CETRA2026 27-30 May 2026, Bol, Island Brač, Croatia
9th International Conference on Road and Rail Infrastructure

ANALYSIS OF THE IMPACT OF REDUCING EVALUATION ITEMS ON URGENCY CLASSIFICATION IN SEWER PIPE INSPECTION

Shuta Notsu[1]**, Makoto Fujiu**[2]**, Yuma Morisaki**[2]
[1] *Kanazawa University, Division of Geosciences and Civil Engineering, Japan*
[2] *Kanazawa University, Faculty of Transdisciplinary Sciences for innovation, Institute of Transdisciplinary Sciences, Japan*

Abstract

In recent years, the deterioration of sewer pipes has become a serious issue, and road collapse accidents caused by pipe failures have occurred in various locations, resulting in casualties and impacts. To prevent such accidents, it is essential to accurately assess the condition of sewer pipes and identify sections that require early repair. Since sewer pipes are buried underground, deterioration cannot be evaluated based on external appearance alone, and regular inspections play a critical role in maintenance. In sewer pipe inspections, a closed-circuit television (CCTV) camera is inserted into the pipe, and multiple damage-related items are assessed based on the recorded footage. These items include breaks, cracks, joint misalignment, water ingress, protruding branch pipes, grease buildup, tree root intrusion, and mortar buildup. The urgency level is determined through a comprehensive evaluation of these items. However, the number of evaluation items is large, and judgments that consider the severity and combination of damage types require advanced expertise. As a result, variability in inspector judgments and inspection delays caused by labor shortages have become significant challenges, increasing the demand for efficient inspection procedures. This study develops a method for classifying urgency levels at the span level using Graph Neural Networks (GNNs) based on sewer pipe inspection records. The input data consists of pipe attribute information and multiple damage-related evaluation items. GNNs can learn relationships between nodes, making them suitable for urgency classification that reflects similarities between pipe spans. Furthermore, the eight damage-related items evaluated through CCTV inspections are exploratively reduced from the input features of the GNN, and changes in urgency classification performance are analyzed as the number of items decreases. Through this process, combinations of evaluation items that achieve classification accuracy comparable to those obtained using all damage-related items are investigated.

Keywords: sewer pipes, Graph Neural Networks, damage indicators, urgency assessment

1 Introduction

In developed countries, the aging of sewer infrastructure, including sewer pipelines, is widely recognized as a major maintenance challenge. Industry associations have reported cases of aging sewer pipes in Europe and the United States [1]. For example, failures of aging sewer pipes have caused road collapses and large-scale sewage spills in various locations, some of which have escalated into accidents involving traffic disruptions and resident evacuations. In addition, residential flooding caused by the combined effects of pipe aging and torrential rainfall has become a serious social issue.

DOI: https://doi.org/10.5592/CO/CETRA.2026.2042

In urban areas, frequent overflows and blockages originating from aging sewer pipes further highlight the insufficient inspection and renewal of sewer infrastructure. In Japan, social infrastructure facilities that were intensively developed during the period of rapid economic growth are progressively aging, making the maintenance and management of sewer pipelines a critical issue. Sewer pipelines form essential infrastructure supporting daily life and urban functions; however, because most pipelines are buried underground, assessing their deterioration from the surface is difficult. As a result, planned inspections and appropriate renewal are indispensable. At the same time, local governments face severe financial constraints and labor shortages, leading to a strong demand for more efficient maintenance of the extensive sewer pipeline stock. According to statistics from the Ministry of Land, Infrastructure, Transport and Tourism, the total length of sewer pipelines in Japan is about 500, 000 km, and the proportion of pipelines exceeding the standard service life of 50 years is expected to increase in the future [2].

Against this background, sewer pipeline inspections are conducted worldwide to assess their condition. In many countries, in-pipe camera surveys are widely adopted as a standard inspection method. However, Japan employs a distinctive evaluation approach: whereas Europe and the United States typically perform numerical assessments based on individual defects, Japan determines urgency at the span level by comprehensively interpreting multiple types of damage within each span. Figure 1 illustrates annual trends in newly installed sewer pipeline length and total sewer pipeline length in Japan. In Japanese sewer inspections, in-pipe camera surveys assess eight damage items: structural damage, cracks, joint misalignment, water ingress, protruding branch pipes, grease buildup, tree root intrusion, and mortar buildup. Based on these assessments, urgency is determined for each span and classified into four levels, ranging from I (highest) to IV (lowest). However, the large number of evaluation items and the requirement for advanced technical knowledge and inspection experience to interpret damage severity and combinations have resulted in variability in judgments and increased practical burdens during inspection and evaluation.

In recent years, numerous studies have been reported on sewer pipe and culvert maintenance, focusing on deterioration prediction and condition assessment using inspection records and structural attributes. Several studies have demonstrated the effectiveness of data-driven approaches by applying machine learning and deep learning techniques to condition assessment and integrity classification based on inspection data [3, 4]. However, many of these previous studies treat pipes as independent entities, and condition assessment that accounts for network structure or relationships with surrounding pipes has not been sufficiently explored. Moreover, the effects of organizing or reducing evaluation items and information volume on assessment performance have not been adequately discussed.

Therefore, this study develops a method for classifying urgency levels at the span level using Graph Neural Networks (GNNs) [5]. The proposed method utilizes inspection records from sewer pipelines managed by a Japanese municipality, incorporating pipe attribute information and multiple evaluation items as input features. In addition, this study analyzes how urgency classification performance changes as the number of evaluation items are progressively reduced and investigates the conditions under which reliable urgency determination remains feasible even with a limited set of evaluation items.

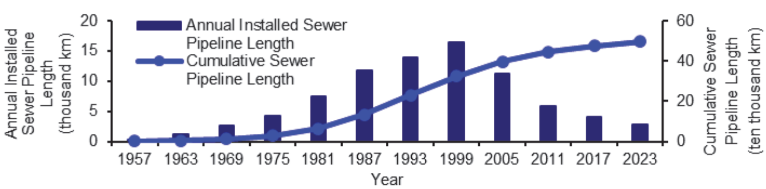

Figure 1 Trends in annual installed length and total length of sewer pipelines

2 Overview of data used

This study aims to determine the urgency level for sewer pipe maintenance using inspection records obtained from actual inspection operations. The dataset consists of television camera survey records collected from fiscal year 2019 to fiscal year 2021, which include information on internal pipe damage and pipe attributes. The analysis unit is defined as a span, where one manhole-to-manhole section constitutes a single span. Information from multiple pipes within each span is aggregated and treated collectively. The dataset comprises 1, 100 spans, of which 1, 000 are used for model construction and 100 for testing. Each span is assigned structural attributes, including installation area, span length, pipe diameter, pipe type, number of pipes, and pipe age. In addition, eight types of internal pipe damage identified by TV camera surveys are recorded: structural damage, crack, joint displacement, infiltration, grease deposition, protruding lateral connection, root intrusion and mortar deposition. Each damage item is evaluated on a three-level scale: a (severe), b (moderate), and c (minor). Each span is also assigned an urgency level based on inspection results, which represents an overall assessment of the span and is used as the output label for the GNN. The urgency levels considered in this study are II, III, and IV, while urgency level I is excluded. The a, b, and c categories represent the severity levels for each damage item and are used as input features for the GNN. A portion of the dataset is presented in table 1.
Because the number of damage occurrences depends on span length, damage assessments are normalized by span length to construct comparable input features across spans. Figure 2 shows the distribution of damage evaluation levels in the normalized training dataset. For Grease deposition and Infiltration, a relatively high proportion of a evaluations is observed, indicating the presence of severe damage. In contrast, structural damage, root intrusion and mortar deposition are dominated by b evaluations, while protruding lateral connection shows a higher proportion of c evaluations. In addition, some damage types, such as root intrusion, occur infrequently, indicating that both occurrence frequency and severity distribution vary among damage types. Based on these considerations, this study uses span-level inspection data, inputs pipe attributes and damage evaluation information for eight damage items into the GNN, and estimates urgency levels.

Table 1 Example of the dataset

Structural attributes (Partial List)		Structural damage			Crack			Joint displacement			Infiltration			Grease deposition			Protruding lateral connection			Root intrusion			Mortar deposition			Urgency level
Pipe age [year]	Pipe diameter [mm]	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	
31	250	0	1	0	0	2	0	0	0	1	1	2	0	0	1	0	0	0	3	0	0	0	0	0	1	III
39	400	1	2	0	0	0	1	0	0	0	2	1	0	0	0	0	0	1	0	0	0	2	1	1	0	II
25	250	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	1	0	2	0	0	0	0	0	0	IV

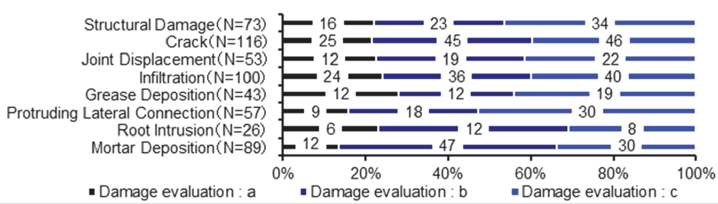

Figure 2 Damage evaluation proportions after normalization

3 GNN-based emergency severity assessment method

3.1 Overview of GNN

GNN is a deep learning model designed to handle graph-structured data composed of nodes and edges, enabling feature learning that accounts for relationships between nodes. Unlike conventional neural networks that process regular data such as images or time series, GNNs can model irregular structures in non-Euclidean spaces. In GNNs, node representations are updated by repeatedly aggregating and transforming information from the node itself and its neighboring nodes. In this study, sewer pipe spans are treated as nodes. Relationships between nodes are defined not by physical connectivity but by similarities in pipe attributes and damage characteristics, allowing spans with similar features to influence each other during learning. Consequently, GNN can incorporate both individual span information and inter-span relationships, making it suitable for sewer pipe urgency classification.

3.2 Training the GNN

This study formulated the urgency determination task as a node classification problem by constructing a graph in which sewer pipe spans were treated as nodes. Node connections were defined based on similarity between feature vectors composed of physical pipe attributes and damage assessment information, rather than physical connectivity in the sewer network. Cosine similarity was used to connect each span to the three most similar spans, allowing similarity relationships between non-adjacent spans to be reflected in the learning process. The dataset consists of 1, 100 spans, with 1, 000 used for model construction and 100 for testing. The training data were further split into training and validation sets at an 8:2 ratio. Input features comprise a 15-dimensional feature vector of physical attributes and span-length-normalized damage evaluation indices, while the output labels are urgency levels (II, III, and IV). The model architecture includes an input layer, two hidden layers, and an output layer. ReLU activation and Dropout (p = 0.5) are applied to the hidden layers to mitigate overfitting, enabling each node to aggregate information from neighboring nodes as well as nodes up to two hops away. The model is trained using the Adam optimizer with a learning rate of 0.01 for 80 epochs, and cross-entropy loss is employed.
As shown in figures 3 and 4, loss values for both training and validation datasets decrease rapidly in the early training stage and converge stably, while accuracy remains high for both datasets. The small gap between them indicates suppressed overfitting and good generalization performance. Overall, the proposed model achieves stable span-level urgency estimation by integrating pipe attributes, damage evaluation information, and inter-span relationships based on cosine similarity.

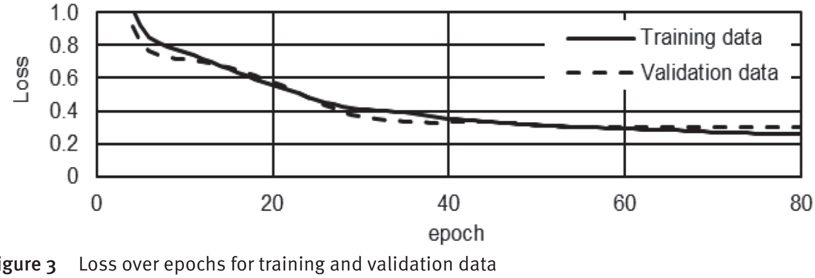

Figure 3 Loss over epochs for training and validation data

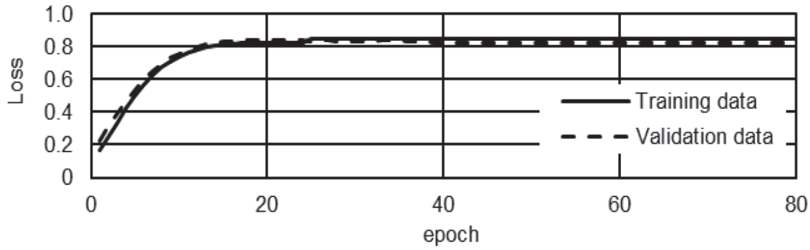

Figure 4 Accuracy over epochs for training and validation data

3.3 Performance evaluation of the constructed model

To verify the predictive performance of the constructed GNN model, emergency level estimations were performed on test data not used for training. Based on the confusion matrix shown in table 1, precision and recall (equations 1 and 2) were calculated to quantitatively evaluate the model's classification performance.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

The classification results for urgency levels (II/III/IV) on the test data yielded a precision of 0.822 and a recall of 0.871, both exceeding 0.8. These results confirm that the constructed model can perform stable urgency classification even on data not used for training, demonstrating consistent generalization performance. Notably, the high recall value indicates that high-urgency spans are being detected without omission, a critical characteristic for maintenance operations. Furthermore, the high precision level suggests that judgments are made while suppressing excessive false detections. A key factor contributing to this classification performance is the use of multiple damage evaluation items as features, in addition to the physical attributes of each span. This approach explicitly captures differences in damage severity and occurrence tendencies as features, likely making it easier for the model to learn the state differences between spans. Furthermore, constructing a graph structure based on cosine similarity allowed the model to learn relationships between spans with similar attributes and damage tendencies, even when physically disconnected. This is thought to have contributed to the stabilization of the urgency assessment. On the other hand, this analysis used all damage evaluation items, which would impose a significant burden on practical inspection work. Therefore, to reduce this burden, it is necessary to consider reducing the number of evaluation items.

Table 2 Confusion matrix for urgency classification

Correct / Prediction	II	II	II
II	TP	FP	FP
III	FN	TP	FP
IV	FN	FN	TN

4 Reducing evaluation items using association analysis

4.1 Overview of association analysis

Association analysis is a data analysis method used to extract co-occurrence tendencies and relationships among multiple items and is widely applied to identify relationships between variables [6]. A key feature of this method is its ability to quantitatively evaluate how frequently one item occurs when another specific item is present, making it suitable for analyzing inspection datasets that include multiple evaluation items. In association analysis, the lift value is used as an indicator of relationship strength between items. Lift compares the conditional probability of one item occurring given the occurrence of another item with the probability expected under independence. A lift value greater than 1 indicates a positive association, and larger values imply a higher likelihood of co-occurrence within the same observation unit.

In this study, association analysis is applied to eight damage items evaluated during sewer pipe television camera inspections to identify co-occurrence relationships. The objective is to clarify the structural relationships among damage items and to explore the potential for reducing the number of evaluation items to alleviate practical inspection burdens.

4.2 Analysis of relationships among damage evaluation items

To investigate relationships among damage items assessed through sewer pipe television camera inspections, association analysis was applied to examine co-occurrence tendencies between items. To focus on relationships with clear relevance, only item pairs with Lift values greater than 1.5 were extracted. Figure 5 illustrates the network of relationships among items based on lift values. Figure 5 shows strong associations among several damaged items. High co-occurrence is observed among structural damage, crack and infiltration, indicating that these damages frequently occur within the same span. These items are closely associated with pipe deterioration and reduced structural integrity and are likely to originate from common degradation processes. In addition, mortar deposit exhibits relatively strong associations with infiltration and protruding lateral connection, while protruding lateral connection is also strongly associated with joint displacement. This suggests that multiple damage types tend to occur sequentially at pipe connection points, which are structurally discontinuous and sensitive to construction conditions and surrounding environmental influences. In contrast, root intrusion and grease deposition show no strong co-occurrence relationships with other damage items, indicating that these damage types may be relatively independent and strongly influenced by local or external factors.

These results demonstrate that distinct co-occurrence relationships and correlation structures exist among sewer pipe damage items. For damage groups with strong associations, such as structural damage, crack and infiltration, the evaluation items may contain overlapping information. Therefore, selecting representative items for urgency assessment could enable a reduction in the number of evaluation items. Similarly, for items sharing structural characteristics, such as protruding lateral connection, joint displacement and mortar deposition, representative item selection may also be feasible. Rather than treating evaluation items as completely independent, integrating items based on co-occurrence tendencies is considered an effective approach for reducing evaluation items.

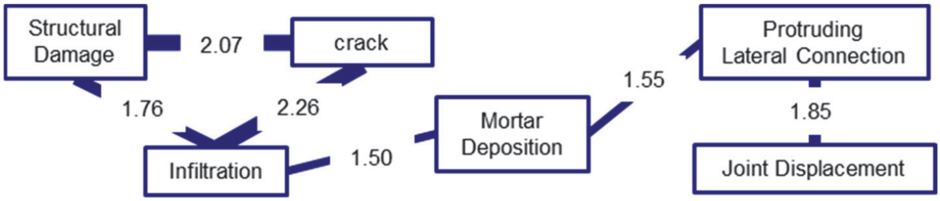

Figure 5 Relationship network based on Lift values

5 Conclusion

This study developed a span-based urgency assessment method using GNN for sewer pipe maintenance, incorporating pipe attributes and damage evaluation information obtained from inspections. By constructing a graph structure based on feature-vector similarity, with each span treated as a node, relationships among spans with similar attributes and damage characteristics are integrated into the learning process. The results demonstrate stable classification performance on test data, confirming the effectiveness of the GNN-based urgency assessment approach. In addition, association analysis was applied to eight damage items evaluated during television camera inspections to examine their co-occurrence relationships. Strong associations were identified among structural damage, crack and infiltration, while associations were also observed among mortar deposition, protruding lateral connection and joint displacement. These findings indicate that grouping damage items based on co-occurrence tendencies and occurrence mechanisms may reduce the number of evaluation items while preserving information required for urgent assessment.

Future work includes reconstructing the GNN model using features derived from a reduced set of evaluation items and quantitatively comparing model performance to assess the impact of item reduction on urgency assessment accuracy. Improving practical applicability will further require incorporating features that account for spatial relationships and temporal progression of damage, as well as graph definitions that partially reflect the actual sewer network structure.

References

[1] Japan Society for Trenchless Technology, Overseas situations and issues related to aging sewer pipelines, https://www.hinkakukyo.jp/rensai/pdf/200910-01.pdf

[2] Sewerage Pipe Maintenance and Management, Ministry of Land, Infrastructure, Transport and Tourism (MLIT), Japan, https://www.mlit.go.jp/mizukokudo/sewerage/crd_sewerage_tk_000135.html

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

[4] Li, X., Khademi, F., Liu, Y., Akbari, M., Wang, C., Bond, L.P., Keller, J., Jiang, G.: Evaluation of data-driven models for predicting the service life of concrete sewer pipes subjected to corrosion, Journal of environmental management, 234 (2019), pp. 431-439

[5] Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G.: The graph neural network model, IEEE transactions on neural networks, 20 (2008) 1, pp. 61-80

[6] Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases, 1993 ACM SIGMOD International Conference on Management of Data, pp. 207–216, 1993.