Infrastructure Renewal and Asset Management. Mc-Graw-Hill Education, New York.

Najafi, M., and Gokhale, S. (2005). Trenchless Technology Pipeline and Utility Design, Construction, and Renewal. McGraw-Hill, New York.

NASSCO. (2018). Pipeline Assessment & Certification Program (PACP). NASSCO, Marriottsvile.

Opila, M. C. (2011). Structural Condition Scoring of Buried Sewer Pipes for Risk-Based Decision Making.

Osama Moselhi, B., and Shehab-Eldeen, T. (2000). Classification of Defects in Sewer Pipes Using Neural Networks.

Peponi, A., Morgado, P., and Trindade, J. (2019). "Combining Artificial Neural Networks and GIS Fundamentals for Coastal Erosion Prediction Modeling." Sustainability (Switzerland), MDPI AG, 11(4).

Rokstad, M. M., and Ugarelli, R. M. (2015). "Evaluating the role of deterioration models for condition assessment of sewers." Journal of Hydroinformatics, IWA Publishing, 17(5), 789–804.

Salman, B., and Salem, O. M. (2004). Infrastructure Management and Deterioration Risk Assessment of Wastewater Collection Systems.

Samuel Ariaratnam, B. T., El-Assaly, A., Members, A., and Yang, Y. (2001). Assessment of Infrastructure Inspection Needs Using Logistic Models.

Selvakumar, A., Matthews, J. C., Condit, W., and Sterling, R. (2015). "Innovative research program on the renewal of aging water infrastructure systems." Journal of Water Supply: Research and Technology - AQUA, IWA Publishing.

Shehab-Eldeen, T., and Moselhi, O. (2001). "A decision support system for rehabilitation of sewer pipes." Canadian Journal of Civil Engineering, 28(3), 394–401.

Sousa, V., Matos, J. P., Almeida, N., and Matos, J. S. (2014). "Risk assessment of sewer condition using artificial intelligence tools: Application to the SANEST sewer system." Water Science and Technology, IWA Publishing, 69(3), 622–627.

Syachrani, S. (2010). Advanced Sewer Asset Management Using Dynamic Deterioration Models.

Syachrani, S., Jeong, H. D., and Chung, C. S. (2013). "Advanced criticality assessment method for sewer pipeline assets." Water Science and Technology, 67(6), 1302–1309.

Tade, O., O'Neill, S., Smith, K. G., Williams, T., Ali, A., Bayyati, A., and See, H. (2019). "Modified sewer asset management to accommodate London's future sustainable development." International Journal of Building Pathology and Adaptation, Emerald Group Publishing Ltd., 37(1), 22–41.

Tscheikner-Gratl, F., Egger, P., Rauch, W., and Kleidorfer, M. (2017). "Comparison of Multi-Criteria Decision Support Methods for Integrated Rehabilitation Prioritization." Water, 9(68).

Tscheikner-Gratl, F., Mikovits, C., Rauch, W., and Kleidorfer, M. (2014). "Adaptation of sewer networks using integrated rehabilitation management." Water Science and Technology, IWA Publishing, 70(11), 1847–1856.

Ugarelli, R., Venkatesh, G., Brattebø, H., and Sægrov, S. (2008). "Importance of investment decisions and rehabilitation approaches in an ageing wastewater pipeline network. A case study of Oslo (Norway)." Water Science and Technology, 58(12), 2279–2293.

USEPA. (2009). Condition Assessment of Wastewater Collection Systems - White Paper.

USEPA. (2010). Report on Condition Assessment Technology of Wastewater Collection Systems. Edison.

Vahidi, E., Jin, E., Das, M., Singh, M., and Zhao, F. (2016). "Environmental life cycle analysis of pipe materials for sewer systems." Sustainable Cities and Society, (27), 167–174.

Vladeanu, G. J., and Matthews, J. C. (2019). "Consequence-of-failure model for risk-based asset management of wastewater pipes using AHP." Journal of Pipeline Systems Engineering and Practice, American Society of Civil Engineers (ASCE), 10(2).

Ward, B., and Savić, D. A. (2012). "A multi-objective optimization model for sewer rehabilitation considering critical risk of failure." Water Science and Technology, 66(11), 2410–2417.

Wirahadikusumah, R., Abraham, D. M., and Castello, J. (1999). "Markov decision process for sewer rehabilitation." Engineering, Construction and Architectural Management.

Wooldridge, J. M. (2013). Introductory Econometrics, A Modern Approach. Cengage Learning, Delhi.

Yang, M. D., and Su, T. C. (2006). "Automation model of sewerage rehabilitation planning." Water Science and Technology, 225–232.

Yin, X., Chen, Y., Bouferguene, A., Al-Hussein, M., Russell, R., and Kurach, L. (2020). “A neural network-based approach to predict the condition for sewer pipes.” Construction Research Congress 2020: Infrastructure Systems and Sustainability - Selected Papers from the Construction Research Congress 2020.

Young, D. S. (2017). Handbook of Regression Methods. CRC Press, Taylor & Francis Group, Boca Raton, FL.

Biographical Information

Earning Ph.D. in Civil Engineering degree has been Daniel Atambo's career dream. Daniel Atambo has been an exemplary Doctoral Student in Civil Engineering. The quality of Daniel's research projects and dissertation stand out and set him apart as a promising researcher and scholar. His distinctive research in his dissertation entitled, "D*evelopment and Comparison of Prediction Model for Sanitary Sewer Pipes Condition Assessment using Multinomial Logistic Regression and Artificial Neural Network,*" addresses unique challenges with sanitary sewer pipes. Employing his genius in civil engineering, Daniel developed a model that predicts sanitary sewer pipes condition. Remarkably, not only can the model be utilized by municipal water utilities to determine the condition of sewer pipes it can also be used to identify pipes that should be rehabilitated or replaced.

Daniel holds two Master's degrees in Environmental Engineering from Lamar University, U.S.A., and Agricultural Engineering from Egerton University, Kenya and undergraduate degree in Agricultural Engineering from the University of Nairobi, Kenya. Daniel has designed, planned, and managed numerous projects in water and wastewater pipelines and irrigation and drainage systems. Daniel is a registered professional engineer (PE) in the State of Texas and a member of the American Society of Civil Engineers (ASCE). Daniel works for the City of Dallas Water Utilities' Pipeline Project Management. He has been volunteering for 15 years on International Community Projects with Engineers Without Borders (EWB). Daniel plans to teach, consult, and conduct research in construction and management of water, and wastewater projects.