

Article

Comprehensive Feature Analysis for Sewer Deterioration Modeling

Bolette D. Hansen ^{1,2,*} , Søren H. Rasmussen ¹, Mads Uggerby ¹, Thomas B. Moeslund ² and David G. Jensen ¹¹ EnviDan A/S, Climate Adaption & Drainage Systems Department, 8600 Silkeborg, Denmark; shr@envidan.dk (S.H.R.); mau@envidan.dk (M.U.); dgj@envidan.dk (D.G.J.)² Visual Analysis and Perception Lab, Aalborg University, 9000 Aalborg, Denmark; tbm@create.aau.dk

* Correspondence: bdha@create.aau.dk

Abstract: Timely maintenance of sewers is essential to preventing reduced functionality and break-down of the systems. Due to the high costs associated with inspecting a sewer system, substantial research has focused on sewer deterioration modeling and identification of the most useful features. However, there is a lack of consensus in the findings. This study investigates how the feature importance depends on the definition of bad pipes and how the feature importance changes between utilities with similar data bases. A dataset containing 318,457 pipes from 35 utilities with a condition state (CS) ranging from one to four was used. The dataset was cleaned, and a backward step analysis (BSA) was applied to two ways of binarizing the CS. Additionally, a BSA was applied for each utility with ≥ 100 pipes in CS four. The results showed that a selective definition of bad pipes reduced the performance and changed the order of which features contributed the most. In each case, either year of construction, age, groundwater, year of rehabilitation, or dimension was the most important feature. On average 6.5 features contributed to the utility-specific models. The feature analysis was sensitive to the inspection strategy, the size of the dataset, and interdependency between the features.



Citation: Hansen, B.D.; Rasmussen, S.H.; Uggerby, M.; Moeslund, T.B.; Jensen, D.G. Comprehensive Feature Analysis for Sewer Deterioration Modeling. *Water* **2021**, *13*, 819. <https://doi.org/10.3390/w13060819>

Academic Editor: Armando Di Nardo

Received: 10 February 2021

Accepted: 14 March 2021

Published: 17 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: sewer; feature analysis; parameter importance; ageing; deterioration model; inspection strategy

1. Introduction

A sewer system is a hidden but very expensive type of infrastructure to maintain [1]. Breakdown of a sewer can result in significant damage to roads and buildings. Furthermore, reduced functionality of the sewers can lead to flooding and exfiltration, for example, which can affect a number of externalities, such as property, traffic disruption, public health and the environment [1,2]. For these reasons, the sewers' operators need to replace them in a timely manner, especially if the sewers are critical. However, sewers' underground location makes them difficult to monitor. Today, monitoring of the sewers is typically done by Closed Circuit Television inspection (CCTV inspection) [1]. CCTV inspection is done by manually sending a TV-inspection robot into the sewer and annotating all observations. As this is very time consuming, expensive, and imprecise due to a number of subjective factors [3,4], much research has been put into automating these processes [1,5]. However, full automatization of sewer inspection is not imminent. The high costs associated with CCTV-inspection forces utilities to prioritize which sewers to inspect. In Denmark, the paradigm for risk-based rehabilitation has been based on area. The areas which should be subject to CCTV-inspection were prioritized based on age of the pipes and the experience of the operators. Based on the findings in the CCTV-inspections, it was chosen whether an area should be rehabilitated or not. This has resulted in rehabilitation of pipes which could have been operational for several years as the inspection showed that the pipe might not be operational for the whole period until the next time the area would be chosen for rehabilitation. Today, for economical optimization and better use of the pipes lift time,

there is a trend toward risk based CCTV-inspection planning and rehabilitation on a pipe level.

Maintenance of sewer systems on pipe level entails new requirements for computer systems to keep track of the individual pipes, as the utilities now need to keep track of several tens of thousands of pipes instead of a limited number of areas. To assist the utilities in choosing which sewers to inspect, several decision support systems have been developed [6–9]. Usually these systems are risk models, consisting of a deterioration model and a consequence model. The deterioration models predict the condition of the sewers or the likelihood of a sewer's condition. The consequence models describe the severity of a potential sewer failure and can include economic, environmental, and social consequences [1]. Generally, the deterioration models suffer from low accuracy.

Development of sewer deterioration models is complicated by a high uncertainty in the data. This uncertainty is influenced, among other things, by subjectivity in the annotation of CCTV inspections, lack of data, and subjective selection of which pipes to inspect [3]. Dirksen et al. [4] found that defects with distinct features like roots were easy to find, while the probability of getting a false negative for other defect types varied around 0.25. The probability of a false positive was found to be around 0.04 [4]. Another issue often affecting the deterioration datasets is a lack of information [1], which results in low quality data. Furthermore, the datasets are affected by the fact that they have typically been collected for a specific purpose, such as quality assurance before asset handover or road renovation, diagnosis of malfunctioning and random inspections. This introduces a selective survival bias in the data [1]. Other factors that complicate deterioration modelling are that the datasets in general are highly skewed, both according to the number of pipes in the different classes and according to the predictor variables [10–12]. Furthermore, the size of the natural variability between sewers is unknown.

A large number of deterioration models have been developed; however, a lack of publicly available datasets due to privacy issues makes it difficult to compare the models [5]. Furthermore, the condition state (CS) is typically based on the local standard for CCTV inspection, which can be based on, for example, the European standard [13], Pipeline Assessment Certification Program [14], or a country specific standard [12,15–17]. Moreover, in order to evaluate the deterioration models many authors tend to classify the multiclass or regression problem as a binary problem [12,13,18–20]. However, the model performance is very sensitive to how picky the evaluation is designed to be. For example, the performance of the precision and recall will increase if considering both pipes in the worst CS and the second worst CS as bad pipes, compared to considering only pipes in the worst CS to be in bad condition.

In addition, when deciding how to define the target variable the developer of the deterioration model needs to decide which predictor variables to use. Several methods have previously been used for parameter selection and the feature importance test. O'Reilly [21] in 1989 investigated the correlation between defects and individual parameters such as age, material, diameter, location, depth, wastewater type, soil type etc. in 180 km of sewers. Hansen et al. [11] investigated the potential benefits of developing deterioration models based on data groups defined by experts but found no improvement in model performance. Yin et al. [22] used a backward variable elimination process, through which they removed a parameter at a time and examined how the performance changed. Davies et al. [23] used a backward selection method and Laakso et al. [13] used the Boruta algorithm and found eight features to be influential.

Carvalho et al. [10] used eight different methods to investigate the feature importance and found that the different methods showed very different results. For example, if analyzing the features by removing the most significant features step by step, the importance of the other features will change, as there is often redundancy in the signal from the different predictor variables. This is not encountered when using the build-in feature analysis in Random Forest [10], however. Due to the uncertainty in the data, Roghani et al. [3] found that using the two or three most informative predictor variables was sufficient to build the

deterioration model. However, using a deterioration model was better than just basing it on the inspection age.

Mohammadi et al. [24] reviewed 24 statistical and AI based papers on sewer deterioration. Nineteen of the reviewed papers provided information on whether a parameter was relevant. Nineteen features were considered, and none of the features were used in all the papers. Furthermore, none of the features considered relevant in more than three of the papers were considered relevant in all the papers. This illustrates a high variability in feature importance. Likewise, none of the features whose significance level was specified in more than one case were irrelevant in all the studies they were used in [24]. Finding the most significant features is important as accessing, extracting, and preprocessing each feature is very time demanding. In a review of deterioration models Hawari et al. [25] concluded that more work needs to be done to identify which data municipalities should collect in order to develop reliable deterioration models [25].

As described above, the performance of the deterioration models is affected by many conditions and a number of choices needs to be made for each model. This makes it possible to develop well performing models within academia. However, to create value, the models must meet the utilities' needs. For example, Guzmán-Fierro et al. [26] worked with a target variable ranging from 1 to 5 but developed a model that encountered only the pipes in CS 1 and CS 5. In reality, it is not possible to leave out the pipes in between, at least during the preliminary inspection.

In summary, sewer deterioration modeling has been a hot topic for the last two decades and myriad factors influence the performance of the models. Finding the optimal model cannot necessarily be done by selecting the model with the highest performance according to the literature. Likewise, there is a great deal of disagreement about which predictor variables are significant. The existing sewer deterioration models presented in the literature are characterized by large deviations in data, methodology, etc. Today researchers tend to perform feature analyses on single datasets. However, a rarely touched perspective is the statistical variation in the features influencing the results when using similar datasets.

The contributions of this study are investigations of:

- The overall feature importance in a dataset containing information from several different utilities, including identification of potential drawbacks
- How the performance and feature importance of the models are affected by how the model developer has distinguished between good and bad pipes
- How the feature importance varies between utilities when the parameters in the datasets have been found in the same way for all utilities.

To the best of the authors' knowledge, this study provides the most comprehensive analysis of feature importance in sewer deterioration modeling and the first investigation of feature importance across several utilities with similar data bases. This information adds value to the process of developing deterioration models for utilities, which have a limited budget.

The following section of the paper, Section 2, provides a description of the data available, preprocessing, model selection, and the method used for feature importance. Section 3 contains three subsections, one for each of the contributions, while Section 4 contains a discussion of the key findings and comparisons to the literature. Section 5 contains a summary of the most important conclusions covered by the paper.

2. Materials and Methods

A dataset containing pipes from 35 utilities across Denmark was extracted from a common database for CCTV inspections. Pipes with suspicious values were not extracted. Examples of suspicious data points included those in which the following criteria were not met: $63 \text{ mm} < \text{dimension} < 3000 \text{ mm}$, $0 \text{ years} < \text{age} < 169 \text{ years}$ and $0.6 \text{ m} < \text{depth} < 10 \text{ m}$. Most of the inspections were performed from the start of the 1990s until today. The full dataset contains CCTV inspection from 318,457 pipes. For each pipe access to 24 different

predictor variables was attempted; however, all predictor variables were only available for 196,174 pipes. An overview of the predictor variables can be seen in Table 1.

Table 1. Overview of predictor variables and the corresponding data types as well as distribution or units. For continuous and numeric data types, the mean and std are presented. For categorical data types, the percentage of pipes in each category is presented and for binary data types, the percentage of true values is presented.

Predictor Variable (Abbreviation)	Data Type	Distribution and Units
Length	Continuous	43.59 ± 26.54 m
Age	Numeric	25.7 ± 21.1 Years
Material	Categorical	Concrete (61.22%), plastic (33.77%), clay (1.50%), full reline (2.23%), other (1.22%)
Dimension	Continuous	306.6 ± 198.9 mm
Wastewater type (Wastewater)	Categorical	Sewage (38.18%), rain (31.66%), combined (29.54%)
Slope	Continuous	12.22 ± 11.61 mm/m
Year of construction (YoConst)	Numeric	Year 1982.2 ± 21.1
Year of rehabilitation (YoRehab)	Numeric	Year 1983.7 ± 21.6. This is set to YoConst if not rehabilitated
Type of rehabilitation (Rehab)	Categorical	Total replacement (5.11%), Full reline, also included as material (2.23%), Punctuate (0.04%), unknown (0.00%)
X coordinate (X)	Continuous	Adjusted UTM (m)
Y coordinate (Y)	Continuous	Adjusted UTM (m)
Utility ID	Numeric	
Ground level	Continues	32.07 ± 24.57 m
Depth	Continuous	2.43 ± 0.89 m
Groundwater level according to pipe (Groundwater)	Continuous	−4.48 ± 3.88 m
Soil type	Categorical	ML ¹ (44.15%), MS ² (19.38%), OPS ³ (9.79%), FDS ⁴ (6.53%), MaS ⁵ (4.28%), MoS ⁶ (4.15%), OMS ⁷ (3.45%), FS ⁸ (2.77%), MC ⁹ (1.00%), MG ¹⁰ (0.46%), Marsk (0.25%), Lake (0.004%)
Road type	Categorical	Tertiary (38.99%), secondary (13.54%), primary (13.00%), traffic (3.99%), other (1.88%), no road (28.60%)
Distance to road center (DistRoad)	Continuous	1.91 ± 1.93 m for pipes less than 10 m from road center. The remaining pipes have been assigned the value 99 m
Distance to nearest trees (Trees)	Categorical	<4 (6.51%), <12 (25.50%), >12 (74.50%)
Number of road grate (NoGrates)	Numeric	3.46 ± 3.56 grates
City type	Categorical	City zone (79% incl. city center and industrial area) city center (18.43%), industrial area (15.22%)
Number of buildings (NoBuildings)	Numeric	6.09 ± 5.22 buildings
Area with tall buildings (BuildingHigh)	Binary	9.44% True
Area with low buildings (BuildingLow)	Binary	77.06% True

¹ Morain clay, ² Meltwater sand, ³ Outwash plain sand, ⁴ Freshwater deposit of sand, ⁵ Marine sand, ⁶ Morain sand, ⁷ Old marine sand, ⁸ Fly sand, ⁹ Meltwater clay, ¹⁰ Marine gravel.

All CCTV-inspections followed the Danish standard for CCTV inspections [27]. The inspections contained information on several observation types and corresponding severity of each. Based on the type of defect and its severity, the observations were categorized as CS 1–4. The way in which each observation should contribute to the CS was based on input from a Danish utility. The CS of a given pipe was then set to the worst of the observations.

An overview of how the different defect types and severities contribute to the CS can be seen in Table 2.

Table 2. Overview of how a CCTV-observation of a defect with severity zero to four entails the condition state to be in state one to four.

	Settled Deposits	Attached Deposits	Deformation	Obstacle	Displaced Joint	Connection	Infiltration	Intruding Sealing Material	Surface Damage	Transitional Component	CNTP ¹	CNCU ²	CNDR ³	Manufacturing Defect	CNCH ⁴	Break and Collapse	Roots	Water Level
Severity 0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Severity 1	1	2	1	2	1	2	2	1	1	2	1	1	1	1	1	1	1	1
Severity 2	2	3	2	3	2	3	3	2	2	2	2	2	2	2	2	2	2	2
Severity 3	3	4	3	3	3	3	4	3	3	3	3	3	3	3	3	3	3	3
Severity 4	4	4	3	4	4	3	4	4	4	3	4	4	4	4	4	4	4	3

¹ Connection with lining defection or intruding connection, ² Connection through cut hole in reline pipe, ³ Connection through drill hole in pipe, ⁴ Connection through chop hole in pipe. The color indicates the CS and goes from green to red.

2.1. Preprocessing of Data

Thirty-five datasets were included in this study: one containing data from all the pipes and one for each of the utilities that had more than 100 bad pipes.

The preprocessing of the datasets was done by first removing features represented in less than 20% of the cases and then removing data points containing NaNs. An overview of the number of pipes available before and after data cleaning, the number of features removed from the dataset, and the number of pipes in bad condition can be seen in Table 3.

Table 3. Overview of the datasets containing more than 100 bad pipes before and after cleaning, as well as the number of features removed in the cleaning process and the number of pipes in condition state (CS) three and four after cleaning.

Dataset	Number of Pipes in Total	Number of Pipes after Cleaning	Number of Features Removed	No. Pipes in CS 3 after Cleaning	No. of Pipes in CS 4 after Cleaning
All pipes	318,457	196,174	0	64,969 (33%)	20,542 (10%)
Utility 1	20,379	17,730	0	5992 (34%)	1758 (10%)
Utility 2	10,062	8270	0	3138 (38%)	872 (11%)
Utility 3	9904	8469	1	3666 (43%)	1103 (13%)
Utility 4	10,116	7913	0	2166 (27%)	615 (8%)
Utility 5	18,745	15,830	0	4477 (28%)	1162 (7%)
Utility 6	17,109	13,867	0	5290 (38%)	1141 (8%)
Utility 7	4669	3108	6	687 (22%)	280 (9%)
Utility 8	4522	3315	0	892 (27%)	790 (24%)
Utility 9	12,163	9451	1	2759 (29%)	1708 (18%)
Utility 10	734	640	1	135 (21%)	103 (16%)
Utility 11	6355	5686	1	1899 (33%)	469 (8%)
Utility 12	20,945	16,453	0	5128 (31%)	987 (6%)
Utility 13	1779	1268	0	312 (25%)	120 (9%)
Utility 14	18,154	14,587	0	6214 (43%)	2063 (14%)
Utility 15	2757	2027	0	580 (29%)	231 (11%)
Utility 16	18,025	15,812	4	4700 (30%)	1560 (10%)
Utility 17	3855	3252	0	1145 (35%)	311 (10%)
Utility 18	9655	7128	0	2516 (35%)	880 (12%)

Table 3. *Cont.*

Dataset	Number of Pipes in Total	Number of Pipes after Cleaning	Number of Features Removed	No. Pipes in CS 3 after Cleaning	No. of Pipes in CS 4 after Cleaning
Utility 19	9253	7006	0	2092 (30%)	806 (12%)
Utility 20	10,520	8370	0	1870 (22%)	893 (11%)
Utility 21	7959	6914	1	1868 (27%)	668 (10%)
Utility 22	4458	4040	7	1685 (42%)	381 (9%)
Utility 23	2974	2427	0	1048 (43%)	492 (20%)
Utility 24	14,300	11,942	1	5953 (50%)	1720 (14%)
Utility 25	18,879	13,978	0	2348 (17%)	1360 (10%)
Utility 26	3864	2812	0	611 (22%)	274 (10%)
Utility 27	7171	6033	1	1956 (32%)	990 (16%)
Utility 28	16,672	14,064	1	4442 (32%)	1673 (12%)
Utility 29	10,939	9428	0	4068 (43%)	761 (8%)
Utility 30	5750	4540	6	1940 (43%)	431 (9%)
Utility 31	4213	3944	1	1765 (45%)	560 (14%)
Utility 32	4877	4073	0	1451 (36%)	253 (6%)
Utility 33	6164	5661	8	1092 (19%)	251 (4%)

All datasets were randomly split with 90% for training and 10% for testing. Due to the high imbalance between good and bad pipes, the training sets were randomly downsampled to contain an equal number of good and bad pipes.

2.2. Model Selection

As shown in Table 1, the predictor variables available for this study have different data types, which is well handled by forest based models. Forest based models can be used to solve either regression or classification problems. They consist of several decision trees, which evaluate the data points according to a treelike structure. The construction of each decision tree is based on statistical variations in the datasets and an introduced randomness. Each decision tree votes for a specific outcome and based on these votes the forest makes a prediction. Two forest based model types were considered for classification: XGBoost [28] and Random Forest [29].

The Random Forest model was implemented using the Python library scikit-learn [30]. The number of decision trees was set to 177 and the max depth was set to 26 based on Hansen et al. [12]. The remaining hyperparameters were set to the default value.

The number of estimators and max depth for the XGBoost model was first defined with inspiration from the settings of the random forest model. Hereafter different ways of setting these parameters were tested. For classification multiclass softmax was used. For the remaining parameters, the default values were used.

XGBoost benefitted from the ability to handle missing data; however, a XGBoost model takes much more time to train than a Random Forest model. XGBoost did not show better results than Random Forest. Furthermore, Random Forest is often used for deterioration modeling in sewer [12,13,19,31] and in water pipes [32]. For this reason, Random Forest was used for this study.

2.3. Feature Importance

In order to make the feature analysis, three methods were considered: (1) The Random Forest built-in feature importance measure [30], (2) Clustering the features in groups of features, training all combinations of the feature clusters, and investigating which feature clusters are most present in the best models and which feature clusters are most present in the bad models and (3) Making a backward step analysis by training a model on all but one feature for all features represented and removing the least contributing feature. This should then be repeated until only one feature is left.

Before selecting which method to use, it is worth considering the redundancy of the features. This has been handled previously by ensuring high heterogeneity between the features [15]. This approach entails removing a large number of features, which might be similar in most cases but could vary in essential cases. An example of this is year of construction and year of rehabilitation. If the pipe has not been rehabilitated, the year of rehabilitation is equal to year of construction, inducing a high redundancy between the two features. However, as rehabilitation is directly related to the condition of the pipe, the feature should be included in the analysis. Moreover, by including all the predictor variables in the analysis it is possible to account for the variations between utilities and obtain knowledge about features otherwise removed from the dataset.

As the built-in Random Forest method calculates the feature importance by number of splits for each feature, it is sensitive to redundancy between features. Clustering the features and training a model for all combinations of the feature clusters was tested initially, but it showed a high variance between the different utilities and did not contribute information on the individual features. The benefit of using the step analysis is that it encounters all the features; however, in the cases where many features are irrelevant it will be random if a feature is the 10th or the 20th least contributing feature. Like the built-in Random Forest method, this approach is sensitive to redundancy in the features, but the influence is of a more transparent character. Based on the above, the decision was made to conduct a backward step analysis.

Backward Step Analysis

To conduct the backward feature step analysis, the dataset was split randomly, and a model was trained for each of the features that was left out. This was repeated 10 times, and the predictor variable, which on average contributed the least to the performance, was removed. This was repeated until only one feature was left. Furthermore, the average performance of a model trained on all the features 10 times was found. For the feature analysis it was necessary to get a single performance measure. For this reason, the performance was calculated as the f1-score which is a balanced evaluation of the precision and recall. The f1-score was calculated using the Python library Scikit-learn [30] and the formula for calculating the f1-score can be seen in Equation (1).

$$f1_{score} = 2 \cdot (\text{precision} \cdot \text{recall}) / (\text{precision} + \text{recall}) \quad (1)$$

A challenge using the f1-score is that it only encounters precision and recall. Thereby it does not encounter that the test set has a skewed distribution. For example, by randomly selecting 50% of the pipes a higher f1-score will be obtained than by selecting a number of pipes corresponding to the number of bad pipes in the dataset. Therefore, to evaluate how well the models performed according to a random selection strategy, the performance was calculated when randomly selecting 50% of the pipes and when randomly selecting a number of pipes corresponding to the number of bad pipes in the dataset. Due to variations in the distribution of bad pipes in the datasets, the F1-score cannot be used to give a fair evaluation of performance between utilities.

An overview of the method used for making the backward step analysis, calculating the performance when using all features were encountered, calculating the performance when randomly selecting 50% of the pipes, and calculating the performance when randomly selecting the same number of bad pipes as present in the dataset can be seen in Figure 1.

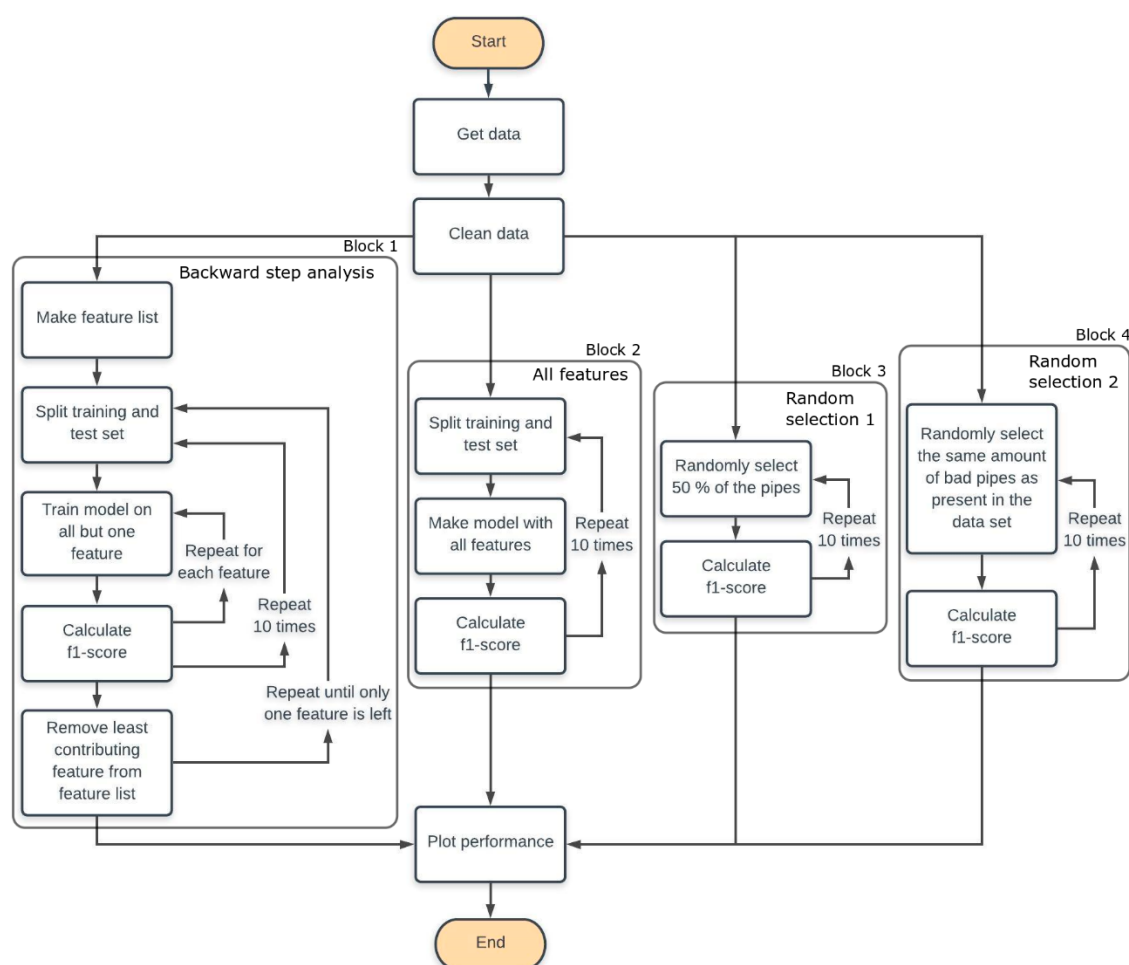


Figure 1. Feature step analysis when removing the features with the smallest contribution one by one. Block 1 shows the backward step analysis, block 2 shows the method for calculating the performance using all features, block 3 shows the method for calculating the performance when randomly selecting 50% of the pipes, and block 4 shows the method for calculating the performance when randomly selecting the same number of bad pipes as present in the dataset.

2.4. Experiments

Three experiments were carried out. The purpose of the first experiment was to identify potential drawbacks of the approach used and take these into account in the remaining experiments. This experiment is referred to as the baseline. The purpose of the second experiment was to investigate how the performance and the feature importance changed when changing the definition of the target variable. The purpose of the last experiment was to investigate how the feature analysis changed between different utilities.

2.4.1. Baseline

This experiment was carried out on the full dataset. The condition of pipes in CS one and two was considered good while that of the pipes in CS three and four was considered bad. In this experiment, the backward step analysis was run for all features and relevant adjustments were incorporated.

2.4.2. Target Variable

In this experiment two backward step analyses were made: in the first analysis both pipes in CS three and four were considered bad pipes. In the second analysis only pipes in CS four were considered bad pipes. To ensure a fair comparison between the two analyses,

the amount of training data in the first analysis was downsampled to the amount of training data available for the second analysis.

2.4.3. Difference between Utilities

A backward step analysis was performed for each of the utilities. This experiment was initially conducted solely considering pipes in CS 4 as being in bad condition; however, there was a relatively high variance in the features found relevant at the different utilities. This was particularly evident for utilities with few bad pipes entailing smaller datasets. For this reason, both pipes in CS three and four were considered bad pipes.

Each of the analyses was manually inspected to determine which parameters were significant for each utility. This would preferably have been an automatic process, but as the results did not show a smoothly decreasing curve in all cases, an automatic approach would have required several assumptions.

An overview of the significant features for the different utilities was made, and the performance of the models was compared to the size of the dataset and the number of significant features.

3. Results

3.1. Baseline

The results of the baseline step analysis can be seen in Figure 2a. The f1-score of the model using all the features is 0.75.

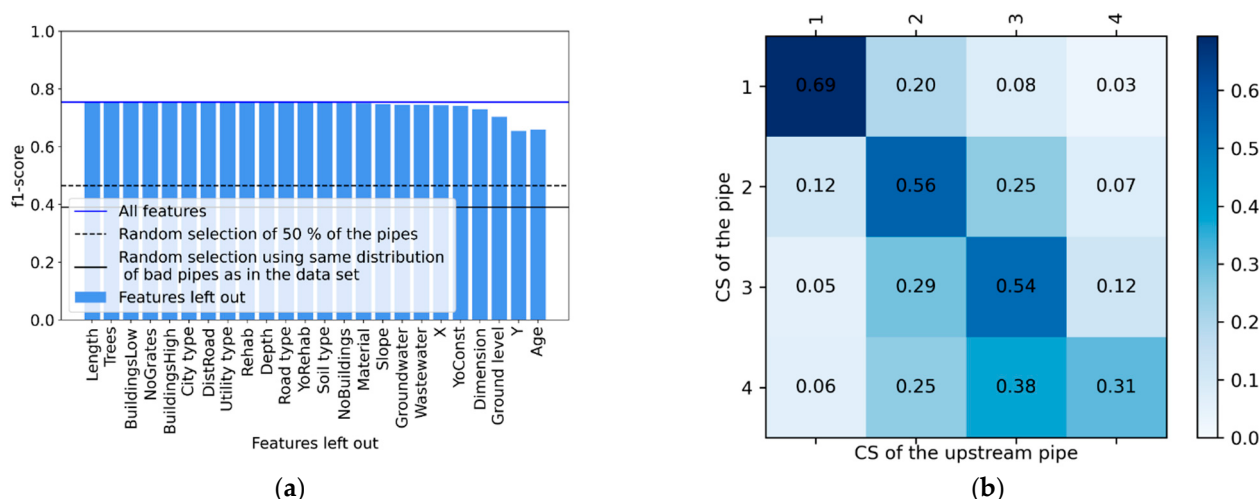


Figure 2. (a) Backward feature step analysis for the baseline. Each chart shows the performance before the least contributing feature remaining has been removed, starting from the left to the right. The dashed and solid lines shows the performance if utilizing a random selection of respectively 50% of the pipes and a percentage corresponding to the distribution of bad pipes in the dataset. (b) Normalized confusion matrix for a pipe and the upstream connected pipe.

From Figure 2a, it can be seen that both Y and X coordinates contribute to the predictions. This could indicate that when these parameters are included, the model learns the position of the pipes rather than that the actual parameters influence the condition state. In other words, this would correspond to using a nearest neighbor approach, which is problematic if applying the method to areas where training data is not available.

To clarify this suspicion, the probability of an upstream pipe present in a certain CS and given the pipe's condition was investigated. The normalized confusion matrix for this can be seen in Figure 2b, and it shows a clear correlation between the CSs of adjacent frames. Calculating the f1-score for pipes in CS three and four gives a f1-score of 0.69. As can be seen in the figure, the confusion matrix is not symmetric, which might be due to systematically occurring changes in the sewers. For example, it is common for an upstream

pipe to be smaller than the downstream pipe but rarely the other way around. It should be noted that the figure is made from the same dataset as used for the feature analysis; however, not all pipes in the dataset have an inspected upstream pipe while others have more than one inspected upstream pipe.

Sewer inspections are not usually performed by taking representative samples from the whole sewer system but rather in subjectively selected areas. Therefore, the performance might be lower when applied to a part of the network that has not previously been inspected. To clarify this, another backward step analysis was applied but, instead of using a random split between training and test data, all pipes from four randomly selected utilities were used for testing and the remaining pipes for training. In so doing, the performance of the model based on all the predictor variable dropped by 10%. Furthermore, when performing the feature analysis, the utility ID and the X and Y coordinates were among the four worst predictor variables. For this reason, features related to location were not included in the remaining experiments. The new baseline can be seen in Figure 3.

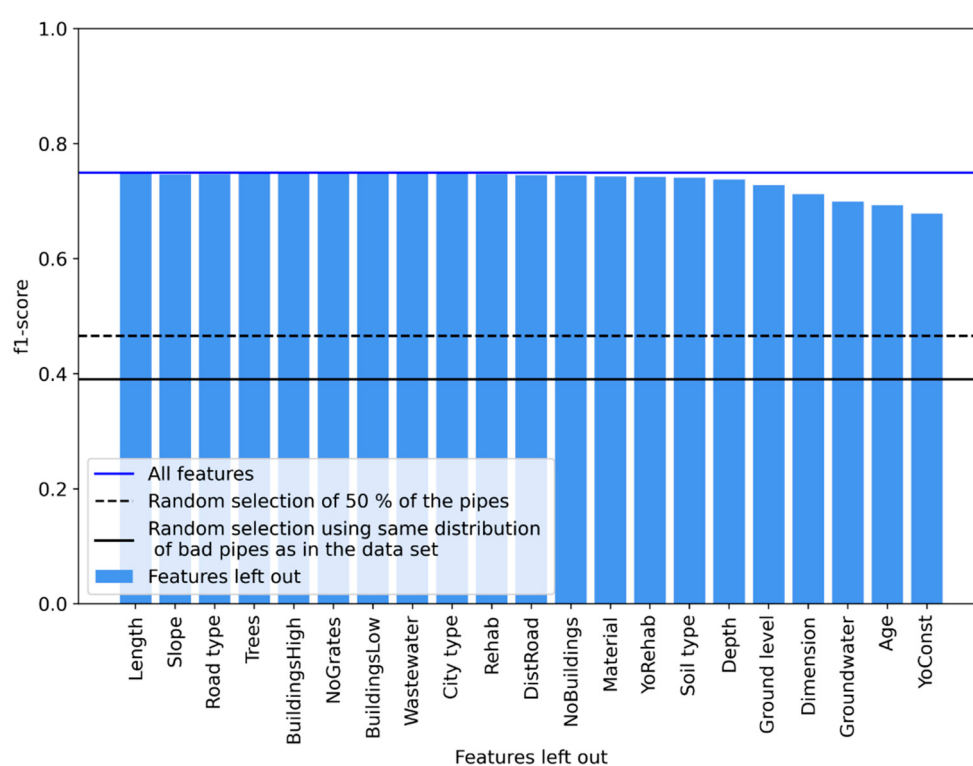


Figure 3. Backward feature step analysis for the new baseline where the geographical information has been removed. Each chart shows the performance when the least contributing feature remaining has been removed starting from the left and progressing to the right. The dashed black line shows the performance when randomly classifying 50% of the pipes as bad pipes and the solid black line shows the performance when randomly classifying a percentage of pipes, corresponding to the number of bad pipes in the dataset, as bad.

3.2. Target Variable

Figure 4a shows the feature step analysis when considering pipes in both CS three and four to be in bad condition when using the same amount of training data as when considering only pipes in class four as being in bad condition. This model obtains a f1-score of 0.73 when using all features. Figure 4b shows the feature step analysis when only considering pipes in CS 4 to be in bad condition. This model obtains a f1-score of 0.35.

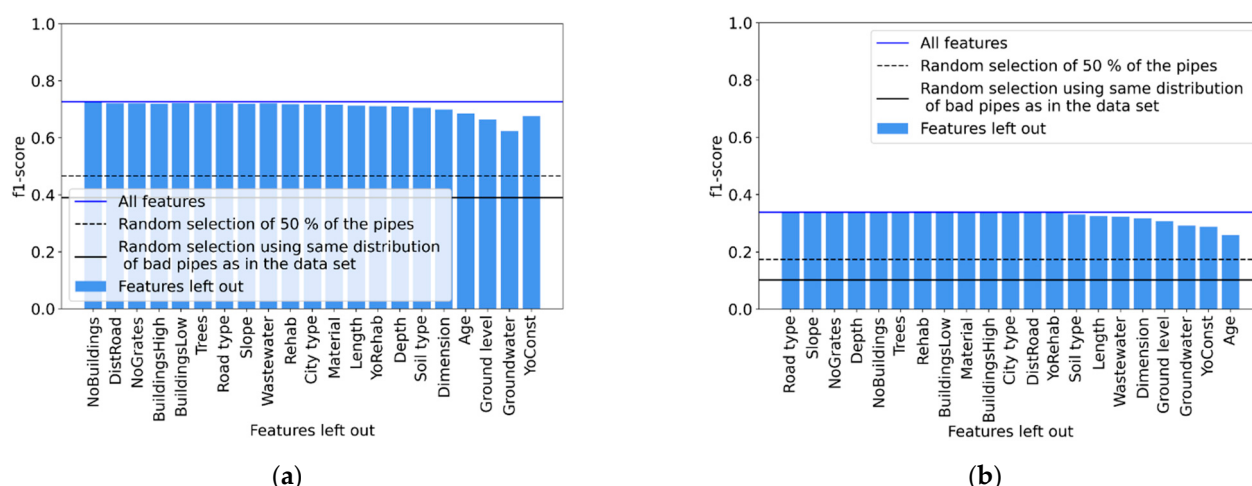


Figure 4. (a) Same as baseline but with the same amount of training data as available when solely considering pipes in CS 4 as being in bad condition. (b) Feature step analysis when solely considering pipes in CS 4 as bad.

Figure 4a shows that the year of construction alone performs better than when combined with the relative groundwater level and ground level. This indicates that the relative groundwater level and ground level contributed positively to a group of features but introduced noise when included individually.

A smaller number of features are found to contribute when solely considering pipes in CS 4 as being in bad condition than when pipes in CS 3 also are considered as being in bad condition. All the parameters that contribute to the model performance in the first case mentioned, aside from wastewater type, also contribute in the second case.

3.3. Difference between Utilities

For each of the utilities in Table 1 a backward step analysis was performed and manually inspected. Some of the utilities were observed to perform better when removing features up to a certain point. This was clearest for utility 10, which is the utility with the smallest number of bad pipes, but the phenomenon could also be observed in some of the other utilities. The feature analysis for utility 10 can be seen in Figure 5a. In most cases, the feature analysis shows a decrease in performance when features are removed. However, for some utilities the performance increases when the second-to-last feature is removed. This most often occurs if the second-to-last feature remaining is ground level or depth, but it has also been observed for groundwater to a smaller extent. An example of this can be seen in Figure 5b, which shows the feature analysis for the utility with the largest number of bad pipes.

An overview of the predictor variables considered relevant for the different utilities can be seen in Table 3. In the table the performance is given as the f1-score when using the optimal number of features. The table also shows how many times a feature is found to be the most important feature.

On average 6.5 features were found to contribute to the performance. The table shows that year of rehabilitation and year of construction contain redundant information, and at least one of them is found to be significant in 24 of the utilities. For this reason, year of construction is considered more relevant than shown in the table if year of rehabilitation is not available and vice versa. Likewise, there might be some redundancy in number of buildings, buildings low, buildings high, and number of grids.

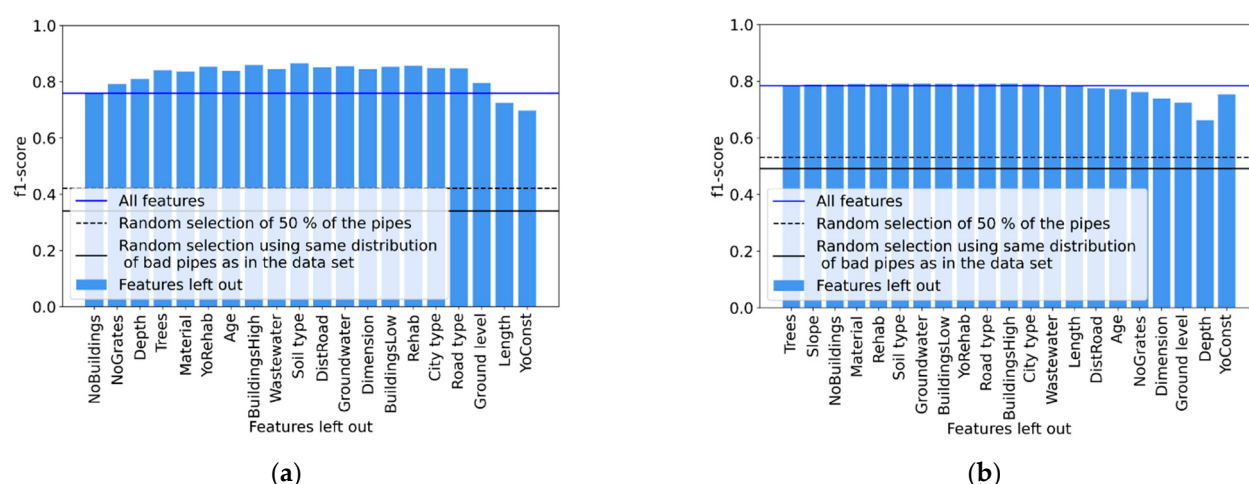


Figure 5. Feature analysis for (a) the utility with the lowest number of bad pipes (utility 10), (b) the utility with the highest number of bad pipes (utility 14).

Table 4 shows that year of construction is the most important feature in 13 of the utilities followed by age (8), groundwater (6), year of rehabilitation (5), and dimension (1). In general, there is a tendency for the continuous variables to be found relevant more often than categorical and binary variables.

To identify general trends between the performance, the size of dataset and the number of relevant features, the relation between the number of bad pipes, the performance, and the number of features contributing to the performance is shown in Figure 6.

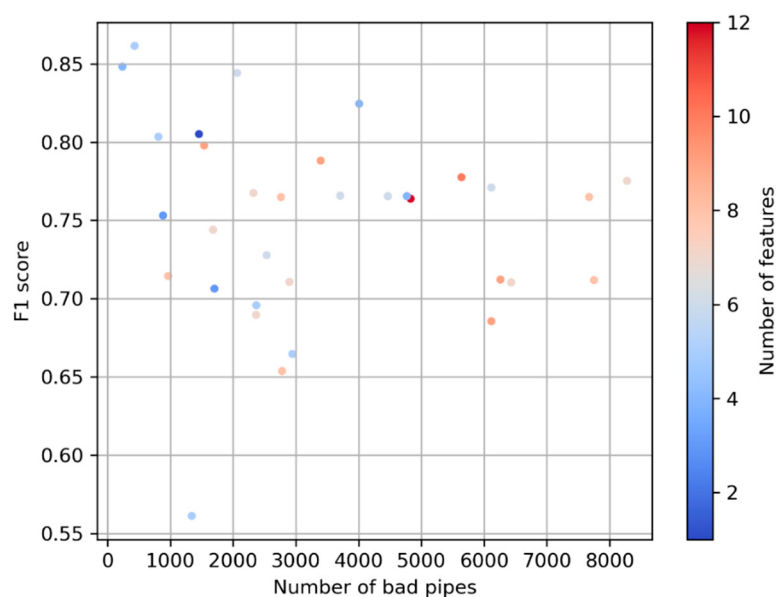


Figure 6. Plot of the relation between the number of bad pipes (first axis), the performance (second axis) and the number of relevant features.

In Figure 6, the number of bad pipes is shown along the first axis, the f1 score is shown along the second axis and the number of relevant features is shown in a color scale ranging from blue to red. For utilities with more than 6000 bad pipes the number of features contributing to the performance staggered at six to nine.

As stated in Section 2.4.3, only pipes in CS four were initially considered as being in bad condition. In that analysis the performance and number of relevant features staggered

for datasets with more than 1000–1500 bad pipes, which indicates that it is not solely the number of bad pipes that influences the results but also the total number of pipes inspected.

Table 4. Overview of which predictor variables contribute to the model performance for each utility. A “●” shows that the predictor variable contributes to the performance, a “○” shows that the predictor variable was included in the analysis but did not contribute to the model performance. The features are sorted in descending order, according to how often they contribute to the performance, and the utilities are sorted in descending order, according to number of pipes in CS 4. In addition, this table includes an overview of the most important feature and the best performance obtained for each utility.

Utility	Ground Level	Age	Groundwater	Wastewater	Length	Dimension	YoConst	YoRehab	Soil Type	Slope	Depth	NoBuildings	NoGrates	Material	DistRoad	Trees	Road Type	Rehab	City Type	Buildings Low	Buildings High	No. Features Relevant	Performance
Utility 14	●	●	○	○	○	●	●	○	○	○	●	○	●	○	●	○	○	○	○	○	○	7	0.78
Utility 1	●	●	●	○	●	●	○	○	●	○	○	○	●	●	○	○	○	○	○	○	○	8	0.71
Utility 24	●	●	●	○	○	●	○	○	●	●	○	○	●	●	○	○	○	○	○	○	○	8	0.76
Utility 6	●	●	●	●	○	○	●	○	○	○	●	○	○	○	○	●	○	○	○	○	○	7	0.71
Utility 16		●	●	●	●	○	●	●	○	○		●	○	○	●	○	○	●	○	○	○	9	0.71
Utility 12	●	○	●	●	●	●	○	●	●	○	○	●	○	○	○	○	○	○	●	○	○	9	0.69
Utility 28	●	●	●	○	●	●	●	○	○		○	○	○	○	○	○	○	○	○	○	○	6	0.77
Utility 5	●	●	●	●	○	●	●	○	●	●	○	●	○	○	○	○	●	○	○	○	○	10	0.78
Utility 29	●	○	●	●	●	●	●	●	●	●	○	○	●	●	○	●	○	○	○	○	○	12	0.76
Utility 3	●	○	●	○	○	●	●	○	○		○	○	○	○	○	○	○	○	○	○	○	4	0.77
Utility 9	●	●	●	●	●	○	●	○	○	○		○	○	○	○	○	○	○	○	○	○	6	0.77
Utility 2	○	●	●	●	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	4	0.82
Utility 25	●	●	●	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	6	0.77
Utility 18	●	●	○	●	●	○	○	○	○	○	●	○	○	○	○	○	●	○	○	○	○	9	0.79
Utility 27	○	○	●	●	●	○	○	●	○		○	○	○	○	○	○	●	○	○	○	○	5	0.66
Utility 19	○	○	●	○	●	○	●	●	○	○	○	●	●	○	○	●	○	○	○	○	○	7	0.71
Utility 4	●	●	○	●	○	●	○	○	●	●	○	○	●	○	○	●	○	○	○	○	○	8	0.65
Utility 20	●	●	○	●	●	●	●	●	○	●	○	○	○	○	○	○	○	○	○	○	○	8	0.76
Utility 21	●	●	○	●	●	●	●	○	○		○	○	○	○	○	○	○	○	○	○	○	6	0.73
Utility 30	●	●		○	●	●	○	○	○	○	○			○		●	○	○	○			5	0.70
Utility 11	○	●	●	●	●	○	●	○	○		●	○	○	○	○	○	●	○	○	○	○	7	0.69
Utility 31	●	●	●	○	●	○	○	○	●		○	○	○	●	●	○	○	○	○	○	○	7	0.77
Utility 22	●	●	●	○	●	●	○	○	○		●			○		○	○	○				6	0.84
Utility 32	○	○	●	●	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	3	0.71
Utility 8	●	○	○	●	○	●	○	●	○	○	○	●	○	○	○	○	○	○	○	○	○	7	0.74
Utility 23	●	●	○	●	○	○	○	○	○	●	●	○	○	○	○	○	○	●	○	○	○	9	0.80
Utility 17	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	1	0.81
Utility 33	●	●		●	○	○	○	●	○		●			○		○	○	○				5	0.56
Utility 7	○	●	●	●	●	○	○	●	●	●	○			●		○	○	○				8	0.71
Utility 26	○	●	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	●	○	○	○	3	0.75
Utility 15	○	○	○	●	○	●	●	○	○	○	○	○	○	○	●	○	○	○	●	○	○	5	0.80
Utility 13	○	○	○	●	○	○	●	●	○	○	●	●	○	○	○	○	○	○	○	○	○	5	0.86
Utility 10	●	○	○	○	●	○	●	○	○		○	○	○	○	○	○	●	○	○	○	○	4	0.85
Total	22	22	20	20	18	17	16	13	12	7	9	6	6	6	5	5	5	3	2	0	0		
Times best	0	8	6	0	0	1	13	5	0	0	0	0	0	0	0	0	0	0	0	0	0		

The smaller variation in performance and number of relevant features for utilities with a higher number of inspected pipes in bad condition, could indicate that these datasets contain a more representative segment of the pipes. Thereby they are less sensitive to a high or low occurrence of defects in an inspected area.

4. Discussion

Sewer deterioration modeling is complicated by several influencing factors. In this section the most prominent factors influencing the results are discussed, and the results are compared to previous findings in the literature.

4.1. Representativeness of Data

The results from the baseline experiment underlined the challenges of using historical data for sewer deterioration modeling, as the CCTV-inspections generally have been performed with a specific purpose, introducing a selective survival bias in the data [1]. However, as the datasets are comprehensive, most utilities do not have the finances to create a new dataset. Instead, the model developers must account for this by excluding the features in which the bias is most prominent, such as features related to geographical position. In the long term, utilities should include some spatial randomness in their strategy for CCTV-inspection.

4.2. Definition of Target Variable

Lack of publicly available data [5], numerous different standards for CCTV-inspections and different methods for evaluation of sewer deterioration models complicate the comparison of deterioration models. This also applies to the performance obtained in experiment two, where the f1-score drops from 0.73 to 0.35 when solely considering pipes in CS four as being in bad condition instead of considering pipes in both CS three and four. However, although the performance was affected, there was a high correlation in the predictor variables relevant for prediction of pipes in CS four and pipes in either CS three or four, which indicates that it is fair to make a binary evaluation of the feature importance.

Today CCTV inspections are performed by an operator who manually annotate the observations found in the sewers according to a given standard for tv inspections. These observations are often transformed into a general measure of the sewers condition. This condition measure can either be based on general standards or they can be utility specific. The benefit of utilizing the general standards are increased comparability between utilities whereas the benefit of utilizing a utility specific performance measure is that it can be adjusted to prioritize the types of defects relevant for the utility. For instance, a utility with limited capacity at the wastewater treatment plant might increase weight on infiltration. Weighting some defects higher can cause the features related to these defects to become more important in a feature analysis. In the CS used in this study a higher weight has been put on attached deposit and infiltration according to other observation types as shown in Table 2. This is consistent with the results showing a high importance of the relative groundwater level. As the groundwater maps available for this study were based on measurements every 500 m, the actual groundwater level can change significantly between the data points. It is likely that the ground level can compensate for these changes, which will induce a higher weight on this feature in the feature analysis.

4.3. Size of Datasets

When considering pipes in both CS three and CS four as being in bad condition, the performance and the number of features relevant staggered for datasets with more than 6000 pipes in bad condition. In the initial analysis only pipes in CS four were considered bad. In that analysis the performance and number of relevant features staggered for datasets with more than 1000–1500 pipes in bad condition. This indicates that the number of bad pipes required for optimal performance is correlated with how the target variable is defined and the total number of pipes inspected.

Furthermore, it is worth noticing that if solely considering the utilities with more than 1000 bad pipes, there is more consensus on which features contribute to the performance. For ground level the percentage of time it is found to be relevant increases from 69% to 78%. Similar tendencies are present for age (67% to 71%) and relative groundwater level (65% to 73%). A full overview is presented in Table 5.

Table 5. Overview of how often a feature contributes to the performance when considering all the utilities and when considering only utilities with more than 1000 bad pipes.

Predictor Variables	All Utilities		Utilities with More Than 1000 Bad Pipes	
	Times Present	Percent of Times Found Relevant	Times Present	Percent of Times Found Relevant
Ground level	32	69	27	78
Age	33	67	28	71
Groundwater	31	65	26	73
Wastewater	33	61	28	61
Length	33	55	28	57
Dimension	33	52	28	57
Year of construction	33	48	28	46
Year of rehabilitation	33	39	28	39
Soil type	33	36	28	36
Slope	23	30	19	32
Depth	31	29	26	31
No. buildings	29	21	25	20
No. grates	29	21	25	24
Material	33	18	28	21
Dist. to road center	29	17	25	16
Dist. to trees	33	15	28	18
Road types	33	15	28	14
Rehabilitation type	33	9	28	7
City type	30	7	26	4
Building low	29	0	25	0
Buildings high	29	0	25	0

4.4. Irregularities in the Step Analysis

For some utilities, the performance improved when predictor variables were removed, indicating overfitting of the model. This was clearest for utility 10, which is also the utility with the smallest amount of training data. For datasets with more than 10,000 pipes, the tendency could still be observed in some cases after cleaning but removing features did not lead to an increase in performance of more than two to three percent.

In a few cases, the performance suddenly increased when removing one parameter. This could not be explained by stochasticity in the performance or overfitting. An example of this can be seen in Figure 5b. This is most likely because some predictor variables perform well when combined but introduce noise when considered individually.

4.5. Comparison to the Literature

Mohammadi et al. [24] reviewed 24 papers, of which 19 had investigated which features were significant. In Table 6 the results of this study are compared to the findings by Mohammadi et al.

Table 6. Comparison of how often the different features contribute to the performance in this study and in the review by Mohammadi et al.

Predictor Variables	Results		Mohammadi et al.	
	Times Present	Percent of Times Found Relevant	Times Present	Percent of Times Found Relevant
Ground level	27	78	-	-
Age	28	71	18	78
Groundwater	26	73	3	100
Wastewater	28	61	6	83
Length	28	57	11	91
Dimension	28	57	17	71
Year of construction	28	46	-	-
Year of rehabilitation	28	39	-	-
Soil type	28	36	5	20
Slope	19	32	12	42
Depth	26	31	16	44
No. buildings	25	20	-	-
No. grates	25	24	-	-
Material	28	21	15	67
Dist. to road center	25	16	-	-
Dist. to trees	28	18	-	-
Road types	28	14	5	40
Rehabilitation type	28	7	-	-
City type	26	4	-	-
Building low	25	0	-	-
Buildings high	25	0	-	-
Location	-	-	5	40
Up-invert	-	-	1	0
Down-invert	-	-	1	0
Bedding type	-	-	2	100
Corrosivity	-	-	2	50
Number of trees	-	-	5	60
Traffic	-	-	1	1
Flow	-	-	3	67
Hydrohalic	-	-	2	100
Location	-	-	5	40
Up-invert	-	-	1	0

In the review by Mohammadi, there is a higher consensus about which predictor variables are significant. The most probable reason for this is that Mohammadi et al. reviewed studies whose authors selected a number of predictor variables. For example, four of the papers investigated between two and eight predictor variables and did not find any insignificant variables. In general, there is a consensus that length, age, dimension, ground water, and wastewater type are often important predictor variables. However, the model developer should consider the specific case when selecting predictor variables as there is no “gold standard”.

4.6. CCTV-Inspection Planning

The still increasing access to pipe specific data and the increasing awareness of the benefits related to risk based pipe inspection and rehabilitation on pipe level are essentials when optimizing the management of sewer systems to save costs and resources. Sewer deterioration modeling is an essential element in this; however, the scientific literature dealing with the underlying parameters influencing the deterioration models is sparse. The findings of this study enlighten some of these shortcomings, and the findings can be incorporated in future model development.

Generally, deterioration models can be used to give a snapshot of the sewer system and is used when no CCTV-inspection has been made or when the CCTV inspection is outdated. Typically, the deterioration models are based on datasets which have been collected over several years. Therefore, users of deterioration models should be aware that the predictions of the CSs are evaluated on historical data and thereby cannot give a fair prediction of future condition states. For example, plastic pipes were rarely used 50 years ago, and plastic pipes older than 50 years have limited representation in the data. Furthermore, the surrounding environment, material quality etc. change over time. Future predictions of CSs are further complicated by variations in the degradation profile of different defect types. Some defect types occur stochastically and do not degrade over time such as defects related to pipe connections or installation of the pipes. Other defects degrade over time such as surface damage. Surface damage is often seen in concrete pipes due to the presence of hydrogen sulphide which erode the surface over time. Hydrogen sulphide is typically formed in pump pipes. Likewise, the degradation profile for defects related to roots in the pipes depends on the surrounding trees and their growth.

5. Conclusions

The primary contribution of this paper is a comprehensive analysis of the feature importance in sewer deterioration modeling. The paper addresses factors that influence sewer deterioration modeling and acknowledges weak or missing information in the literature, such as handling of biased datasets, the impact of how bad pipes are defined, and the variations in feature importance between utilities.

Deterioration models are usually based on CCTV-inspections performed over several years with a specific purpose in mind. This is problematic due to a selective survival bias in the data whereby the models do not perform as well on noninspected areas as they do on inspected areas. Ideally the datasets should be random in character, but due to economic constraints this is often infeasible. Instead, model developers should avoid utilization of geographically related parameters. Moreover, utilities should include randomness in their strategy for CCTV inspection.

Changing the definition of when a pipe is in bad condition produced large deviations in model performance. However, in the feature analysis it was the same features that contributed to the performance, although more features contributed when both pipes in CS three and four were considered bad than when only pipes in CS four were considered bad. This indicates that it is fair to use an advantageous split between good and bad pipes when making a feature analysis.

Comparison of feature analysis from 33 different utilities showed a relatively high variance in the number of features contributing to the performance, which features contributed, and the performance obtained by the models. These variations were especially high for utilities with fewer than 6000 pipes in bad condition. It is worth noting that the number of bad pipes depends on the definition of bad pipes. When solely considering pipes in CS four as “bad”, the high variations were primarily present for utilities with fewer than 1000–1500 pipes in bad condition.

No feature was considered relevant in more than 69% of the utility specific models; however, when only considering utilities with more than 1000 bad pipes there was a higher consensus on which features were relevant (up to 78%). For these utilities, the features that contributed to the performance most of the time were ground level (78%), age (71%), groundwater level (73%), wastewater type (61%), length (57%), dimension (57%), year of construction (46%), and year of rehabilitation (39%). As there is a high redundancy between year of construction, year of rehabilitation, and age, removing one of these as a possible predictor variable would most likely induce the others to contribute to the performance in more cases. In 26 out of 33 cases the most important feature was related to either age, year of construction, or year of rehabilitation. On average 6.5 features contributed to the utility specific models.

The overall trends in feature importance found in this work showed consensus with the findings in a review by Mohammadi et al. [24]; however, due to variations in study design of the articles reviewed by Mohammadi et al. the two papers are not comparable on a detailed level.

The added value of this paper is a better understanding of the underlying parameters influencing sewer deterioration modeling and knowledge of feature importance when encountering the statistical variations between utilities. The exact results related to feature importance are specific to the condition measure used in the study, however, the overall trends are comparable to findings in the literature and can be used to assist the feature selection for sewer deterioration modeling, which is important because feature extraction is a labor intensive process.

Author Contributions: Conceptualization, B.D.H., M.U., T.B.M. and D.G.J.; methodology, B.D.H., T.B.M., D.G.J.; software, B.D.H.; validation, B.D.H.; formal analysis, B.D.H. and S.H.R.; investigation, B.D.H., S.H.R., M.U., T.B.M. and D.G.J.; resources, B.D.H.; data curation, B.D.H. and S.H.R.; writing—original draft preparation, B.D.H.; writing—review and editing, B.D.H., S.H.R., M.U., T.B.M. and D.G.J.; visualization, B.D.H.; supervision, S.H.R., M.U., T.B.M. and D.G.J.; project administration, B.D.H. and D.G.J.; funding acquisition, B.D.H., M.U., T.B.M. and D.G.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Innovation Fund Denmark.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tscheikner-Gratl, F.; Caradot, N.; Cherqui, F.; Leitão, J.P.; Ahmadi, M.; Langeveld, J.G.; Le Gat, Y.; Scholten, L.; Roghani, B.; Rodríguez, J.P.; et al. Sewer Asset Management—State of the Art and Research Needs. *Urban Water J.* **2019**, *16*, 662–675. [\[CrossRef\]](#)
2. Marlow, D.; Pearson, L.; MacDonald, D.H.; Whitten, S.; Burn, S. A Framework for Considering Externalities in Urban Water Asset Management. *Water Sci. Technol.* **2011**, *64*, 2199–2206. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Roghani, B.; Cherqui, F.; Ahmadi, M.; Le Gauffre, P.; Tabesh, M. Dealing with Uncertainty in Sewer Condition Assessment: Impact on Inspection Programs. *Autom. Constr.* **2019**, *103*, 117–126. [\[CrossRef\]](#)
4. Dirksen, J.; Clemens, F.H.L.R.; Korving, H.; Cherqui, F.; Le Gauffre, P.; Ertl, T.; Plihal, H.; Müller, K.; Snaterse, C.T.M. The Consistency of Visual Sewer Inspection Data. *Struct. Infrastruct. Eng.* **2013**, *9*, 214–228. [\[CrossRef\]](#)
5. Haurum, J.B.; Moeslund, T.B. A Survey on Image-Based Automation of CCTV and SSET Sewer Inspections. *Autom. Constr.* **2020**, *111*, 103061. [\[CrossRef\]](#)
6. Altarabsheh, A.; Ventresca, M.; Kandil, A. New Approach for Critical Pipe Prioritization in Wastewater Asset Management Planning. *J. Comput. Civ. Eng.* **2018**, *32*, 04018044. [\[CrossRef\]](#)
7. Elmasry, M.; Zayed, T.; Hawari, A. Multi-Objective Optimization Model for Inspection Scheduling of Sewer Pipelines. *J. Constr. Eng. Manag.* **2019**, *145*, 04018129. [\[CrossRef\]](#)
8. Ghavami, S.M.; Borzooei, Z.; Maleki, J. An Effective Approach for Assessing Risk of Failure in Urban Sewer Pipelines Using a Combination of GIS and AHP-DEA. *Process Saf. Environ. Prot.* **2020**, *133*, 275–285. [\[CrossRef\]](#)
9. Elmasry, M.; Hawari, A.; Zayed, T. An Economic Loss Model for Failure of Sewer Pipelines. *Struct. Infrastruct. Eng.* **2018**, *14*, 1312–1323. [\[CrossRef\]](#)
10. Carvalho, G.; Amado, C.; Brito, R.S.; Coelho, S.T.; Leitão, J.P. Analysing the Importance of Variables for Sewer Failure Prediction. *Urban Water J.* **2018**, *15*, 338–345. [\[CrossRef\]](#)
11. Hansen, B.D.; Rasmussen, S.H.; Moeslund, T.B.; Uggerby, M.; Jensen, D.G. Sewer Deterioration Modeling: The Effect of Training a Random Forest Model on Logically Selected Data-Groups. *Procedia Comput. Sci.* **2020**, *176*, 291–299. [\[CrossRef\]](#)
12. Hansen, B.D.; Getreuer Jensen, D.; Rasmussen, S.H.; Tamouk, J.; Uggerby, M.; Moeslund, T.B. General Sewer Deterioration Model Using Random Forest. In Proceedings of the 2019 IEEE Symposium Series on Computational Intelligence (SSCI), Xiamen, China, 6–9 December 2019; pp. 834–841.
13. Laakso, T.; Kokkonen, T.; Mellin, I.; Vahala, R. Sewer Condition Prediction and Analysis of Explanatory Factors. *Water* **2018**, *10*, 1239. [\[CrossRef\]](#)
14. Park, H.; Ting, S.H.; Jeong, H.D. Procedural Framework for Modeling the Likelihood of Failure of Underground Pipeline Assets. *J. Pipeline Syst. Eng. Pract.* **2016**, *7*, 04015023. [\[CrossRef\]](#)
15. Le Gat, Y. Modelling the Deterioration Process of Drainage Pipelines. *Urban Water J.* **2008**, *5*, 97–106. [\[CrossRef\]](#)
16. Elmasry, M.; Hawari, A.; Zayed, T. Defect Based Deterioration Model for Sewer Pipelines Using Bayesian Belief Networks. *Can. J. Civ. Eng.* **2017**, *44*, 675–690. [\[CrossRef\]](#)
17. Hawari, A.; Alkadour, F.; Elmasry, M.; Zayed, T. Condition Assessment Model for Sewer Pipelines Using Fuzzy-Based Evidential Reasoning. *Aust. J. Civ. Eng.* **2018**, *16*, 23–37. [\[CrossRef\]](#)

18. Kabir, G.; Balek, N.B.C.; Tesfamariam, S. Sewer Structural Condition Prediction Integrating Bayesian Model Averaging with Logistic Regression. *J. Perform. Constr. Facil.* **2018**, *32*, 04018019. [[CrossRef](#)]
19. Rokstad, M.M.; Ugarelli, R.M. Evaluating the Role of Deterioration Models for Condition Assessment of Sewers. *J. Hydroinform* **2015**, *17*, 789–804. [[CrossRef](#)]
20. Fuchs-Hanusch, D.; Günther, M.; Möderl, M.; Muschalla, D. Cause and Effect Oriented Sewer Degradation Evaluation to Support Scheduled Inspection Planning. *Water Sci. Technol.* **2015**, *72*, 1176–1183. [[CrossRef](#)] [[PubMed](#)]
21. O'Reilly, M.P.; Rosbrook, R.B.; Cox, G.C.; McCloskey, A. *Analysis of Defects in 180 Km of Pipe Sewers in Southern Water Authority*; Research Report (Transport and Road Research Laboratory); Ground Engineering Division, Structures Group, Transport and Road Research Laboratory: Berkshire, UK, 1989.
22. Yin, X.; Chen, Y.; Bouferguene, A.; Al-Hussein, M. Data-Driven Bi-Level Sewer Pipe Deterioration Model: Design and Analysis. *Autom. Constr.* **2020**, *116*, 103181. [[CrossRef](#)]
23. Davies, J.P.; Clarke, B.; Whiter, J.; Cunningham, R.; Leidi, A. The Structural Condition of Rigid Sewer Pipes: A Statistical Investigation. *Urban Water* **2001**, *3*, 277–286. [[CrossRef](#)]
24. Malek Mohammadi, M.; Najafi, M.; Kermanshachi, S.; Kaushal, V.; Serajiantehrani, R. Factors Influencing the Condition of Sewer Pipes: State-of-the-Art Review. *J. Pipeline Syst. Eng. Pract.* **2020**, *11*, 03120002. [[CrossRef](#)]
25. Hawari, A.; Alkadour, F.; Elmasry, M.; Zayed, T. A State of the Art Review on Condition Assessment Models Developed for Sewer Pipelines. *Eng. Appl. Artif. Intell.* **2020**, *93*, 103721. [[CrossRef](#)]
26. Guzmán-Fierro, J.; Charry, S.; González, I.; Peña-Heredia, F.; Hernández, N.; Luna-Acosta, A.; Torres, A. Bayesian Network-Based Methodology for Selecting a Cost-Effective Sewer Asset Management Model. *Water Sci. Technol.* **2020**, wst2020299. [[CrossRef](#)] [[PubMed](#)]
27. Laden, B.; DANVA. *Fotomanualgruppen Fotomanualen: TV-Inspektion af Afløbsledninger*; DANVA: Skanderborg, Denmark, 2010; ISBN 978-87-90455-81-1.
28. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13 August 2016; pp. 785–794.
29. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
30. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
31. Harvey, R.R.; McBean, E.A. Predicting the Structural Condition of Individual Sanitary Sewer Pipes with Random Forests. *Can. J. Civ. Eng.* **2014**, *41*, 294–303. [[CrossRef](#)]
32. Regression Methods for Predicting Rate and Type of Failures of Water Conduits. *Ecol. Chem. Eng. A* **2017**. [[CrossRef](#)]