

A machine learning approach for predicting and localizing the failure and damage point in sewer networks due to pipe properties

Mohammad Reza Goodarzi ^{a,*} and Majid Vazirian^b

^a Department of Civil Engineering, Yazd University, Yazd, Iran

^b Department of Civil Engineering, Water Resources Management Engineering, Yazd University, Yazd, Iran

*Corresponding author. E-mail: Goodarzimr@yazd.ac.ir

 MRG, 0000-0002-6296-2315

ABSTRACT

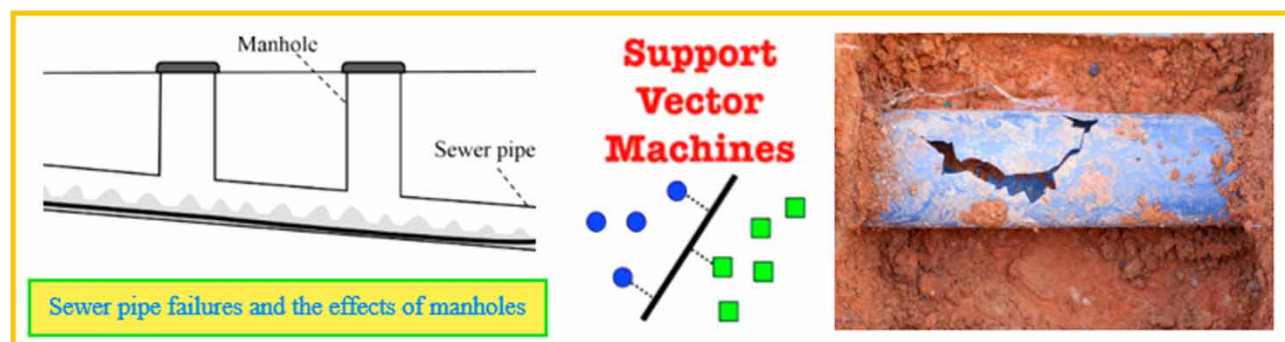
As a basic infrastructure, sewers play an important role in the innards of every city and town to remove unsanitary water from all kinds of livable and functional spaces. Sewer pipe failures (SPFs) are unwanted and unsafe in many ways, as the disturbance that they cause is undeniable. Sewer pipes meet manholes frequently, unlike water distribution systems, as in sewers, water movement is due to gravity and manholes are needed in every intersection as well as through pipe length. Many studies have been focused on sewer pipe failures and so on, but few investigations have been done to show the effect of manhole proximity on pipe failure. Predicting and localizing the sewer pipe failures is affected by different parameters of sewer pipe properties, such as material, age, slope, and depth of the sewer pipes. This study investigates the applicability of a support vector machine (SVM), a supervised machine learning (ML) algorithm, for the development of a prediction model to predict sewer pipe failures and the effects of manhole proximity. The results show that SVM with an accuracy of 84% can properly approximate the manhole effects on sewer pipe failures.

Key words: machine learning approach, manhole proximity effect, sewer network, sewer pipe damage, sewer pipe failure decision, SVM (support vector machine)

HIGHLIGHTS

- A machine learning approach was carried out to predict the sewer pipe damage in sewage networks.
- This study investigates the applicability of a support vector machine (SVM) for the development of a prediction model to predict the failure of sewer pipes.
- Diagnosis of the effects of manhole proximity on sewer pipe failure is important to identify and predict the effect of manholes on pipe failure.

GRAPHICAL ABSTRACT



This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

1. INTRODUCTION

1.1. Water and health

Clean and hygienic water and proper management of clean drinking water and sanitary services are vital for humans and ecosystems. The lack of clean water has endangered human health and the sustainability of the environment and ecosystems. Studies show that about four billion people live in extreme water shortage conditions for at least 1 month of the year (Shehata *et al.* 2023). That is why collecting wastewater from urban areas and accessing adequate safe drinking water and sanitation are fundamental and imperative to human development, health, and well-being, and are internationally accepted human rights (Zheng *et al.* 2022). Studies have been devoted to the issues of water resources such as limited water resources, climate change, drought, groundwater discharge, water and wastewater networks, and declining quality of water all over the world (Alotaibi *et al.* 2023; Barbieri *et al.* 2023; Engelbert & Scheuring 2023; He & Rosa 2023; Huang *et al.* 2023; Liu *et al.* 2023).

Water supply issues generally are not only related to the scarcity of water resources, but also the lack of appropriate technologies for water supply, improper treatment and distribution networks, insufficient use of national or international financial resources, and lack of implementation of necessary strategies in accordance with national, regional, and local conditions (Goodarzi *et al.* 2023). For example, inadequate technology may reduce the efficiency of water extraction and distribution, which may lead to virtual scarcity even if water resources are sufficient. Improper distribution and treatment networks are susceptible to significant water loss and pollution. Insufficient use of financial resources may lead to insufficient investment in water infrastructure and insufficient water supply. Lack of implementation of necessary strategies may lead to inefficient water management and exacerbate water shortage problems.

1.2. Wastewater infrastructures

As a basic infrastructure, sewer pipes are vitally important and play an unobtrusive but significant role in modern cities. The sewer pipes are expected to work unstopably to improve the public health and environmental protocols. Hydraulic implications of inefficient sewer networks lead to flow dynamics complexities and potential cascading effects on public health.

Inefficient sewer networks are those that have inadequate capacity, poor layout, or structural damage that affects their hydraulic performance and reliability, which can cause flow obstacles, such as blockages, surcharges, or overflows, consequently reducing the conveyance of wastewater and increasing the risk of flooding, pollution, and health hazards (Damvergis 2014; Sourabh & Timbadiya 2018; Duque *et al.* 2020). Flow dynamics complexities in sewer networks are influenced by factors such as rainfall intensity and duration, infiltration and inflow, pipe geometry and roughness, and network topology. The potential cascading effects of inefficient sewage networks on public health include exposure to pathogens, chemicals, and sewage odors, contamination of water resources and ecosystems, and damage to infrastructure and property (Nasrin *et al.* 2017).

A sewer pipe failure (SPF) has the potential to lead to human life problems and a failure of public life (McLaren *et al.* 2022), as it can cause human health risks, environmental damage, infrastructural challenges, and economic impacts because failures of water and wastewater networks can lead to severe economic consequences. Therefore, it is crucial to invest in the treatment of sewage because proper maintenance and timely repair of sewer systems can help mitigate these issues (Laakso *et al.* 2018a; Li & Achal 2020). Some factors may cause a wastewater pipe to reduce its efficiency years after being in operation, such as pipe age, traffic loads, poor construction, excessive capacity use, environmental conditions, and poor operating management (Xu *et al.* 2022).

Therefore, sewage system failure (SSF) leads to many economic, social, environmental, and health consequences that may affect human life and environmental protocols. Renewing the sewer network system (SNS) is often neglected as most of the pipes are buried and far from the supervisor and pedestrian point of view (Ariaratnam *et al.* 2001).

A sewer network is affected by several damages during its lifetime, such as tree root penetration, floods, earthquakes, and blockages. Over the years, the condition of the sewer network pipes deteriorates because of situations such as operating weaknesses, poor maintenance, and irregular supervision of the sewer pipes (Moeini & Zare 2021). Implementing a strong maintenance strategy can help strengthen the sewer network. This may include regular inspections to quickly identify potential problems, prompt repairs to address any damage, and preventative measures to protect the network from common issues such as tree root intrusion and blockages (Stip *et al.* 2019).

There are two general approaches to programming the renovation of sewer pipes, such as the project ‘Sealing of Sewers – Effects on the Cleaning Performance of Sewage Treatment Plants and Influence on the Local Water Balance,’ undertaken by the Institute for Underground Infrastructure (IKT)

The approaches gather enough data from sewer pipes over the years, then apply statistical methods to predict pipe failures and damage-related points of the network to manage the planning of repairs (Hoseingholi & Moeini 2023).

This is essential, as lack of knowledge about the conditions of the pipes leads to failure and sudden and unpredictable accidents, which in addition to stopping the service, disrupts the activities on the ground and traffic activities (Fontecha *et al.* 2021). Therefore, to prevent the destruction and collapse of the network, its condition should be checked at appropriate times, before an accident occurs, through conditions such as periodic visits and video-metric visual operations (CCTV). In most countries, for example in Germany, sewage networks must be inspected once every 10 years. This means that, on average, a sewer pipe is inspected 8–10 times during its lifetime, depending on the development of the sewer system and sewage network, the population growth rate, and effects of inflation on increasing pipe age cost and priority of pipe repairs emphasizing the accurate forecasting and prediction of wastewater network conditions (Anbari & Tabesh 2015).

In many cases, infrastructure maintenance is postponed because of the limited budget and high cost (Hoseingholi & Moeini 2023). Obviously, neglecting it will reduce the performance of the sewer system and in some cases the sewer service could stop completely.

Neglecting the maintenance of sewage pipes leads to defects in the sewage network. These defects may lead to excess water entering the pipes from unwanted sources, be it surface water, groundwater through the defect, or a direct illegal connection. In addition, the roots of nearby plants may penetrate through cracks in sewer lines and lead to blockages. More severe signs of SPF include foundation problems, such as cracks, failures, and in some cases sinkholes. Therefore, regular maintenance of sewer pipes is very important to avoid these issues and ensure the efficient functioning of the sewer system (Elwira & Zielińska 2017). Conversely, following a regular maintenance program for the network leads to an increase in the useful life of the network, which in most cases should be prioritized despite the limited budget of the sewer network inspection activities (Sattar *et al.* 2017). The factors affecting sewer pipe efficiency are as follows:

1.2.1. Pipe age

Over time, pipes are subjected to corrosion which can lead to leaks, cracks, or blockages. Gedam *et al.* proposed a model to investigate effective parameters and demonstrated that pipe age is very significant to the model but the depth is not of significant impact (Gedam *et al.* 2016).

1.2.2. Traffic loads

If the sewer network is located under a road with heavy traffic, the weight and vibration can cause the pipes to crack or collapse. Sewer pipes are prone to deterioration and aging, as well as damage under excessive stresses from traffic (Zamanian & Shafieezadeh 2023).

1.2.3. Poor construction

Infrastructures, especially SNSs, are usually not only out of sight, but also out of mind. Reports show that the current performance of many sewer systems is poor and many systems have received minimal maintenance over several years (Damvergis 2014). If a pipe is not properly installed, due to improper slope, poor joint construction, or the use of incorrect materials, it will not be able to handle the wastewater flow properly.

1.2.4. Excessive capacity use

If the amount of wastewater exceeds the capacity of the sewer pipe, it can lead to backups and overflows, which is often a problem in areas with rapid population growth.

1.2.5. Environmental conditions

Factors such as soil type, groundwater levels, and temperature can affect the integrity of a pipe. For example, in cold regions, water in pipes can freeze and lead to the bursting of pipes.

1.2.6. Poor operating management

Lack of regular inspection and maintenance may lead to problems going unnoticed until they become serious. Regular cleaning of pipes prevents the buildup of grease, hair, and other materials that can cause blockages.

1.3. Investigating sewer network properties

As mentioned in the previous section, some conditions applied to sewer pipes can affect their life and performance. Conversely, some characteristics of sewer pipes are also effective in their health and durability, which are discussed in this section. The health and durability of a sewer network may be affected by various parameters such as the type of sewer pipes and the material quality, pipe diameter, depth of pipes in the soil, slope of the pipe, and so on. To prevent failure, the sewage conditions must be checked at appropriate times, which requires the correct prediction of the network conditions. Furthermore, due to the fact that the wear and tear in the sewer network has a great impact on its performance, monitoring the condition of sewer pipes is one of the low-cost methods that have attracted the attention of many researchers recently.

Many studies have been undertaken in the field for predicting the state of the sewage network. Hahn *et al.* described the development of an expert system that prioritizes sewer network inspection, identified potential hazards and also their consequences, provided the appropriate inspection test method, and warned the user regarding accurate determination. Prioritizing sewer network inspections will help identify the critical points of the system to reduce the number of inspections and the costs of emergency preventions and repairs. The logic of this expert system is based on the information provided by the water and wastewater companies (Hahn *et al.* 1999).

A new classification system was investigated to evaluate the leakage in a sewage network. For classification, data such as gender, length, diameter, underground water level, soil type, and age of pipes were used (Baik *et al.* 2006).

A logistic regression model was predicted to investigate the probability of failure in the sewer network of Edmonton. In this model, parameters such as diameter, age, type of pipe, type of sewage, and installation depth were considered. Analysis of the results revealed that the parameters of age, diameter, and type of sewage systems have a significant impact on the likelihood of network failure, and the effect of the two parameters of the depth of burial and the type of the network is insignificant. The output of this model is used to plan and schedule inspections and periodic repairs. This means that network failure is not affected by how deep the pipes are buried or whether they are made of concrete, polyethylene, or metal. The results are very useful for planning and scheduling periodic inspections and repairs of the sewer network. They can be used to estimate the probability of failure for each pipe section in the network and to prioritize those with the highest risk of failure. Also, the results can be used to identify the factors that contribute to network failure and suggest solutions to improve network performance and reliability (Guo *et al.* 2022).

The conditions of the sewers and the prediction of the remaining life of the pipes were investigated. To evaluate the condition of the network, the six parameters of location, soil type, burial depth, pipe dimensions, type of sewage, and seismic condition of the area were considered. In this research, the decision to reconstruct the pipes was based on the classification of network conditions (Anbari & Tabesh 2015).

According to a study by Rezaei *et al.* (2015), water pipe failures are associated with pipe characteristics, material properties, and environmental and loading conditions. Their investigations found that the steady and unsteady state of hydraulic conditions may impose excessive loading on the assets. The resulting cyclic loadings when acting upon degraded (e.g. corroded) pipes could contribute to the development and acceleration of failures. The study also investigated the impact of dynamic hydraulic conditions on pipe failures by analyzing historical burst records, designing and implementing an extensive experimental program, and investigating case studies by correlating dynamic pressure and pipe failures (Rezaei *et al.* 2015).

1.4. Sewer network failure prediction

There have been specific challenges in sewer network management, such as structural vulnerabilities, which can be a consequence of population growth, urbanization, and climate change (Jia *et al.* 2021); however, the literature shows a lack of predictive models. Some studies have been done by experimental laboratory investigations (Nalluri *et al.* 1994; Vongvises-somjai *et al.* 2010), using artificial neural network (ANN) in predicting the sediment transport (Ebtehaj & Bonakdari 2013; Kakoudakis *et al.* 2018), and some were Markov chain studies which are introduced in the following, but few studies contain machine learning (ML) and deep learning.

Sewerage network planning for its inspection was done with the help of predicting the state of the network. In this research, parameters such as the age and type of pipe, diameter, slope, and type of sewage were used to predict the time of critical conditions of the sewer network. The proposed function was developed based on the wear rate of the pipes and it was used to plan the inspection time of the network (McDonald & Zhao 2020).

As mentioned previously, in some of the studies the conditions of the sewer network infrastructure were predicted by using the Markov chain method. A model was proposed to predict the condition of sewer pipes and related repair costs, and three models were proposed for the maintenance and repair program. In the first model, the current condition of the pipe was predicted and scored according to the age and type of the pipe and its length (Malek Mohammadi *et al.* 2021). In the second model, the future condition of the pipe was predicted, and in the third model, the repair and reconstruction costs were investigated.

Najafi and Kulandaivel predicted the conditions of sewage collection pipelines using the ANN model to predict the condition of sewer pipes based on historical condition assessment data. The parameters examined in this research included length, diameter, gender, age, depth of coating, slope of pipes, and type of network. This model can optimize the number of inspections and costs of reconstruction and repairs. The developed model is intended to help identify damaged sections of the sewer pipeline network using a set of input values, then can be directed toward assessing and prioritizing maintenance actions needed to prevent the acceleration of future problems and the eventual failure of sewer pipes (Najafi & Kulandaivel 2005).

Baik *et al.* (2006) evaluated the conditions of the sewage system using the Markov chain model. The variables considered in this model included the length, diameter, type, age, and slope of the pipe. It is worth mentioning that the advantage of the proposed method is ease of use and accuracy in estimating the possibility of producing matrices based on the experiences of workers. A Markov chain is a discrete-time stochastic process in which the conditional probability of any future event depends only on the current state and is independent of past states. A Markov chain model needs data from assessments of existing systems conditions. The first step in evaluating the status of the sewer network is to determine the current structural and hydraulic status of the network. Evaluation of the structural condition of pipes determines the severity of the defects. The adequacy of the capacity of the existing sewer network is evaluated through the assessment of hydraulic conditions. Structural conditions are monitored via internal inspections, while hydraulic conditions are analyzed by hydraulic modeling. Infiltration and inflow are checked to identify the causes of structural failure and hydraulic overloads. Using the structural condition evaluation results, the transition probabilities for the deterioration model based on the Markov chain can be estimated. The researchers applied and evaluated the model using the condition data of sewer pipes and demonstrated that the developed model offers many advantages in estimating transmission probabilities in comparison with previously developed approaches, including the approach based on nonlinear optimization, in terms of versatility in implementation, data accuracy, and appropriateness of assumptions in the model (Baik *et al.* 2006).

By presenting a model to investigate the condition of the infrastructure, Chughtai and Zayed predicted the conditions of the sewage pipes. Using the multiple regression method and historical data, they investigated and evaluated the functional and structural conditions of a sewage network. Parameters, including the type of pipe, bed material, type of street, and other characteristics of the pipe, were considered. Finally, based on the type of pipes, different regression models were proposed and forecasted.

They declared that the developed regression models have shown a good accuracy of 82–86% and are used to generate deterioration curves for concrete, asbestos cement, and polyvinyl chloride pipes in relation to traffic loads, bedding materials, and other pipe characteristics. They considered the developed models are suitable to assist municipal engineers in identifying critical sewer conditions, prioritizing sewer inspections and rehabilitation requirements (Chughtai & Zayed 2008).

Mashford *et al.* (2011) employed a support vector machine (SVM) to predict sewer network conditions, categorizing the network on a scale from 1 to 5. Their model incorporated parameters such as age, diameter, pipe type, soil characteristics, and groundwater level. This approach demonstrated that the SVM achieves a good predictive performance. Accessing a representative set of training data, the SVM approach can be used to allocate a condition grade to sewer networks with a good confidence and identify high-risk sewer pipes for subsequent inspection (Mashford *et al.* 2011).

Anbari & Tabesh (2015) calculated the probability of failure in sewage collection networks using the Bayesian network, which was applied because of its capabilities and high efficiency, and the characteristics of the sewage system. Using the probability of the failure event obtained from the model, the sewer pipes were divided into five groups according to the priority of inspection and maintenance programs. The output of this model was the probability of a failure event in each pipeline (Anbari & Tabesh 2015). Gedam *et al.* (2016) performed a linear regression method to predict the conditions of a sewage network. For this purpose, information such as the type and age of the pipe used and the important and influential factors in predicting the failure of the network were investigated. The results showed that the proposed model can be used to predict the state of other system networks (Gedam *et al.* 2016).

Laakso *et al.* (2018b) examined the status of the sewage network by combining inspection results, network characteristics, and environmental factors. Utilizing the Boruta algorithm, they modeled pipe conditions and assessed the significance of various variables. Their research delved into screening sewer pipes for future inspections (Laakso *et al.* 2018b). Kabir *et al.* (2018) applied a regression model to investigate the structural condition of a sewage network and the effect of certain parameters on the destruction of the network. The results showed that the age and length of the sewer have the greatest effect on the destruction of concrete and cement pipes, as well as the age and diameter of the metal and polyethylene sewers (Kabir *et al.* 2018). Model scalability is perhaps the most complex challenge in pipe failure assessment. The desired output should determine the scalability of the model, although short-term predictions at pipe levels perform poorly. At low spatial and temporal scales, unbalanced data must be carefully considered (Barton *et al.* 2022). Pipe failure mechanisms are unique to different pipe materials (Barton *et al.* 2020). It is necessary to investigate pipe failure modes and mechanisms from historical data for predictive modeling as it avoids errors from purely data-driven approaches that are unreasonable, which leads to unrealistic assumptions as well as extra work by measuring the wrong thing (Barton *et al.* 2019).

1.5. Geostatistical prediction of pipe failures

Geostatistical approaches developed by the means of statistical continuous random functions generated to estimate a continuous surface parameter used to represent a particular attribute are also used to determine the most susceptible areas suffering from pipe failure (Goodarzi & Vazirian 2023).

Potyralla (2019) investigated the modeling of water supply network loads and its subsequent forecasting, which is a necessary element for making an optimum decision in the process of planning the development and operation of water networks, and demonstrated that the aim of their research was to present the advantages of geographic information systems (GIS) for studying water supply systems (Potyralla 2019). Vinson (2017) estimated the potential of a geostatistical methodology for locating rain-derived infiltration and also inflow to wastewater treatment systems in the Paul Metropolitan Area, Minnesota, USA, and revealed that this methodology can locate areas of high risk (Vinson 2017).

The literature shows that there have been a variety of investigations in many fields of infrastructures and water distribution systems, but few studies in the field of sewer networks failures and their affecting parameters or factors. The literature obviously demonstrates that there is a significant gap in ML as a tool in sewer pipe failures. Therefore, this study focuses on the effects of the existence of a manhole as an intersection on sewer pipe failures, which has not been investigated yet.

The literature on the research also shows that determining the probability of failure and predicting the structural and hydraulic failure of sewage networks is one of the topics of interest to researchers in the field of water and sewage, for which various methods have been proposed. Manholes are the connection points of the sewer network pipes and play a very important role in sewer networks; hence it is necessary to locate them on intersections or after some distance in sewer pipes to facilitate the wastewater flow or network periodic visits. So far, the effect of the presence of manholes on the extent of sewage failures and clogging has not been taken into account because of the lack of data in this field. Predicting the location of a blockage and its distance to the manhole can be very effective, and while saving time to remove the blockage, it prevents unnecessary excavations and high costs. Investigations show that some blockages are near the entrance or occur near the outlet of the sewer pipe where the hydraulic of the flow changes. Therefore, estimating the location of failure can help in dealing with accidents and to predict and reduce them. In the realm of sewer network research, the literature has seen numerous investigations across various infrastructural domains, particularly in water distribution systems. However, there has been limited exploration into the impact of manholes on sewer pipe failures yet, possibly due to a lack of available data in this domain. This research seeks to fill this void by investigating the influence of manholes on sewer pipe failures, with a focus on predicting the number of blockages in a specific segment of the sewer network of Isfahan City, Iran, by using ML as a tool. Leveraging the SVM model and considering parameters such as age, type, depth, and diameter of pipes, our study aims to shed light on the previously unexplored relationship between manholes and sewer pipe failures.

2. METHODS

2.1. Case study

The city of Isfahan is located in the central part of Iran and has a longitude of 51° 39' 40" E and a latitude of 32° 38' 30" N. It is about 1,570 m above sea level. With a population of about 2 million people, the city is located in the center of the Isfahan Province, with an area of about 512 km² (Statistical Center of Iran 2016). Isfahan is a metropolitan and an ancient city with very ancient monuments that attract many tourists all through the year. Figure 1 shows the study area and the location of

Isfahan in the country, province, city region, and the map of selected locations. Because of the importance of maintaining the ancient monuments of Isfahan, the population, and the traffic, it is necessary for sewer pipe infrastructures to remain safe and work uninterrupted.

In this research, a part of the sewage network of the city of Isfahan, Iran, was used as a case study to evaluate the proposed models. When a pipe failure or damage occurs, pedestrians or people who live in the neighborhood of a damage point call the number 122, which is related to sewage network incidents. We obtained the data from 122, the water and wastewater bureau of Isfahan.

In each sewer network, sewage comprises much information on sewage collection area characteristics and the health of the people (Guo *et al.* 2022). In 1335, a plan was presented for the sewage collection system of the city of Isfahan. This plan included the implementation of about 55 km of networks of pipes with a diameter of 250 mm. At present, Isfahan's 50-year-old sewage network with a length of about 3,500 km is responsible for collecting and directing the production of sewage from about 380,000 branches. The 50-year life of the sewage network of Isfahan has caused wear and tear and its condition has become critical (Saleh & Khakivatan 2009). In this research, a part of the sewage network of Isfahan was considered as a case study to determine the number of clogged pipes. Figure 2 shows the schematic map of the sewer network in the considered districts of Isfahan. In this figure, the location and scope of the wastewater treatment plant area are shown. It is worth mentioning that the area of this region is 227 km² and the length of the network in operation is about 700 km.

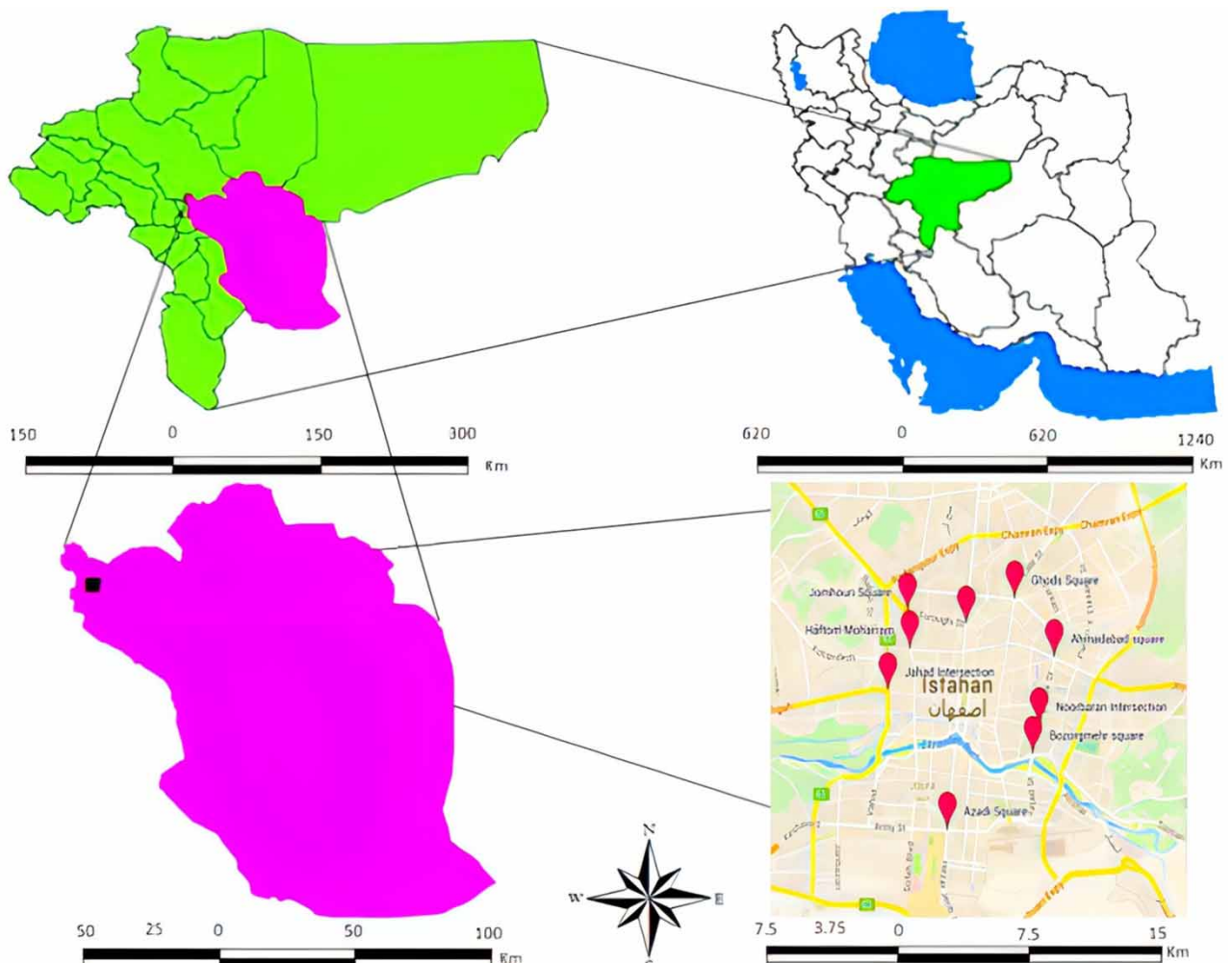


Figure 1 | Location of Isfahan in the country, province, city region, and the map of selected areas.

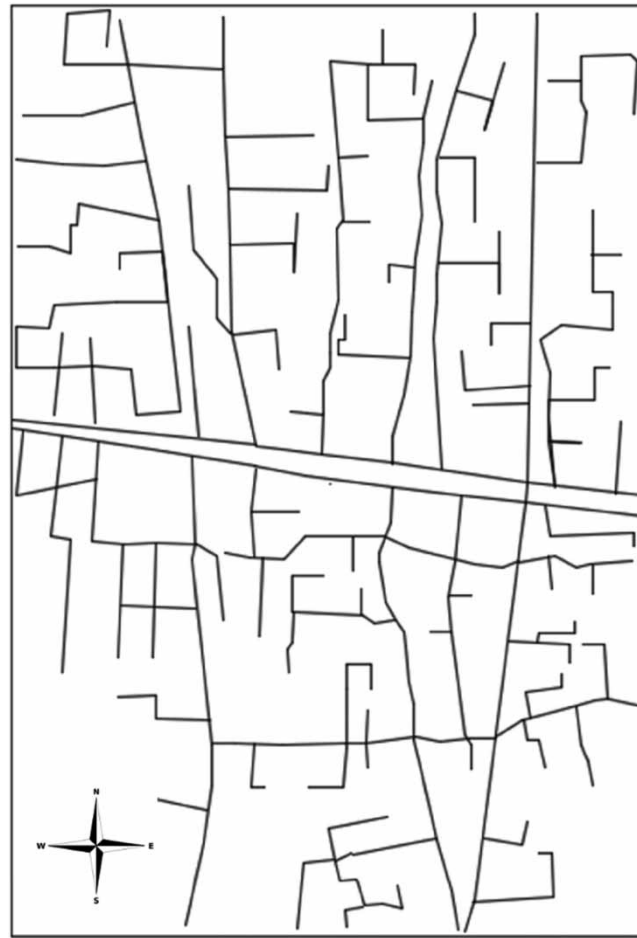


Figure 2 | Schematic map of the wastewater network layout of Isfahan.

2.2. Failure of sewer pipes

In general, various factors and parameters cause the failure of the sewage network. They can be divided into three types: environmental, functional, and physical factors. Environmental factors include parameters such as soil type, sewage type, bed conditions, freezing, proximity to other underground facilities, traffic volume, and underground water level. For example, a high temperature of the water may facilitate the erosion rate of pipes. Also, functional factors include maintenance strategies such as proper management, renewing old pipes, and developing pipe failure prediction strategies. Physical factors include age, diameter, length, type, depth, and slope, which may affect the safety of sewer pipes. For example, the old pipes are more susceptible to suffer for failure than the new pipes (Chughtai & Zayed 2008).

In general, the types of failure in the sewer grid can be divided into two categories: structural failure and hydraulic failure. Structural failure is related to the physical conditions of the pipe, while hydraulic failure is related to the inability of the pipe to meet the design capacity and is caused by an error in the design of the pipe slope. Another factor affecting hydraulic failures can be external infiltration currents, which lead to a decrease in the hydraulic capacity of the network. Sometimes, the reduction of the hydraulic capacity can be caused by tree roots or blockages in the pipeline (Abraham *et al.* 1998).

Structural failure is also applied to any type of failure that depends on the physics and structure of the network. Types of structural failures include erosion, corrosion, cross-sectional shape change, and cracking.

Each of the hydraulic and structural failures is dependent on parameters such as age, length, diameter, type, and depth. (Anbari & Tabesh 2015).

At the end, it is necessary to mention that blockage is considered one of the types of hydraulic failure. Clogging in the sewers can be caused by sediments and grease or the penetration of tree roots. Among the factors affecting the probability

of clogging in the sewer network, we can mention the age of the pipe, type, diameter, depth of burial, and number of connections (Anbari & Tabesh 2015).

2.3. Support vector machine

ML approaches are widely studied in the field of water and environment (Niazkar *et al.* 2023). SVM is one of the supervised learning methods that are used for linear and nonlinear classification as well as multidimensional regression. This method is one of the relatively new methods that has shown good performance in recent years compared with older classification methods such as neural networks.

The same as the neural networks or somewhat similarly to them, SVMs possess the well-known ability of being approximators of any multivariate function to any desired degree of accuracy. Consequently, they are of particular interest for modeling the unknown, or partially well-known, highly nonlinear, complex systems, plants, or processes (Kecman 2005).

The working basis of the SVM classifier is the linear classification of the data, which chooses the line that has a higher confidence margin. The problem of finding the optimal line for data classification is done by quadratic programming (QP) methods, which are known for solving constrained problems. In order for the machine to be able to categorize data with high complexity, the data are moved to a space with much higher dimensions by means of kernel transformation. To be able to solve the problem of very high dimensions using these methods, we use Lagrange's dual theorem to transform the minimization problem into its dual form. Different kernel functions can be used, including exponential kernels, polynomials, and sigmoids. The SVM algorithm is classified as a pattern recognition algorithm. The SVM algorithm can be used wherever there is a need to recognize a pattern or classify objects into special classes. The methods of least-squares support vector machines and twin support vector machines are improved versions of this method in terms of speed and performance, which we will describe in detail in the following.

This algorithm was first introduced in 1963 by Vladimir N. Vapnik and Alexey Ya Chervonenkis. and in 1992, Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik proposed this algorithm to build a nonlinear classifier using hyperplanes by using a kernel. The standard version that is used today was proposed in 1993 by Corinna Cortes and Vladimir N. Vapnik and published in 1995. Also in 1999, the method of least-squares support vector machines was proposed by Suykens *et al.*, who provided very interesting results in terms of time and performance. In 2007, the method of twin support vector machines was introduced by Jayadeva *et al.* and its performance was compared with common classification methods and the advantages of this method were shown. Least-squares and twin support vector machines (LST-SVM) were presented, which provided better results than the mentioned methods. Various versions of the twin SVM method were introduced from 2007 to 2016 based on the same method with a slight change, and sometimes for certain data, better results were introduced (Kutyłowska & Kowalski 2021). The SVM model has been widely used for predicting rainfall (Pham *et al.* 2018) and drought (Mafarja *et al.* 2019). In the SVM model, the input data should be divided into testing and training samples. The selected input vector (training sample) which is mapped into a high-dimensional feature space, generates the optimal decision function.

As shown in the figure of the SVM classifier, the purple dashed line classifies blue and green data points.

2.4. What are support vectors?

- Support vectors are the data points that lie closest to the decision surface (or hyperplane);
- They are the data points most difficult for classification;
- They have direct bearing on the optimum location of the decision surface;
- We can show that the optimal hyperplane stems from the function class with the lowest capacity of independent features/parameters we can adjust.

Figure 3 shows that in general, lots of possible solutions for a,b,c (an infinite number) can be imagined, but an SVM finds an optimal solution.

But SVMs maximize the margin around the separating hyperplane, as shown in Figure 4. So the decision function is fully specified by a (usually very small) subset of training samples, (the support vectors). This becomes a quadratic programming problem, which is easy to solve by standard methods.

As shown in Figure 5, for linear separability, nodes can be separated by a line, but in higher dimensions, hyperplanes are used.

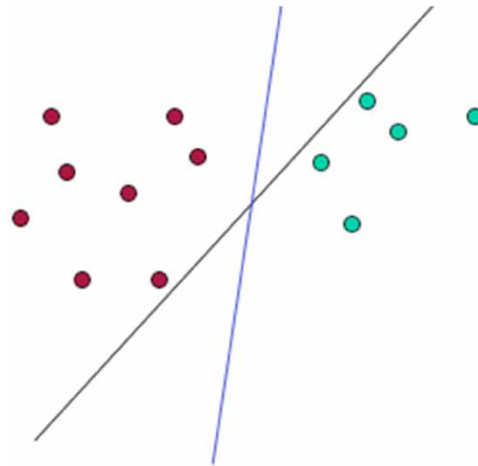


Figure 3 | Support vectors which are classifying the nodes.

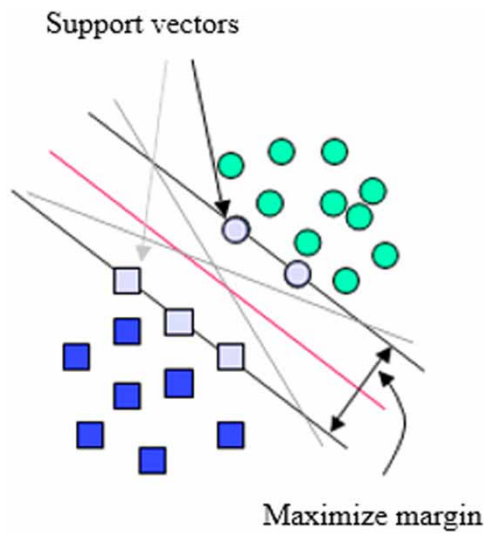


Figure 4 | Support vectors.

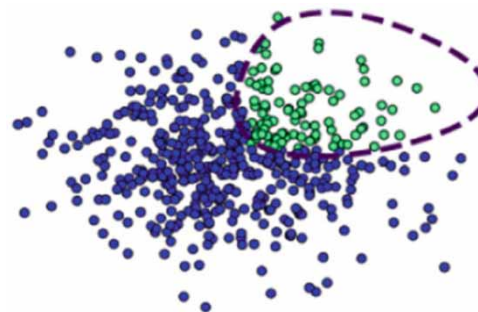


Figure 5 | Hyperplane shown as the purple dashed line classifying blue and green data points (Cai *et al.* 2022).

2.5. SVM classifier

What we mean by the classifier is actually finding a line in the form of $w \cdot x - b = 0$ to be able to correctly separate the data of two or more classes from each other. Since the word line is more reminiscent of a plane or, at best, a three-dimensional curve, for data with higher dimensions, a more general concept should be used. We will use the mathematical word hyperplane instead. In fact, in mathematical terms, the hyperplane is a subspace of the studied space, which differs from the studied space by only one dimension, for example, let's consider a 2D plane as a space, the corresponding hyperplane is a normal line, and for the 3D space, the hyperplane is a 2-dimensional plane, and in the same way, this concept is generalized for higher dimensions.

The test data set D contains n members (points), which are defined as follows:

$$D = \{(x_i, y_i) | x_i \in R^p, y_i \in \{-1, +1\}\}_{i=1}^n \quad (1)$$

where x_i are observed features, y_i are predicted data, and y is -1 or 1 and the goal is to find the separating hyperplane with the largest distance from the marginal points that separates the points $y = -1$ and $y = 1$. Any hyperplane can be written as a set of points x under the following conditions:

$$w \cdot x - b = 0 \quad (2)$$

where w is the normal vector, which is perpendicular to the hyperplane. We want to choose b and w so that the maximum distance between the parallel hyperplanes separates the data. These hyperplanes are described using the following relation.

Any data above the separating hyperplane is marked with label 1 :

$$w \cdot x - b = +1 \quad (3)$$

And any data at the bottom of the separating hyperplane, with the label -1 :

$$w \cdot x - b = -1 \quad (4)$$

If the training data are linearly separable, we can consider two hyperplanes on the edge of points such that they have no points in common, and then try to maximize their distance.

SVMs work by mapping the data into a higher-dimensional space using the kernel trick and then solving the ML task as a convex optimization problem. This allows for a clear margin of separation between layers, which can be a straight line (in two dimensions) or a hyperplane (in higher dimensions).

2.6. Confusion matrix

In supervised classification of a data set using classification methods like SVM, the goal is to achieve the highest possible precision and accuracy in the classification and recognition of categories. In some cases, it is more important to correctly identify the samples related to one of the categories. For example, consider a study in which the goal is to identify people with a certain type of a dangerous disease. Suppose there is a risk of death for people who are suffering from this disease and they need to receive some kind of special medicine to eliminate this risk. In this situation, the correct diagnosis of patients is very important.

In other words, the error in the diagnosis of healthy people can be ignored, but this possibility cannot be accepted for the identification of sick people. In other words, our expectation is to detect all sick individuals, without bias, even if a healthy individual is mistakenly classified as a sick individual. In such cases, when the accuracy of the diagnosis of a category is more important compared with the accuracy of the overall diagnosis, the concept of 'Confusion Matrix' becomes important.

Based on the example stated previously, let's consider belonging to the category of sick people as positive and not belonging to this category as negative. Each sample or individual in reality belongs to one of the positive or negative classes; and no matter the algorithm that is used to categorize the data, in the end, each member sample will be classified into one of these two 'classes'. Therefore, for each data sample, one of the following four states may occur (Table 1).

TP = The sample is a member of the positive category and is recognized as a member of the same class (true positive).

FN = The class member sample is positive and the class member is recognized as negative (false negative).

Table 1 | Confusion matrix

Predicted values	Actual values	
		TP
	FN	TN

TN = The sample is a member of the negative class and is recognized as a member of the same class (true negative).

FP = And finally, the class member sample is negative and the class member is recognized as positive (false positive).

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FN} + \text{FP} + \text{TN}) \quad (5)$$

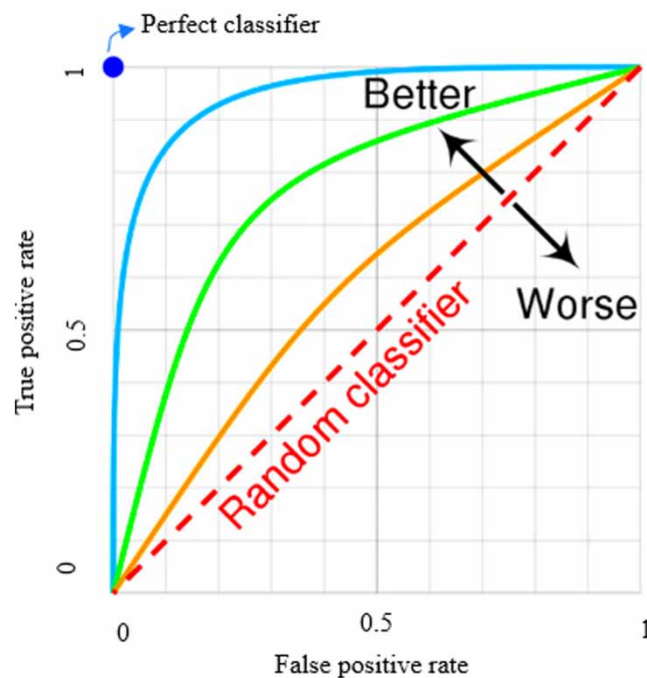
$$\text{Sensitivity (TPR)} = \text{TP} / (\text{TP} + \text{FN}) \quad (6)$$

$$\text{Specificity (TNR)} = \text{TN} / (\text{TN} + \text{FP}) \quad (7)$$

It is clear that an excellent prediction is a prediction whose sensitivity and specificity values are both 100%; however, the probability of this happening in reality is very low and there is always a minimal margin of error. The parameters of sensitivity and specificity, according to their nature, are always in competition with each other. That is, the increase of one is associated with the decrease of the other and vice versa. This situation has led to the production of another tool to evaluate the quality of categories.

2.7. Receiver operating characteristic

Receiver operating characteristic (ROC) is a curve that expresses the relationship between two parameters, sensitivity and property. The vertical axis of this graph shows the true positive rate (sensitivity), and the horizontal axis shows the value of the false positive rate. Different classification results indicate different points on this graph and finally form a curve. In the best case and assuming 100% correct classification in both categories, the corresponding point is the upper left corner point, i.e. point (0,1) and also assuming random classification, the corresponding point in the curve is one of the points

**Figure 6** | Receiver operating characteristic curve.

on the line between point (0,0) and point (1,1). In reality, the curve resulting from a classification is the curve between these two states (Figure 6).

The area under this graph (area under curve) is used as a measure to evaluate the performance of the category. According to the explanations given earlier, it is obvious that in the ideal case, the area under the curve is equal to its maximum value, that is 1. Therefore, the closer the area under the graph is to the number 1, the better the performance of the classifier. In addition to the two parameters of sensitivity and property, other parameters are also extracted from the confusion matrix, each of which expresses a concept and has different applications.

3. RESULTS AND DISCUSSION

3.1. Data acquisition and model structure

In this research, the accidents in a part of the network, including information such as pipe length, burial depth, slope, age, and failures that occurred in the network in 2014 and 2015 were investigated. Parameters such as slope, age, length, and depth of the burial were considered as inputs to the model SVM, and other influencing parameters such as groundwater level, soil type, number of branches, and type of sewage were omitted because of lack of sufficient information. Also, in each breakdown, the location of the clogging and its distance to the manhole were examined so that it could be used to find out whether the clogging was affected by the presence of the manhole or not. It is worth mentioning that the type of pipe in the network is concrete and its diameter is 250 mm, the distance between the pipes varies from 6 to 69 m, and the age of the pipes is 40–55 years.

According to the available incident information from the Supervisory Control and Data Acquisition (SCADA) system report, there were about 248 rows of data related to the number of failures that occurred in the network, The available data were divided into two parts of which one part is used for training process, and the other part is used for testing and the validation process. Approximately 70% of data are used for training and 30% for test and validation processes.

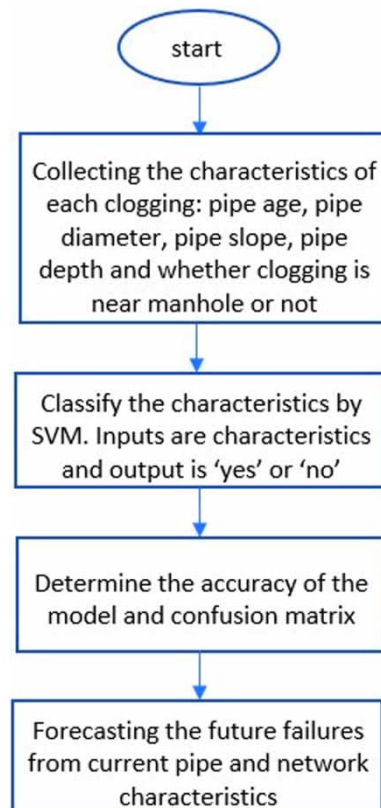


Figure 7 | SVM programming flowchart.

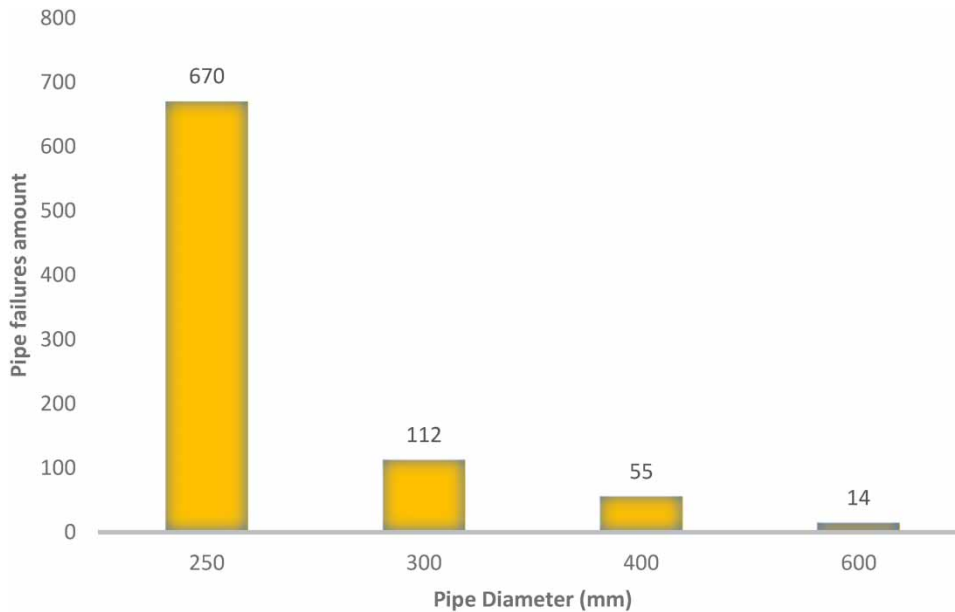


Figure 8 | Number of pipe failures based on the pipe diameter range.

Figure 7 shows the flowchart diagram of the SVM programming steps.

In this section, the number of blocked pipes in a part of the sewage network in Isfahan region was determined using the SVM model and the results were analyzed.

Various parameters were investigated to determine the factors affecting the failure from the SVM modeling. The results are presented in Figure 8, which shows the number of failures by sewer pipe diameter.

As Figure 8 shows, the highest number of failures is for the pipe with a diameter of 250 mm and the lowest number of failures for the pipe with a diameter of 600 mm.

Figure 4 shows the number of failures according to the slope of the sewer pipe.

Some other parameters were investigated by the SVM modeling to determine the factors affecting failure. The results are presented in Figure 9, which shows that most pipes of the network have slopes between 0.001 and 0.005.

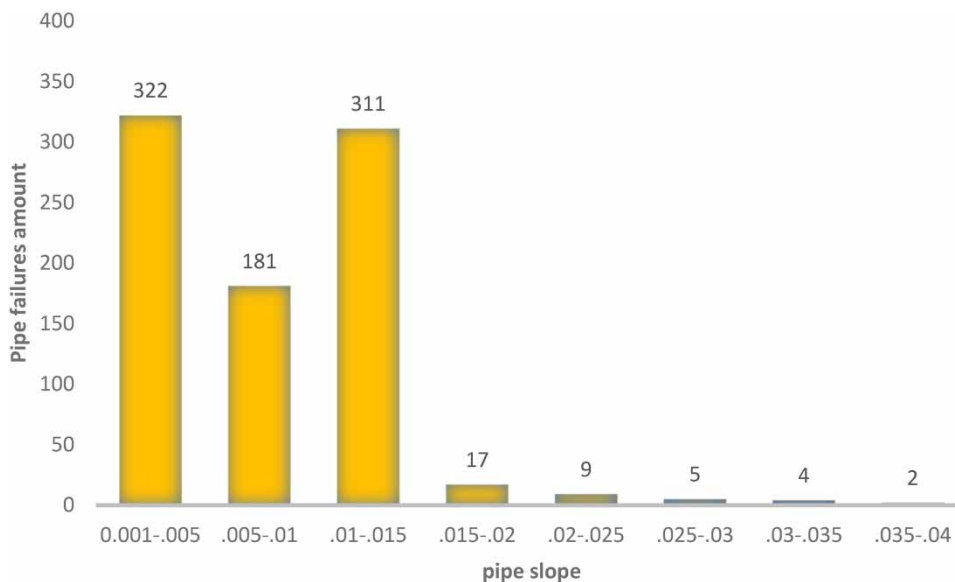


Figure 9 | Number of pipe failures based on the pipe slope range.

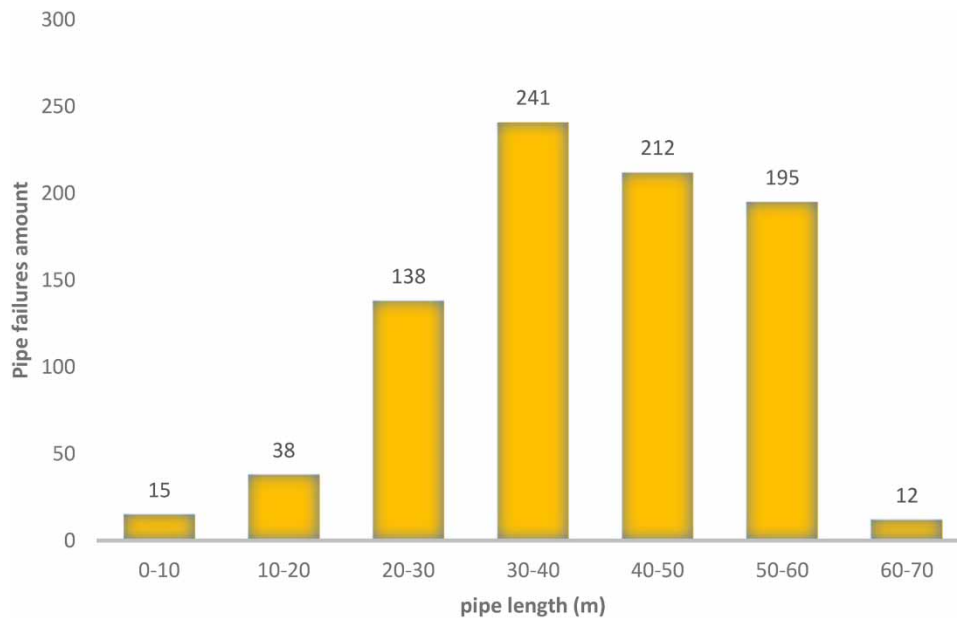


Figure 10 | Number of pipe failures based on the pipe length between manholes.

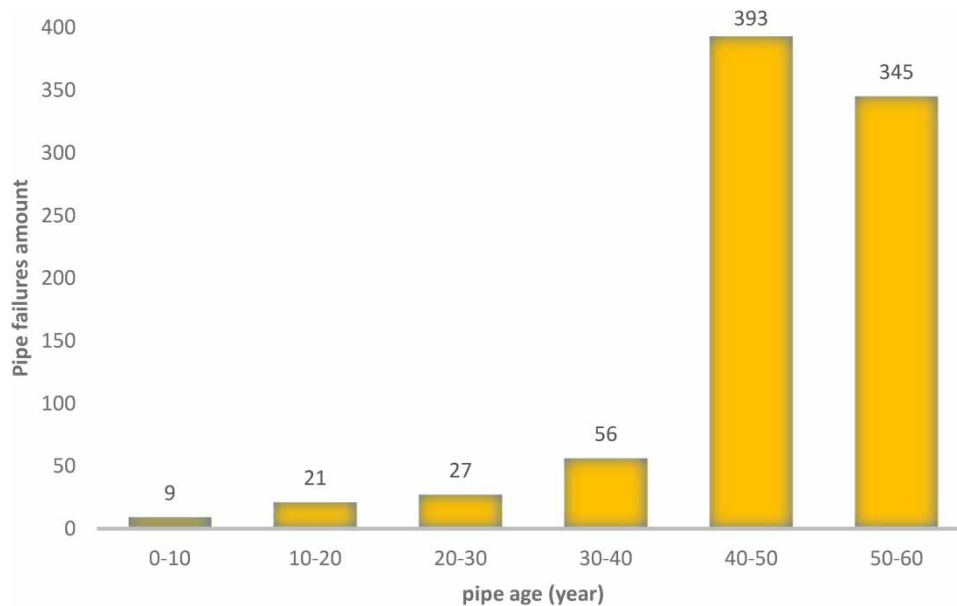


Figure 11 | Number of pipe failures based on the pipe age.

From [Figure 10](#), we can notice the number of failures that occurred by pipe lengths. Pipes with a length of 30–40 m account for the largest number of failures.

Other parameters were investigated by SVM modeling to determine the factors affecting failure. The results are presented in [Figure 11](#), which shows the number of failures by sewer pipe age. As shown in this picture, more failures occur in older pipes.

Some parameters like sewer pipe depth can also be investigated using SVM modeling to determine other factors affecting the failure. The results are presented in [Figure 12](#), which shows the number of failures by sewer pipe depth. Pipes at the depth of 3–4 m have the most failures as at that depth the combined force of traffic loads and soil pressure is maximum.

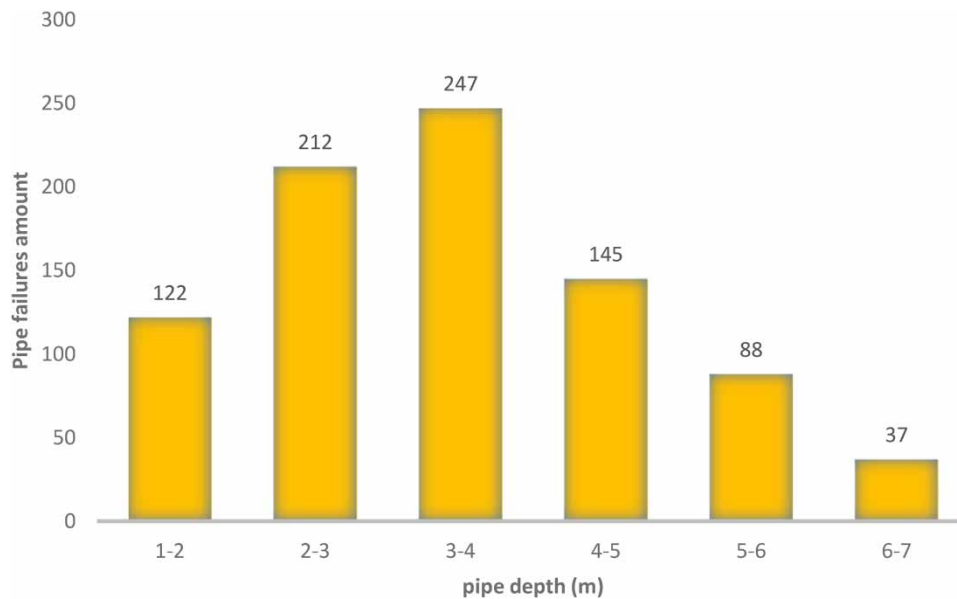


Figure 12 | Number of pipe failures based on the pipe depth.

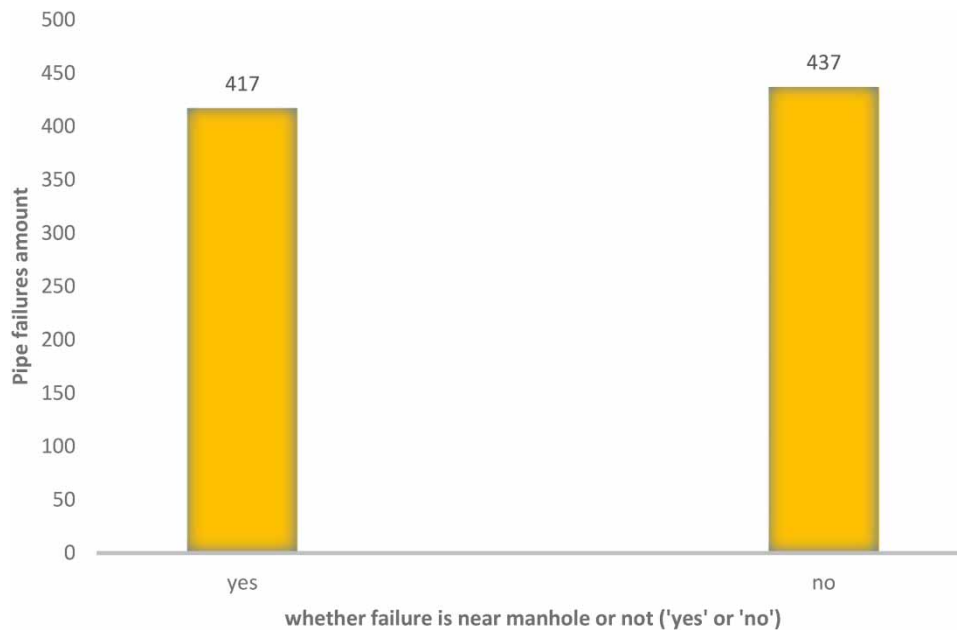


Figure 13 | Number of pipe failures based on whether the failure is near a manhole or not.

Parameters like effects of the existence of a manhole on pipe failure can also be investigated by SVM modeling to determine the factors affecting failure. The results are presented in [Figure 13](#), which shows how many pipe failures occur near the manhole and how many failures occur away from the manhole. Similarly, we can determine the effects of manhole on pipe failure.

In [Table 2](#), we can see the pipe parameters and if the failure occurs near a manhole (yes) or not (no).

[Figures 8–13](#) display the histogram for the pipe diameter, pipe slope, pipe length, pipe age, pipe depth, and whether the pipe is affected by the proximity of a manhole. It can be seen that the histograms of pipe length and pipe depth follow a normal distribution.

Table 2 | Details of first 53 obtained data of studied sewer network

Pipe ID	Diameter (mm)	Slope	Average cover depth (m)	Length of pipe (m)	Pipe age (year)	Manhole effect yes/no
1	250	0.003	2.05	50	40	Yes
2	250	0.005	2.34	50	42	Yes
3	250	0.005	2.104	50	42	No
4	250	0.005	2.34	50	42	No
5	250	0.004	1.47	31.5	42	No
6	250	0.004	1.84	50	42	Yes
7	250	0.0025	3.65	30	42	Yes
8	250	0.0025	3.61	30	42	No
9	250	0.0025	1.44	31.5	42	Yes
10	250	0.0025	2.06	40	42	No
11	250	0.0025	2.06	40	42	No
12	250	0.0025	2.06	40	42	Yes
13	250	0.0025	2.015	40	42	No
14	250	0.0025	2.015	40	42	Yes
15	250	0.0025	2.065	40	42	No
16	250	0.0025	2.015	40	42	No
17	250	0.0025	2.015	40	42	No
18	250	0.0025	2.015	40	42	No
19	250	0.0025	2.065	40	42	Yes
20	250	0.0025	1.95	45	42	Yes
21	300	0.003	2.04	43	40	No
22	300	0.0025	3.8	39	43	No
23	300	0.0025	3.8	39	43	Yes
24	300	0.0025	3.8	39	43	No
25	300	0.0025	4.24	39	43	No
26	400	0.0025	2.8	50.1	43	No
27	400	0.0028	3.77	63.8	43	No
28	400	0.0025	2.16	50	43	Yes
29	400	0.025	2.74	50	43	Yes
30	400	0.0025	2.94	50	43	No
31	400	0.0025	3.12	50	43	Yes
32	250	0.004	3.3	51	43	Yes
33	200	0.005	1.6	50	43	Yes
34	400	0.004	3.77	63.8	43	No
35	400	0.002	4.68	42.5	43	Yes
36	400	0.01	2.8	58.2	46	Yes
37	600	0.005	2.46	30	43	Yes
38	600	0.002	5.76	75	43	Yes
39	600	0.001	2.57	50	46	No
40	600	0.001	2.7	42	47	Yes
41	600	0.001	2.65	50	47	No
42	250	0.006	2.54	37	35	No

(Continued.)

Table 2 | Continued

Pipe ID	Diameter (mm)	Slope	Average cover depth (m)	Length of pipe (m)	Pipe age (year)	Manhole effect yes/no
43	250	0.004	2.59	50	35	No
44	250	0.006	2.48	16	36	Yes
45	250	0.005	2.55	50	58	Yes
46	250	0.01	1.56	36	39	No
47	250	0.01	3.5	30	40	No
48	250	0.01	2.5	50	40	No
50	250	0.01	1.99	50	41	Yes
51	300	0.0066	2.4	24	42	Yes
52	300	0.01	1.54	40	42	No
53	300	0.005	1.94	50	42	No

3.2. Statistical indices for evaluation of models

In this study, the following indices were used to evaluate the performance of models:

Root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

R-squared:

$$R^2 = \frac{\left(\sum_{i=1}^n (y_i - \bar{y}_i)(\hat{y}_i - \bar{\hat{y}}_i) \right)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2 \sum_{i=1}^n ((\hat{y}_i - \bar{\hat{y}}_i))^2} \quad (9)$$

Present bias:

$$\text{PBIAS} = 100 * \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{\sum_{i=1}^n y_i} \quad (10)$$

In [Table 3](#), the confusion matrix values for TP and TN are presented as 405 and 412, respectively, of the total 851 failures. Also, the values for FP and FN are 22 and 12, respectively.

[Table 4](#) provides the statistical outputs of the SVM model for the 851 input data. The accuracy of the model is 84% with a sensitivity of 88% and specificity of 80%. In another trial, instead of a random split for training/testing/validation set, the 10-fold cross-validation was used and an accuracy of 86% was obtained with the Area under ROC curve (AUC) of 0.905.

Table 3 | The confusion matrix amounts

Predicted values	Actual values	
	405	22
12	412	

Table 4 | Summary statistics parameters

Statistical parameters	Number of cases	851
Train		70%
Test and validation		30%
Test/train/validation accuracy		84%
Ten-fold cross-validation accuracy		86%
Sensitivity		88%
Specificity		80%
AUC (area under ROC curve)		0.905

From Table 4, we can see that the proposed classification method is appropriate for the quality analysis of registered damages of sewer pipes.

Figure 14 shows the effect of each of the parameters (pipe age, pipe slope, pipe depth, pipe diameter, and pipe length) on sewer pipe failures.

Table 5 provides the obtained results of the model (RMSE, R^2 , and PBIAS) from the SVM model compared with the results of related works.

The actual and predicted number of pipe failures for training and testing data are presented in Figures 15(a) and 15(b).

4. CONCLUSIONS

Sewer pipes are vitally important and play a very important role in modern cities because they are directly related to human health and the environment as an infrastructure network for the disposal of wastewater from the human living environment. Sewer pipes are expected to work unstopably to improve the public health and environmental protocols. An SPF has the potential to lead to a failure of public life. If sewer pipes fail, there is a risk of sewage overflow, which can contaminate local water supplies and endanger public health. For example, most of the US sewer infrastructure that is more than 100 years old, because of a combination of aging, chemical, and environmental factors, causes at least 23,000–75,000 of sanitary sewer overflow incidents per year. Artificial intelligence techniques such as SVM models can improve the accuracy and reduce the uncertainty of current forecasting models.

It seems necessary to check if the application of an SVM model as a selected ML prediction method could be useful for the rational management of buried infrastructure. Studies point out that an appropriate arrangement of the operational data that are registered in water and wastewater companies is necessary and should result in greater possibilities of using such information for the construction of a reliability model. The purpose of the correct arrangement of the operational data is to record

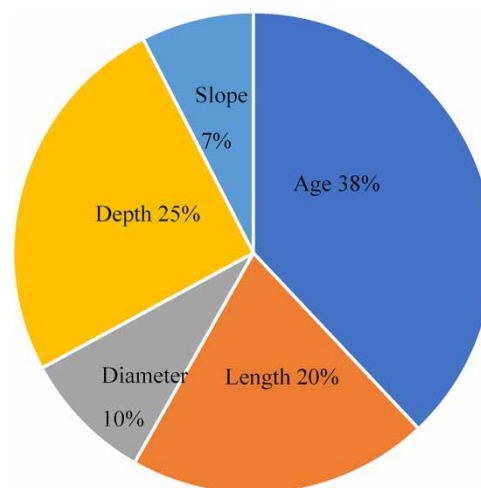
**Figure 14** | Influence of different parameters on pipe failures.

Table 5 | Summary statistics parameters

Methods	Model	Training			Test and validation		
		RMSE	R ²	PBIAS (%)	RMSE	R ²	PBIAS(%)
SVM (this work)		0.336	0.961	15	0.341	0.952	17
GP (Hoseingholi Moeini & Zare 2020)	1st	1.286	0.995		1.629	0.991	
	2nd	0.804	0.971		0.695	0.944	
	3rd	0.941	0.997		3.209	0.987	
GP (Hoseingholi & Moeini 2023)	1st	0.0036	0.969		0.0017	0.982	
	2nd	0.0031	0.966		0.0015	0.989	
	3rd	0.00074	0.996		0.0019	0.941	
	4th	0.0014	0.986		0.0055	0.615	
ANN	1st	0.0029	0.97		0.0017	0.97	
	2nd	0.0047	0.95		0.0019	0.98	
	3rd	0.0008	0.96		0.0014	0.96	
	4th	0.0019	0.97		0.0057	0.58	

Statistical characteristics of the applied hybrid models.
 ANN, artificial neural network; GP, genetic programming; SVM, support vector machine.

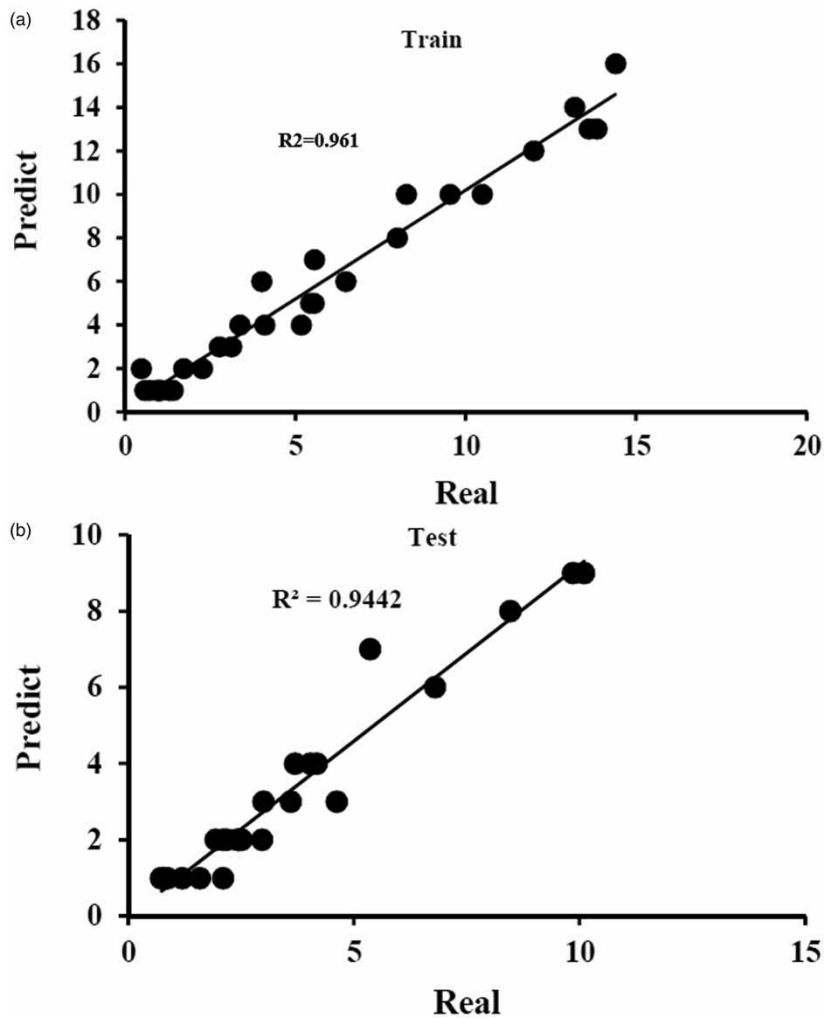


Figure 15 | Results of R² of the data, (a) train and (b) test.

the detailed information about the failure after the repair operation. Hence, in the water and wastewater companies, the failure of the sewer pipes and their types are accurately recorded, so that it can be used to predict future risks.

Manholes are the connection points of the sewer pipes and therefore play an important role in sewer networks. Hence, it is necessary to locate them on intersections or after some distance in sewer pipes to facilitate the wastewater flow. That's why it is important for us to know what the effect of manholes can be on the location of failures.

In this research, considering the novelty of the manhole effect on sewer pipe failures, investigations have been carried out to predict and determine the condition of the sewer network, and also the special benefits of knowing the distance to manholes. This was used to determine the number of clogged pipes in the sewer network as one of the indicators of hydraulic failure. As a case study, the sewage network of the city of Isfahan was investigated. It is worth mentioning that, considering most of the incidents that have occurred in these areas are of blockage type, in this research, the number of blockages in the network was predicted by presenting an SVM model. In this sense, by combining the parameters affecting the clogging of pipes, that is the age, type, depth, and diameter of the pipe, the number of clogging incidents of pipes in a part of the sewage network of Isfahan was predicted using the SVM model. The results show that an SVM model can properly approximate the manhole effects on sewer pipe failures with an accuracy of 86% and an AUC of 0.905. Hernandez *et al.* also used an SVM applicable tool for the main cities of Colombia, Bogotá and Medellín, for predicting sewer pipe structural conditions. Models were applied resulting in a deviation of less than 6% in the prediction of structural conditions in both cities (Hernández *et al.* 2021).

The accuracy of an SVM model is a measure of how well the model can predict the correct class or value for new data points. In terms of comparing the accuracy of SVM models with other methods, it is important to note that the accuracy of a model depends on the specific data set and the problem being solved. However, studies have shown that SVM models can outperform other algorithms. Classification works better. According to the literature on this research, the accuracy of 86% may be sufficient for the intended purpose of the model. However, to check more accurate methods, it is possible to use a comparison of different ML methods to check the performance accuracy of the manhole effect on the failure of sewer pipes as future research directions.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- Abraham, D. M., Wirahadikusumah, R., Short, T. J. & Shahbahrani, S. 1998 Optimization modeling for sewer network management. *Journal of Construction Engineering and Management* **124** (5), 402–410.
- Alotaibi, B. A., Baig, M. B., Najim, M. M. M., Shah, A. A. & Alamri, Y. A. 2023 Water scarcity management to ensure food scarcity through sustainable water resources management in Saudi Arabia. *Sustainability* **15** (13), 10648.
- Anbari, M. J. & Tabesh, M. 2015 Calculate the probability of failure event in wastewater collection networks using the Bayesian network. *Journal of Water and Wastewater* **27** (3), 48–61.
- Ariaratnam, S. T., El-Assaly, A. & Yang, Y. 2001 Assessment of infrastructure inspection needs using logistic models. *Journal of Infrastructure Systems* **7** (4), 160–165.
- Baik, H.-S., Jeong, H. S. & Abraham, D. M. 2006 Estimating transition probabilities in Markov chain-based deterioration models for management of wastewater systems. *Journal of Water Resources Planning and Management* **132** (1), 15–24.
- Barbieri, M., Barberio, M. D., Banzato, F., Billi, A., Boschetti, T., Franchini, S., Gori, F. & Petitta, M. 2023 Climate change and its effect on groundwater quality. *Environmental Geochemistry and Health* **45** (4), 1133–1144. <https://doi.org/10.1007/s10653-021-01140-5>.
- Barton, N. A., Farewell, T. S., Hallett, S. H. & Acland, T. F. 2019 Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks. *Water Research* **164**, 114926. <https://doi.org/10.1016/j.watres.2019.114926>.
- Barton, N. A., Farewell, T. S. & Hallett, S. H. 2020 Using generalized additive models to investigate the environmental effects on pipe failure in clean water networks. *npj Clean Water* **3** (1), 31. <https://doi.org/10.1038/s41545-020-0077-3>.
- Barton, N. A., Hallett, S. H. & Jude, S. R. 2022 The challenges of predicting pipe failures in clean water networks: A view from current practice. *Water Supply* **22** (1), 527–541. <https://doi.org/10.2166/ws.2021.255>.
- Cai, Z., Dziedzic, R. & Li, S. S. 2022 Water Distribution System Leak Detection Using Support Vector Machines. pp. 491–499. https://doi.org/10.1007/978-981-19-1065-4_41.

- Chughtai, F. & Zayed, T. 2008 Infrastructure condition prediction models for sustainable sewer pipelines. *Journal of Performance of Constructed Facilities* **22** (5), 333–341.
- Damvergis, C. N. 2014 Sewer systems: Failures and rehabilitation. *Water Utility Journal* **8**, 17–24.
- Duque, N., Duque, D., Aguilar, A. & Saldarriaga, J. 2020 Sewer network layout selection and hydraulic design using a mathematical optimization framework. *Water* **12** (12), 3337. <https://doi.org/10.3390/w12123337>.
- Ebtehaj, I. & Bonakdari, H. 2013 Evaluation of sediment transport in sewer using artificial neural network. *Engineering Applications of Computational Fluid Mechanics* **7** (3), 382–392. <https://doi.org/10.1080/19942060.2013.11015479>.
- Elwira, T. & Zielińska, A. 2017 Example of sewerage system rehabilitation using trenchless technology. *Ecological Chemistry and Engineering S.* **24** (3), 405–416.
- Engelbert, E. A. & Scheuring, A. F. 2023 *Water Scarcity: Impacts on Western Agriculture*. Univ of California Press.
- Fontecha, J. E., Agarwal, P., Torres, M. N., Mukherjee, S., Walteros, J. L. & Rodríguez, J. P. 2021 A two-stage data-driven spatiotemporal analysis to predict failure risk of urban sewer systems leveraging machine learning algorithms. *Risk Analysis* **41** (12), 2356–2391.
- Gedam, A., Mangulkar, S. & Gandhi, B. 2016 Prediction of sewer pipe main condition using the linear regression approach. *Journal of Geoscience and Environment Protection* **4** (05), 100–105. <https://doi.org/10.4236/gep.2016.45010>.
- Goodarzi, M. R. & Vazirian, M. 2023 A geostatistical approach to estimate flow duration curve parameters in ungauged basins. *Applied Water Science* **13** (9), 178. <https://doi.org/10.1007/s13201-023-01993-4>.
- Goodarzi, M. R., Niknam, A. R., Barzkar, A., Niazkar, M., Zare Mehrjerdi, Y., Abedi, M. J. & Heydari Pour, M. 2023 Water quality index estimations using machine learning algorithms: A case study of Yazd-Ardakan Plain, Iran. *Water* **15** (10), 1876. <https://doi.org/10.3390/w15101876>.
- Guo, Z., Hatakeyama, T., Yoshimura, C., Wang, T. & Hatano, Y. 2022 Basic influent sewage quality reflects sewershed characteristics in Tokyo city. *Journal of Water and Health* **20** (6), 972–984. <https://doi.org/10.2166/wh.2022.022>.
- Hahn, M., Palmer, R. N. & Merrill, M. S. 1999 Prioritizing sewer line inspection with an expert system. WRPMD'99: Preparing for the 21st Century, pp. 1–9.
- He, L. & Rosa, L. 2023 Solutions to agricultural green water scarcity under climate change. *PNAS Nexus* **2** (4), pgad117.
- Hernández, N., Caradot, N., Sonnenberg, H., Rouault, P. & Torres, A. 2021 Optimizing SVM models as predicting tools for sewer pipes conditions in the two main cities in Colombia for different sewer asset management purposes. *Structure and Infrastructure Engineering* **17** (2), 156–169. <https://doi.org/10.1080/15732479.2020.1733029>.
- Hoseingholi, P. & Moeini, R. 2023 Pipe failure prediction of wastewater network using genetic programming: Proposing three approaches. *Ain Shams Engineering Journal* **14** (5), 101958. <https://doi.org/10.1016/j.asej.2022.101958>.
- Hoseingholi, P., Moeini, R. & Zare, M. R. 2020 Number of blockage prediction for sanitary sewer networks (Case study: Isfahan Region 2). *Water and Wastewater (in Persian)* **31** (1), 86–98. <https://doi.org/10.22093/wwj.2019.143553.2731>.
- Huang, Z., Yuan, X., Liu, X. & Tang, Q. 2023 Growing control of climate change on water scarcity alleviation over northern part of China. *Journal of Hydrology: Regional Studies* **46**, 101332.
- Jia, Y., Zheng, F., Maier, H. R., Ostfeld, A., Creaco, E., Savic, D., Langeveld, J. & Kapelan, Z. 2021 Water quality modeling in sewer networks: Review and future research directions. *Water Research* **202**, 117419. <https://doi.org/10.1016/j.watres.2021.117419>.
- Kabir, G., Balek, N. B. C. & Tesfamariam, S. 2018 Sewer structural condition prediction integrating Bayesian model averaging with logistic regression. *Journal of Performance of Constructed Facilities* **32** (3), 4018019.
- Kakoudakis, K., Farmani, R. & Butler, D. 2018 Pipeline failure prediction in water distribution networks using weather conditions as explanatory factors. *Journal of Hydroinformatics* **20** (5), 1191–1200. <https://doi.org/10.2166/hydro.2018.152>.
- Kecman, V. 2005 'Support vector machines – an introduction BT – support vector machines: Theory and applications. L. Wang (ed.). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–47. Available at: https://doi.org/10.1007/10984697_1.
- Kutyłowska, M. & Kowalski, D. 2021 'Application of regression methods for classification of sewers' damages'. *Applied Water Science* **11** (9), 156. <https://doi.org/10.1007/s13201-021-01488-0>.
- Laakso, T., Ahopelto, S., Lampola, T., Kokkonen, T. & Vahala, R. 2018a Estimating water and wastewater pipe failure consequences and the most detrimental failure modes. *Water Supply* **18** (3), 901–909. <https://doi.org/10.2166/ws.2017.164>.
- Laakso, T., Kokkonen, T., Mellin, I. & Vahala, R. 2018b Sewer condition prediction and analysis of explanatory factors. *Water* **10** (9), 1239. <https://doi.org/10.3390/w10091239>.
- Li, W. & Achal, V. 2020 Environmental and health impacts due to e-waste disposal in China – a review. *Science of the Total Environment* **737**, 139745.
- Liu, J., Fu, Z. & Liu, W. 2023 Impacts of precipitation variations on agricultural water scarcity under historical and future climate change. *Journal of Hydrology* **617**, 128999.
- Mafarja, M., Aljarah, I., Faris, H., Hammouri, A. I., Al-Zoubi, A. M. & Mirjalili, S. 2019 Binary grasshopper optimisation algorithm approaches for feature selection problems. *Expert Systems with Applications* **117**, 267–286. <https://doi.org/10.1016/j.eswa.2018.09.015>.
- Malek Mohammadi, M., Najafi, M., Serajiantehrani, R., Kaushal, V. & Hajyalikhani, P. 2021 Using machine learning to predict condition of sewer pipes. *Pipelines 2021: Planning – Proceedings of Sessions of the Pipelines 2021 Conference*. American Society of Civil Engineers (ASCE), pp. 185–195. <https://doi.org/10.1061/9780784483602.022>.
- Mashford, J., Marlow, D., Tran, D. & May, R. 2011 Prediction of sewer condition grade using support vector machines. *Journal of Computing in Civil Engineering* **25** (4), 283–290.

- McDonald, S. E. & Zhao, J. Q. 2020 Condition assessment and rehabilitation of large sewers. In: *Underground Infrastructure Research*. CRC Press, pp. 361–369.
- McLaren, S. J., Sahli, M. W., Selig, S., Masten, S. J. & Page, W. H. 2022 The drinking water crises of Flint and Havelock North: A failure of public health risk management. *Journal of Water and Health* **20** (9), 1314–1328. <https://doi.org/10.2166/wh.2022.035>.
- Moeini, R. & Zare, M. R. 2021 Reduce repair cost of wastewater pipelines for using trenchless and open cut technology. *Ain Shams Engineering Journal* **12** (1), 423–435.
- Najafi, M. & Kulandaivel, G. 2005 Pipeline condition prediction using neural network models. *Pipelines 2005: Optimizing Pipeline Design, Operations, and Maintenance in Today's Economy*, pp. 767–781.
- Nalluri, C., Ghani, A. A. & El-Zaemey, A. K. S. 1994 Sediment transport over deposited beds in sewers. *Water Science and Technology* **29** (1–2), 125–133. <https://doi.org/10.2166/wst.1994.0658>.
- Nasrin, T., Sharma, A. & Muttil, N. 2017 Impact of short duration intense rainfall events on sanitary sewer network performance. *Water* **9** (3), 225. <https://doi.org/10.3390/w9030225>.
- Niazkar, M., Goodarzi, M. R., Fatehifar, A. & Abedi, M. J. 2023 Machine learning-based downscaling: Application of multi-gene genetic programming for downscaling daily temperature at Dogonbadan, Iran, under CMIP6 scenarios. *Theoretical and Applied Climatology* **151** (1–2), 153–168. <https://doi.org/10.1007/s00704-022-04274-3>.
- Pham, B. T., Tien Bui, D. & Prakash, I. 2018 Bagging based support vector machines for spatial prediction of landslides. *Environmental Earth Sciences* **77** (4), 146. <https://doi.org/10.1007/s12665-018-7268-y>.
- Potyralla, M. 2019 Geostatistical methods in water distribution network design – a case study. *Ecological Chemistry and Engineering S* **26** (1), 101–118. <https://doi.org/10.1515/eces-2019-0008>.
- Rezaei, H., Ryan, B. & Stoianov, I. 2015 Pipe failure analysis and impact of dynamic hydraulic conditions in water supply networks. *Procedia Engineering* **119**, 253–262. <https://doi.org/10.1016/j.proeng.2015.08.885>.
- Saleh, M. & Khakivatan, R. 2009 Problems of sewage networks operation and the role of video meter in better operation of networks. In *Second National Water and Wastewater Conference with Operation Approach*.
- Sattar, A. M. A., Gharabaghi, B., Sabouri, F. & Thompson, A. M. 2017 Urban stormwater thermal gene expression models for protection of sensitive receiving streams. *Hydrological Processes* **31** (13), 2330–2348. <https://doi.org/10.1002/hyp.11170>.
- Shehata, N., Egirani, D., Olabi, A. G., Inayat, A., Abdelkareem, M. A., Chae, K.-J. & Sayed, E. T. 2023 Membrane-based water and wastewater treatment technologies: Issues, current trends, challenges, and role in achieving sustainable development goals, and circular economy. *Chemosphere* **320**, 137993. <https://doi.org/10.1016/j.chemosphere.2023.137993>.
- Sourabh, N. & Timbadiya, P. V. 2018 Hydraulic and condition assessment of existing sewerage network: A case study of an educational institute. *Journal of The Institution of Engineers (India): Series A* **99** (3), 555–563. <https://doi.org/10.1007/s40030-018-0296-5>.
- Statistical Center of Iran 2016 Available from: <https://www.amar.org.ir>.
- Stip, C., Mao, Z., Bonzanigo, L., Browder, G. & Tracy, J. 2019 *Water Infrastructure Resilience: Examples of Dams, Wastewater Treatment Plants, and Water Supply and Sanitation Systems*. World Bank, Washington, DC.
- Vinson, A. W. 2017 *Evaluating the Potential of A Geospatial/Geostatistical Methodology for Locating Rain-Derived Infiltration and Inflow Into Wastewater Treatment Systems in the Minneapolis/St. Paul Metropolitan Area, Minnesota, USA*. Minnesota State University, Mankato.
- Vongvisessomjai, N., Tingsanchali, T. & Babel, M. S. 2010 Non-deposition design criteria for sewers with part-full flow. *Urban Water Journal* **7** (1), 61–77. <https://doi.org/10.1080/15730620903242824>.
- Xu, X., Zhou, Q., Chen, X., Li, Y. & Jiang, Y. 2022 The efficiency of green technology innovation and Its influencing factors in wastewater treatment companies. *Separations* **9** (10), 263. <https://doi.org/10.3390/separations9100263>.
- Zamanian, S. & Shafieezadeh, A. 2023 Age-dependent failure probabilities of corroding concrete sewer pipes under traffic loads. *Structures* **52**, 524–535. <https://doi.org/10.1016/j.istruc.2023.03.132>.
- Zheng, M., Wutich, A., Brewis, A. & Kavouras, S. 2022 Health impacts of water and sanitation insecurity in the Global North: A scoping literature review for U.S. Colonias on the Mexico border. *Journal of Water and Health* **20** (9), 1329–1342. <https://doi.org/10.2166/wh.2022.085>

First received 15 August 2023; accepted in revised form 21 January 2024. Available online 5 February 2024