14th International Congress on Civil Engineering, 21-23 October 2025
Sharif University of Technology, Tehran, Iran

Evaluation of Decision-Making Prediction Models for Sewer Pipes Asset Management

Salar Shirkhanloo[1], Sadegh Ghavami[2], Mohammad Najafi[3], Parang Aminifard[4]

1- Clinical Assistant Professor, Department of Mechanical Engineering, University of North Texas, Texas, United States

2- Assistant Professor, Faculty of Civil Engineering, Sahand University of Technology, Tabriz, Iran

3- Associate Professor, Department of Civil Engineering, University of Texas at Arlington, Texas, United States

4- M.Sc. Graduate, Faculty of Civil Engineering, Sahand University of Technology, Tabriz, Iran

salar.shirkhanloo@unt.edu

Abstract

Wastewater collection systems deteriorate over time, requiring continuous adjustments and the development of asset management frameworks by utility owners to maintain the performance of their assets. While closed-circuit television (CCTV) is the primary method for inspecting sewer pipes in the U.S., it is both costly and time-consuming. Therefore, the primary objective of this research is to develop predictive models based on various machine learning algorithms that can forecast the future conditions of sewer pipes. The results of the models can be used to prioritize the need for sanitary sewer pipe inspections, rehabilitation, and replacements. Data collected from the cities of Dallas and Tampa were combined in this research. This dataset included nine independent variables: pipe age, size, length, material, surrounding soil type, soil pH, depth, slope, and surface conditions. The dependent variable was the condition rating of the sewer pipe based on PACP scores, which ranged from 1 to 5. Among models, tree-based models performed better than other models, and the bagging approach was more efficient than boosting techniques. Additionally, it was found that the age and length of the pipes had the greatest effect on the condition rating of sewer pipes, while the pipe location had the least impact.

Keywords: Sewer pipe, Asset management, Prediction model, Machine learning.

1- INTRODUCTION

The underground infrastructure networks cover thousands of kilometers and are an essential part of the overall infrastructure of the United States [1]. Some elements of the wastewater infrastructure in the United States are over a century old, and a combination of age, malfunctions, and accidents results in at least 23000 to 75000 sanitary sewer overflows per year [2]. Deterioration in sewer pipes, as an essential part of wastewater systems, could have various social, environmental, and economic consequences, including poor water quality, chemical or biological pollution, disease, and high maintenance costs [3]. The inspection and monitoring of all sewer pipelines in metropolitan locations, which are often hidden and underground, is nearly impossible due to financial, time, and assessment technology limitations. Pipeline Assessment and Certification Program (PACP) is the North American standard for pipeline defect identification and assessment to identify the pipe condition and manage the sewer pipe networks. PACP assesses the condition of pipes on a scale of 1 to 5 based on the results obtained from closed-circuit television (CCTV) inspections and operator judgments: (1) minor defects; (2) defects that have not begun to deteriorate; (3) moderate defects that will continue to deteriorate; (4) severe defects that will become Grade 5 defects within the foreseeable future; (5) defects requiring immediate attention. The PACP condition grading method assigns a ranking to pipe segments depending on the severity of the identified defects and problems. Several attempts have been made in recent years to assess the state of sewage pipes and to identify the factors that impact degradation [4, 5]. Table 1 shows these factors.

Condition prediction models may be used to estimate the condition rating of infrastructure using data from inspection databases. Deterioration models for sewer pipelines are classified into different categories. Thus, existing sewer deterioration models can be classified into two groups: statistical and artificial intelligence (AI) models. Statistical models include linear and logistic regressions. According to Luger (2009) [6], artificial intelligence can be decomposed into several categories, including game playing, automated reasoning and theorem proving, expert systems, natural language understanding and semantics, modeling human performance, planning and robotics, languages and environments for AI, machine learning, neural networks and genetic algorithms, AI and philosophy.

Table 1- Factors Affecting Wastewater Pipes Deterioration [4, 5]

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

Bakry et al. (2016) [7, 8] developed a condition prediction model for sewage pipes that had previously been restored using the cured-in-place pipe (CIPP) approach, employing regression analysis techniques. Generally, the linear regression model is too simple to describe the probabilistic nature of pipe degradation and is not suitable for predicting discrete condition values [9, 10]. Using logistic regression, Ana et al. (2009) [11] investigated the impact of sewage physical parameters on sewer pipeline structural degradation. In this study, the age, material, and length of the sewers were determined to be important, and no validation procedure was utilized to confirm the findings. Malek Mohammadi (2019) [12] used logistic regression to predict the condition of sewer pipes for the city of Tampa, Florida. According to the results of his model, the condition of 65.8% of sanitary sewer pipes was predicted correctly; however, pipes in condition levels 2, 3, and 4 were not estimated accurately. Najafi and Kulandaivel (2005) [13] created a prediction model using an artificial neural network. The study revealed that using a neural network to construct a condition prediction model for pipelines is viable; however, model accuracy is strongly dependent on a larger and more inclusive sample set. Tran et al. (2007) [9] developed a neural network deterioration model to forecast the serviceability of underground stormwater pipelines. In this work, the model was calibrated using the Markov Chain Monte Carlo simulation. According to the findings of the study, the performance of a neural network calibrated using the Markov chain approach outperforms that of a neural network calibrated with the backpropagation method. Mashford et al. (2011) [14] used a support vector machine to forecast sewage pipeline condition grades. The study's findings revealed that the support vector machine has extremely strong prediction performance with 91 percent accuracy and may be utilized as a novel method to simulate sewage pipe degradation. To estimate the structural integrity of individual sanitary sewage pipes, Harvey and McBean (2014) [15] employed a random forests model. According to the findings, random forest models can accurately estimate the state of individual sewage pipes with an area under the receiver operating characteristic (ROC) curve of 0.81. Random forest prediction models have the potential to cut project costs and time in half, and this method may be used to assess the status of uninspected sewage pipelines. To forecast sewage pipeline condition ratings, Laakso et al. (2018) [16] used random forest and binary logistic regression. The models' accuracy was 62% for binary logistic regression and 67% for random forest, respectively. The study found that both logistic regression and random forest models may be utilized to forecast sewage pipeline condition in the future.

Pipe degradation is a complicated process, as detailed in earlier sections, and no one element may be the cause of pipe deterioration. Furthermore, wastewater agencies and municipalities are frequently short on funds to examine the status of all pipes in the network regularly. As a result, an alternate method must be adopted to cut inspection costs while still providing a thorough plan for prioritizing and scheduling inspections. Condition prediction models for individual sewage pipes have yet to be thoroughly investigated, and the majority of research has concluded that novel data analysis methodologies can be used to estimate future sewer conditions and behavior. The first objective of this study is to establish a decision-making support tool as a condition prediction model for sanitary sewer asset management. Assessment of the prediction results of different models for the case study could help cities design strategic plans for their sewer networks. The secondary objective of this research is to analyze the differences in the identified significant factors affecting sewer pipe deterioration. Agencies and municipalities can gather fewer data points during inspections by identifying influencing variables.

2- DATA COLLECTION

This study is based on the combined data collected from the Dallas Water Utilities Wastewater Collection System (Dallas, TX) and the City of Tampa's Wastewater Department (Tampa, FL). The purpose of combining the two separate datasets was to have more diverse data about sewer pipes and their environmental conditions to develop a more comprehensive model. Additionally, increasing the data to achieve a more accurate model was another reason for combining the datasets.

CCTVs are widely used in the United States to inspect sewer pipes [17]. They are employed in both cities' sanitary sewer pipe inspection and condition assessment processes. Pipeline Assessment and Certification Program (PACP) guidelines were used to evaluate the condition of pipes in both cities on a scale of 1 to 5. The inventory of the sewer system is stored using geographic information system (GIS) databases. The recorded database inventory includes information such as pipe installation details, pipe location concerning geographical maps, and so on.

One of the important steps in preparing data for further analysis is pre-processing. The collected data is processed to avoid any null values as part of the data preparation process. The dataset's null values were excluded from further analysis. Additionally, some pipes with minority materials such as asbestos-cement (AC), cast iron (CI), ductile iron (DI), prestressed concrete cylinder pipe (PCCP), and high-density polyethylene (HDPE) were excluded to avoid any misclassification or error during the modeling process. Individual features such as age were calculated based on collected data as the next step in data preparation to include in the model development phase. Ultimately, the surface condition of each pipe segment (the road type that the pipe is buried beneath) was identified using GIS and added to the dataset. The final dataset for Dallas City, containing 3104 data points, was used for analysis and model development. In the Dallas City dataset, the soil types were sand, loam, clay, and rock, and the pipe materials were polyvinyl chloride (PVC), vitrified clay pipe (VCP), and reinforced concrete (RC).

In the case of the Tampa City dataset, the sewer inventory dataset included 5144 manholes and manhole pipe segments. To make it easier to track individual pipes, each pipe segment was assigned a unique number (Facility ID). In addition, the dataset came with a shapefile (GIS file). The dataset included pipe attributes such as installation date, material, diameter, length, depth, down elevation, up elevation, and location. As a first step, pipes with missing information on installation date, depth, material, length, and condition scales were excluded from the dataset, totaling around 2000 segments. Then, pipe ages and slopes were calculated based on available data. In the next step, pipe materials with a low population in the dataset, such as ductile iron, reinforced concrete, and plastic pipes, were removed. The total number of these pipes was approximately 200. Also, using the mentioned shapefile, soil data of pipes, including soil pH and the type of soil surrounding pipe segments and surface conditions, were extracted. Finally, the condition rating of each pipe was obtained from the dataset. The overall condition of pipes existed in this database, in addition to some information such as pipe rating, quick rating, and pipe rating index for structural and operational conditions. The final dataset of Tampa City contains 2,944 individual pipe segments. The soil type in Tampa City included sand and gravel, and the pipe materials were polyvinyl chloride (PVC) and vitrified clay pipe (VCP).

Table 2- Variables Included in Sewer Pipe Dataset

Category	Variable	Description
Physical	Age (years)	Time difference between the installation date of the pipe segment and the date of inspection
	Material	Type of pipes material (PVC, VCP, and RC)
	Diameter (in)	Diameter of the pipe segment
	Depth (ft)	Depth of overburden above the pipe segment
	Slope (%)	Vertical displacement of the pipe segment per horizontal displacement
	Length (ft)	Length of the pipe segment between two manholes
Environmental	Soil Type	Type of soil surrounding the pipe (Sand, Gravel, Loam, Clay, and Rock)
	Soil pH	A numerical expression of the relative acidity or alkalinity of a soil sample
	Pipe Location	Category of the ground surface where the pipe is located (Highway, Street, Alley, and Easement)

Finally, two datasets were combined into one set, and boxplot techniques were used to remove outliers from the dataset. Outliers numerically distant from the rest of the data are frequently found in observed datasets. Outliers are typically larger or smaller than the observed values in the dataset. A boxplot is a graphical tool for displaying the variation of continuous data. The boxplot identifies the median, lower quartile, upper quartile, and upper extreme. Table 2 shows that the final dataset contains 4,803 individual pipe segments with various physical and environmental parameters. The final dataset for analysis consisted of nine independent variables and one dependent variable (Condition Rating).

3- RESULTS AND DISCUSSIONS

Logistic Regression

The first method developed by Python using a resampled dataset was Multinomial Logistic Regression (MLR). It can be used as an extension of binary logistic regression when the dependent variable is categorical and contains more than two levels. Firstly, the dependent and independent variables were defined. Then, using the train_test_split method of Python, 80% of the data were set for training the model and 20% for testing. The sklearn_linear_model Logistic Regression library of Python was used to implement the Logistic Regression model.

To be consistent with other AI methods implemented in this study, the evaluation metric for the developed Multinomial Logistic Regression model was selected to be the confusion matrix instead of the classification table. Figure 1 shows the confusion matrix of this model. The high values in the non-diagonal cells of the confusion matrix below demonstrate that misclassification was high in the developed MLR prediction model. Specifically, pipes with condition ratings of 3 had significant misclassification with pipes having condition ratings of 2 and 4. Another important metric to evaluate the performance of the developed model is the ROC curve, including the AUC criteria. As can be seen in Figure 1, the ROC curves of classes 2, 3, and 4 are not close to the upper left corner. However, detailed measurements should be investigated to assess the model's effectiveness. They are calculated based on the confusion matrix and are shown in Table 3.

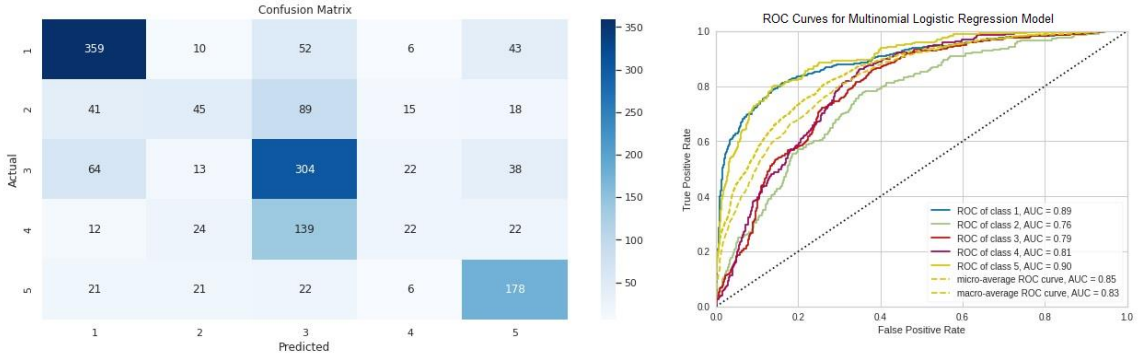

Figure 1- Confusion Matrix and ROC Curves for Multinomial Logistic Regression Model

Table 3- Precision, Recall, and F1-Score Metrics for Multinomial Logistic Regression Model

Condition Rating	Precision	Recall	F1-Score
1	0.72	0.76	0.73
2	0.39	0.23	0.29
3	0.50	0.68	0.57
4	0.30	0.10	0.15
5	0.59	0.71	0.64
Macro - Average	0.49	0.48	0.48

F1-scores of pipes with condition ratings of 2 and 4 are very low, indicating the weak performance of the developed model in predicting these classes. A precision value of 0.59 for class 5, which is an important class in sewer prediction modeling, shows that the model only correctly identified 59% of pipes with a condition rating of 5 out of all the pipes classified as having a condition rating of 5. Finally, the macro-average F1-score of 0.48, the summary result of the model's efficiency, indicates that the developed MLR model was unsuccessful.

KNN

A 5-fold cross-validation method was used to develop the models, including KNN. It randomly selected 80% of the data for training and 20% for testing the model. The Python Scikit-learn library's K-neighbors classifier parameters are shown in Table 4.

Table 4- Parameters of the Developed KNN Model

Parameters	Description
n_neighbors	Specify number of neighbors: 7
weights	weight function used in prediction: uniform, distance
algorithm	Algorithm used to compute the nearest neighbors: auto, ball_tree
leaf_size	This parameter is estimated by ball_tree

To maintain the model's consistency, the weight parameter was set uniformly. Some parameters, including leaf size, were set to default values to create the KNN model. The nearest neighbors were calculated using the auto algorithm because this function tries to discover the best algorithm. The most crucial step in the KNN model's development is determining the number of neighbors (K). When smaller values are chosen, there is generally a high risk of overfitting. This parameter can be determined manually. Thus, the model was run using different K values from 3 to 11. The K value should be odd because of the voting issue during the KNN model development. Therefore, using the numbers 3, 5, 7, 9, and 11 as the values of K, various KNN models were implemented, and the highest accuracy was achieved when the number of neighbors was 7.
The performance of the KNN model developed using resampled data is reviewed in this section. The number of instances in each class in the oversampled dataset matches that of the dominant class. As seen in Figure 2, a confusion matrix was produced using the generated KNN model. It should be noted that the confusion matrix below is based on the testing portion of the developed model and serves as validation for the model.

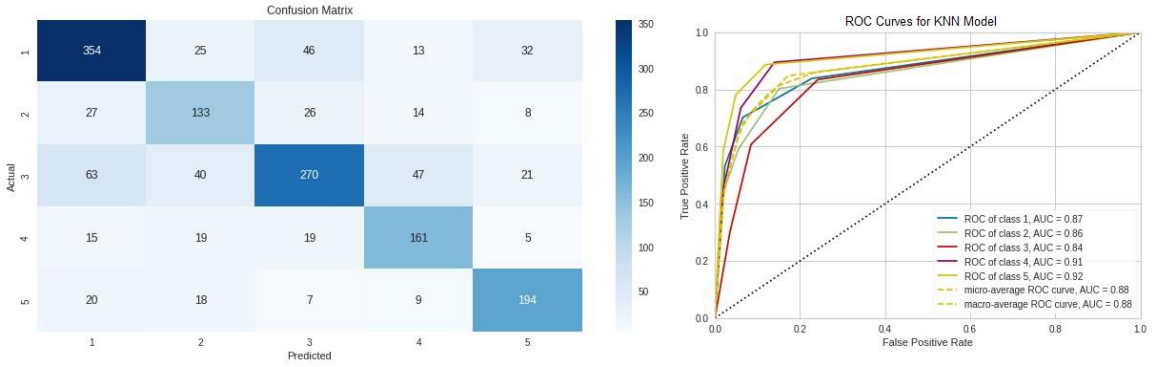

Figure 2- Confusion Matrix and ROC Curves for KNN Model

The sum of the values in the matrix should equal 1588, which is valid for the confusion matrix in Figure 2. The confusion matrix's main element is the diagonal cells showing the model's performance for each class. Higher numbers and darker backgrounds in these cells indicate better performance of the model for each class. Consequently, lower numbers and brighter backgrounds in other cells are preferred since they represent misclassified instances. ROC curves were plotted, as shown in Figure 2, to better understand the performance of the KNN model. The model has higher overall accuracy when the ROC curve is closer to the upper left corner; consequently, the AUC value is close to 1.
AUC values were closer to 1 for curves corresponding to condition ratings of 4 and 5, demonstrating the model's high accuracy in predicting these classes. AUC was found to be 0.87, 0.86, and 0.84 for classes 1, 2, and 3, respectively. Evaluation metrics, including precision, recall, and finally, the F1-score, were calculated using the generated confusion matrix. Table 5 displays the recall, precision, and F1-score of the developed KNN model.

Table 5- Precision, Recall, and F1-Score Metrics for KNN Model

Condition Rating	Precision	Recall	F1-Score
1	0.74	0.75	0.74
2	0.57	0.64	0.60
3	0.74	0.61	0.67
4	0.66	0.74	0.70
5	0.75	0.78	0.76
Macro - Average	0.70	0.71	0.70

To explain Table 5, the metrics for pipes with a condition rating of 4 are described. A precision value of 0.66 shows that the model correctly identified 66% of pipes with a condition rating of 4 out of all the pipes classified as having a condition rating of 4. The recall value of 0.74 shows that the model correctly identified 74% of pipes with a condition rating of 4 out of all the pipes having a condition rating of 4. Finally, the F1-score, by combining them into a single measure, evaluates the model's effectiveness. An F1-score of 0.70 shows the ability of the model to predict pipes with a condition rating of 4. An F1-score closer to 1 indicates better performance of the model. An overall macro-average F1-score of 0.70 shows an acceptable performance of the developed KNN model for the available sewer pipes dataset. It was found that the KNN model had better performance for pipes with a condition rating of 5 (F1 = 0.76) rather than for others and had the lowest accuracy in predicting pipes with a condition rating of 2 (F1 = 0.60).

Tree-Based Models

To compare the performance of different machine learning methods to achieve the most accurate model, one of this study's main goals, the tree-based models were developed. In the beginning, the regular Decision Tree classifier was tested.

The main parameters in training a Decision Tree model are maximum depth and splitting criterion. For the criterion, the Gini index was selected as the default. For maximum depth, researchers suggest a trial-and-error process or a depth equal to the number of the dataset's attributes, which means the number of variables [18, 19]. A trial-and-error process was conducted, and the maximum depth of 11 was selected as the best. The resultant Decision Tree model was used to create a confusion matrix, as shown in Figure 3. The interesting point is that despite the model's good performance regarding classes 1 and 3, the misclassification between these two is a little high. Figure 4 illustrates the ROC curve of the model. According to the figure, the model has a higher AUC for pipes with poor condition ratings (PACP 4 and 5) than for other classes, which is a good sign. According to Table 6, the developed model performs less well for pipes with a condition rating of 2 than it does for other pipes. This might be because there are only a few pipes in the dataset in this situation. Additionally, it can be observed that pipes in class 5 have the greatest F1-scores, and the model can accurately forecast them. Overall, the Decision Tree classifier model, which has an F1-score of 0.73, outperforms the KNN model.

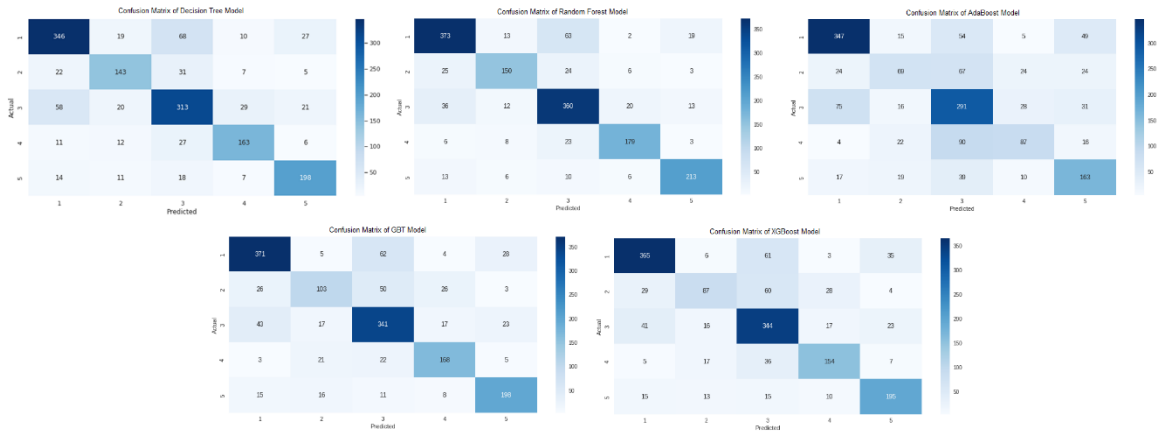

Figure 3- Confusion Matrix of Tree-Based Models

The first ensemble of the Decision Tree algorithm to be tested was the Random Forest. The Random Forest algorithm is based on the Bagging approach. It constructs different decision trees and reaches the best result by taking the average of the final results of the trees. The randomness in this method means that each tree is constructed with a random dataset and a random variable. The main parameters for this method that should be assigned are the number of decision trees (n_estimators) and the number of features to be analyzed in each tree (n_features). A higher number of constructed trees usually leads to higher accuracy [19]. Therefore, after the trial-and-error process, the number of trees was set to 100. Thus, 100 distinct trees were constructed using 100 different datasets randomly selected from the original dataset. Next, in each tree, the algorithm randomly selects a subset of the features (independent variables) and looks for the best possible test involving one of these features. The number of features that are selected is controlled by the max_features parameter. Max_features is a critical parameter in this method. If we set it equal to the number of independent variables (which in our case is 9), that means that each tree can look at all variables in the dataset, and no randomness will be injected into the feature selection (though the randomness due to the bootstrapping of data remains). If we set it to 1, the trees have no choice on which feature to test and can only search over the randomly selected variable. As a result, a high max_features value indicates that the trees in the random forest will be relatively similar and easily fit the data using the most distinguishing features. The trees in the random forest will be significantly different if the max_features is low, and each tree may need to be very deep to fit the data adequately [18]. It is possible to leave the selection of the best value for this parameter to the algorithm itself. Consequently, this parameter was decided to be determined by the algorithm, and it had to be a number between 1 and 9. Finally, the prediction model for sewer pipe condition assessment by the RF algorithm was trained via Python, and the results are shown in this section. Figure 3 depicts the confusion matrix of the testing part of the developed model. Relatively low values in the non-diagonal cells are a sign of the model's outstanding performance. ROC curves for all classes were plotted, as shown in Figure 4. It can be seen that the AUC for the ROC curves of condition ratings 4 and 5 are close to the unit. Generally, in sewer pipe prediction modeling,

high values in metrics related to minority conditions (PACP 4 and 5) show the high efficiency of the model. ROC curves of other classes are also in excellent condition. Table 6 illustrates other metrics for the developed RF model. F1-scores of classes 1, 4, and 5 are high, indicating excellent performance in predicting them. The precision of 0.82 for class 1 reveals that the model correctly identified 82% of pipes with a condition rating of 1 out of all the pipes classified as having a condition rating of 1. Since a high quantity of pipes belongs to this class, this precision value is considered high accuracy. Furthermore, a recall value of 0.86 for class 5 indicates that the model correctly identified 86% of pipes having a condition rating of 5 out of all the pipes having a condition rating of 5. It is a symbol of low misclassification of the developed model. Finally, the overall macro-average F1-score of 0.8 shows that the model had a very good performance in predicting all classes and outperforms the KNN and Decision Tree models.

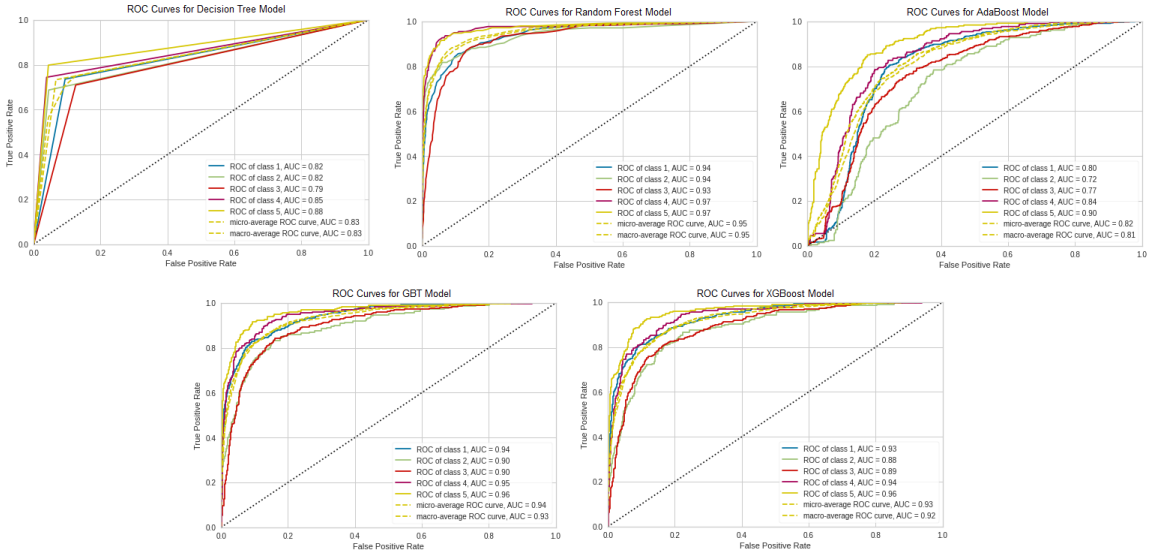

Figure 4- ROC Curves for Tree-Based Models

AdaBoost is the first algorithm among the boosting algorithms to be tried. To compare their findings to those of other tree-based models, boosting techniques were applied. AdaBoost simply calculates the predictions from each predictor and weighs them using the predictor weights (the higher the weight of the predictor, the more accurate the predictor is). The class that receives the majority of weighted votes is the predicted class [19]. The number of trees and the learning rate, which controls how much each tree is permitted to correct the errors of the previous trees, are the two key parameters of the boosted models. These two factors are related because more trees are required to construct a model with the same level of complexity at a lower learning rate. In boosting models, adding more estimators makes the model more complex, which might result in overfitting, in contrast to random forests, where having more estimators (predictors) (trees) is always beneficial. It is common practice to fit several estimators based on the time and memory budget before looking through various learning rates. A common practice is to fit some estimators depending on the time and memory budget and then search over different learning rates [18]. By these explanations, the number of predictors was set to 50, and the learning rate was set to 1 as default. It should be noted that decreasing the learning rate would cause an increase in the number of trees. The confusion matrix of the developed model is shown in Figure 3. The values in non-diagonal cells are high, illustrating relatively high misclassification among different classes. The misclassification between classes 3 and 4 is higher than in others. The ROC curve of this model is depicted in Figure 4. It can be seen that the AUC of classes 2 and 3 is relatively low. Other metrics should be discussed to evaluate the performance of the model. The precision, recall, and F1-score of all classes are shown in Table 6. The low F1-scores for pipes with condition ratings of 2 and 4 show the weak performance of the developed model for these classes. Also, the precision and recall values of the model for pipes with a condition rating of 5, which has always been a sign of the model's performance, are low. Finally, the overall macro-average F1-score of 0.57 indicates that AdaBoost was not a suitable model for the available dataset.

One significant drawback of this sequential learning technique is that it cannot be parallelized (or only partially) since each predictor can only be trained after the previous predictor has been trained and evaluated. As a result, it does not scale as well as bagging. The other disadvantage of boosting is that it is sensitive to outliers since

every classifier must fix the errors of its predecessors. Thus, the method is too dependent on outliers [18, 19]. These two disadvantages could be the reason for the weak performance of the developed AdaBoost model.
Another boosting approach tried in this study was the Gradient Boosting Tree. This method operates similarly to AdaBoost. The only difference is that this method modifies the residual errors instead of modifying instance weights. Again, the main parameters are the number of trees and the learning rate. The learning rate was set at one as the default. You can employ early stopping to determine the ideal number of trees. Using the staged_predict Python code, which returns an iterator over the predictions made by the ensemble at each stage of training, is an easy method to accomplish this (with one tree, two trees, etc.). The code first trains a GBT ensemble with 120 trees, measures the validation error at each training stage to determine the ideal number of trees, and then repeats the process with the ideal number of trees [19]. The confusion matrix of the developed Gradient Boosting Trees model is shown in Figure 3. The values on the diagonal elements are evidence of a relatively proper result for the developed model, specifically for pipes with condition ratings of 1 and 3. Another tool to evaluate the developed GBT model was the ROC curve, which is shown in Figure 4. Unlike the AdaBoost model, AUC values for all classes are higher than 0.9, which indicates that the model had better performance than the previous boosting model. Table 6 explains other evaluation metrics for the GBT model. Except for pipes with a condition rating of 2, precision and recall values for other classes are acceptable. The calculated F1-score shows that this model is much more accurate than the AdaBoost model. One reason for the better performance of GBT compared to AdaBoost could be the optimal number selection of trees, which was set to be done automatically. The other reason could be that GBT uses a loss function (such as squared error) to correct the errors of the prior tree, which has been more effective than the AdaBoost approach, which uses modifying instance weights to correct the previous trees.

Table 6- Precision, Recall, and F1-Score Metrics for Tree-Based Models

Model	Condition Rating	Precision	Recall	F1-Score
Decision Tree	1	0.74	0.74	0.74
	2	0.70	0.69	0.69
	3	0.69	0.71	0.70
	4	0.75	0.74	0.74
	5	0.77	0.80	0.78
	Macro - Average	0.73	0.74	0.73
Random Forest	1	0.82	0.80	0.81
	2	0.80	0.72	0.76
	3	0.75	0.81	0.78
	4	0.84	0.82	0.83
	5	0.85	0.86	0.85
	Macro - Average	0.81	0.80	0.80
AdaBoost	1	0.74	0.74	0.74
	2	0.49	0.33	0.39
	3	0.54	0.66	0.60
	4	0.56	0.40	0.47
	5	0.58	0.66	0.62
	Macro - Average	0.58	0.56	0.57
GBT	1	0.81	0.79	0.80
	2	0.64	0.50	0.56
	3	0.70	0.77	0.73
	4	0.75	0.77	0.76
	5	0.77	0.80	0.78
	Macro - Average	0.73	0.73	0.73
XGBoost	1	0.80	0.78	0.79
	2	0.62	0.42	0.50
	3	0.67	0.78	0.72
	4	0.73	0.70	0.71
	5	0.74	0.78	0.76
	Macro - Average	0.71	0.70	0.70

The XGBoost method is the same as GBT but faster. This model's confusion matrix is shown in Figure 3. For pipes with condition ratings of 1 and 3, which are the most numerous, the non-diagonal cell values show no significant misclassification. This indicates that the model is operating within acceptable bounds. The ROC curve of the developed XGBoost model is illustrated in Figure 4. It can be seen that the AUC of all classes is relatively high, indicating low misclassification among different classes. The macro-average AUC value of 0.92 is close to GBT's. The generated XGBoost model's final evaluation metrics are described in Table 6 at the end. Similar to the GBT model, satisfactory results for pipes with condition ratings of 1, 3, 4, and 5 are

visible. A precision value of 0.80 for class 1 and a recall value of 0.78 for class 5 are evidence of low misclassification for a crowded group of pipes with a condition rating of 1 and the model's high ability to capture the minority group of pipes with a condition rating of 5, respectively. The overall F1-score of 0.70 shows an acceptable performance of this model, finally.

Based on a combined historical inspection dataset gathered from the cities of Tampa and Dallas, this study developed seven different models to predict the condition level of sewer pipes. To establish the prediction models, nine independent variables were included: pipe age, material, diameter, length, depth, slope, soil type, soil pH, and pipe location. Sewer pipe condition ratings were the target variable and were evaluated using the PACP method. The intention was to develop a model to anticipate each pipe's five condition ratings.

Utilizing the confusion matrix, ROC curve, and macro-average F1-score as three different validation techniques, all the models were tested for accuracy. The effectiveness of the models implemented in this investigation is shown in Figure 5. As seen, the Random Forest model had the best accuracy, with an F1-score of 80%, whereas Multinomial Logistic Regression had the lowest accuracy (F1 = 48%). It can be seen that tree-based models had better performance than others. However, the Bagging approach was more efficient than the Boosting approach. Furthermore, determining the significant variables is a critical component of condition prediction modeling. These factors have a significant impact on sewer pipe conditions. Therefore, leaving them out of the model could reduce its accuracy. One advantage of tree-based models is the ability to prioritize the significance of the independent variables for both regression and classification goals. Generally, feature importance gives a score indicating how helpful a variable is in putting the model into practice. Figure 6 displays the relative importance of various factors obtained from the feature importance attribute of this study's best model (Random Forest). This figure demonstrates that the factors of age and length of pipes had the highest effect on their condition. The material was the next most important parameter. Pipe size, surrounding soil type, pH of surrounding soil, depth of buried pipe, and pipe slope had a lower impact, respectively. The pipe location (road type) had the least effect on the condition of sewers.

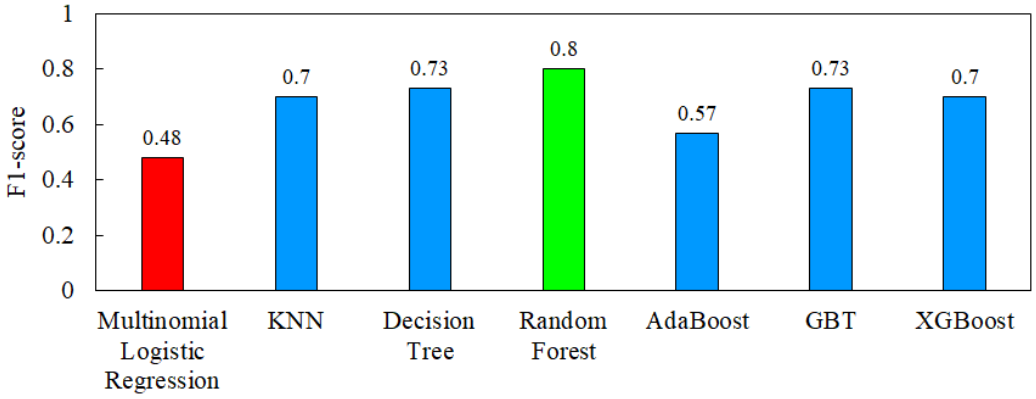

Figure 5- Comparison of Model Performances

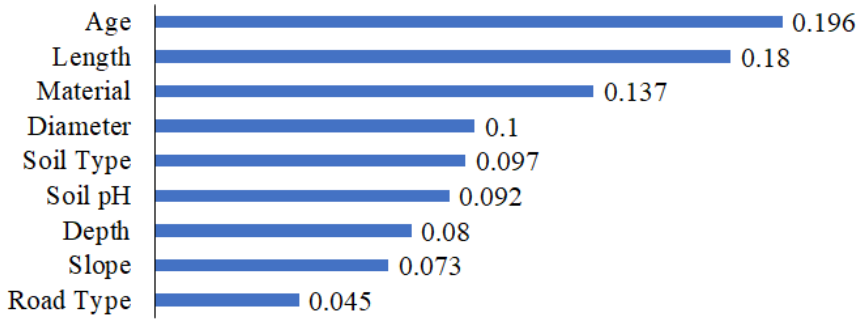

Figure 6- Relative Importance of Independent Variables

4- CONCLUSIONS

Sewer pipe network datasets from two cities, Dallas, TX, and Tampa, FL, were combined in this study. The reasons for the combination were to increase the number of data points to enhance the model's accuracy, to

develop a more comprehensive model than previous studies, and to avoid the model's overfitting. The overall macro-average F1-score, as a summary indicator of the model's performance, was used to compare the developed models. The Random Forest model had the best accuracy in predicting the condition rating of pipes in all five conditions, with an F1-score of 0.8. In contrast, Multinomial Logistic Regression had the lowest accuracy, with an F1-score of 0.48. It could be seen that tree-based models had better performance than others. However, the Bagging approach was more efficient than the Boosting approach. Additionally, the results of the most accurate model developed in this research (Random Forest) demonstrated that the factors of age, length of pipes, and pipe material had the highest effect on the condition ratings of sewers, which was consistent with the results of the Binary Logistic Regression model, except for the surrounding soil type. On the other hand, pipe location (road type) had the least effect.

5- REFERENCES

1. Najafi, M., Gokhale, S., Calderón, D.R., & Ma, B. (2021). Trenchless technology: Pipeline and utility design, construction, and renewal. Second Edition, McGraw-Hill Education.
2. EPA, (2002), Asset Management for Sewer Collection Systems. United States Environmental Protection Agency, Office of Wastewater Management.
3. Opila, M.C. (2011). Structural condition scoring of buried sewer pipes for risk-based decision making. Ph.D. Thesis, University of Delaware, Newark, DE, USA.
4. Davies, J., Clarke, B., Whiter, J., & Cunningham, R. (2001). Factors influencing the structural deterioration and collapse of rigid sewer pipes. Urban Water, 3(1-2), 73-89.
5. Malek Mohammadi, M., Najafi, M., Kermanshachi, S., Kaushal, V., & Serajiantehrani, R. (2020). Factors Influencing the Condition of Sewer Pipes: State-of-the-Art Review. Journal of Pipeline Systems Engineering and Practice, 11(4), 03120002.
6. Luger, G. F. (2009). Artificial Intelligence. Structures and Strategies for Complex Problem Solving. 6th Edition, Addison Wesley, Boston MA.
7. Bakry, I., Alzraiee, H., Kaddoura, K., El Masry, M., & Zayed, T. (2016). Condition Prediction for Chemical Grouting Rehabilitation of Sewer Networks. Journal of Performance of Constructed Facilities, 30(6), 04016042.
8. Bakry, I., Alzraiee, H., Masry, M. E., Kaddoura, K., & Zayed, T. (2016). Condition Prediction for Cured-in-Place Pipe Rehabilitation of Sewer Mains. Journal of Performance of Constructed Facilities, 30(5), 04016016.
9. Tran, D., Ng, A., & Perera, B. (2007). Neural networks deterioration models for serviceability condition of buried stormwater pipes. Engineering Applications of Artificial Intelligence, 20(8), 1144-1151.
10. Abdel Moteleb, M. (2010). Risk Based Decision Making Tools for Sewer Infrastructure Management Ph.D. Thesis, University of Cincinnati, Cincinnati, OH, USA.
11. Ana, E., Bauwens, W., Pessemier, M., Thoeye, C., Smolders, S., Boonen, I., & De Gueldre, G. (2009). An investigation of the factors influencing sewer structural deterioration. Urban Water Journal, 6(4), 303-312.
12. Malek Mohammadi, M. (2019). Development of Condition Prediction Models for Sanitary Sewer Pipes. Ph.D. Thesis, University of Texas at Arlington, Arlington, TX, USA.
13. Najafi, M., & Kulandaivel, G. (2005). Pipeline condition prediction using neural network models. In Pipelines 2005: Optimizing Pipeline Design, Operations, and Maintenance in Today's Economy, 767-781.
14. 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.
15. Harvey, R.R., & McBean, E.A. (2014). Comparing the utility of decision trees and support vector machines when planning inspections of linear sewer infrastructure. Journal of Hydroinformatics, 16(6), 1265–1279.
16. Laakso, T., Kokkonen, T., Mellin, I., & Vahala, R. (2018). Sewer Condition Prediction and Analysis of Explanatory Factors. Water, 10(9), 1239.
17. National Association of Sewer Service Companies (NASSCO). (2018). Pipeline Assessment Certification Manual. Marriottsville, MD, USA.
18. Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists. O'Reilly Media, Inc.: Newton, MA, USA.
19. Géron, A. (2017). Hands-on machine learning with scikit-learn and tensorflow: Concepts. Tools, and Techniques to build intelligent systems. First Edition, O'Reilly Media, Inc.