

# 機械学習を用いた下水道管渠における不良発生率ランクの推定に関する分析

諏訪 太紀<sup>1</sup>・藤生 慎<sup>2</sup>・森崎 裕磨<sup>3</sup>・福岡 知隆<sup>4</sup>・吉倉 麻衣<sup>5</sup>

<sup>1</sup> 学生会員 金沢大学大学院 自然科学研究科地球社会基盤学専攻 (〒920-1192 石川県金沢市角間町)

E-mail: s.taiki@stu.kanazawa-u.ac.jp

<sup>2</sup> 正会員 金沢大学准教授 融合研究域融合科学系 (〒920-1192 石川県金沢市角間町)

E-mail: fujiu@se.kanazawa-u.ac.jp (Corresponding Author)

<sup>3</sup> 正会員 金沢大学特任助教 融合研究域融合科学系 (〒920-1192 石川県金沢市角間町)

E-mail: morisaki@staff.kanazawa-u.ac.jp

<sup>4</sup> 正会員 金沢大学博士研究員 融合研究域融合科学系 (〒920-1192 石川県金沢市角間町)

E-mail: tfukuoka@se.kanazawa-u.ac.jp

<sup>5</sup> 学生会員 金沢大学大学院 自然科学研究科環境デザイン学専攻 (〒920-1192 石川県金沢市角間町)

E-mail: m.yoshikura@stu.kanazawa-u.ac.jp

現在、我が国における下水道管渠の総延長は約 490,000km となり、膨大なストックを抱えている。また、標準耐用年数である 50 年を経過する下水道管渠の割合は今後加速度的に増加していく。一方で、我が国では地方自治体の財政難や少子高齢化、人口減少により、財源・人材・技術の三つの不足が問題となっている。このような状況下で膨大な下水道管渠に対して、均等に調査を行うことは困難である。したがって、ストックマネジメント計画を策定し、少ない現地調査結果をもとに、現地調査を実施していない下水道管渠の健全性の推定を行い、優先度付けを行った上で、施設の点検・調査、修繕・改築を実施することで、下水道管渠における維持管理の最適化を図る必要がある。そこで、本研究では、より効率的な維持管理手法として、下水道管渠の点検結果を活用し、下水道管渠における不良発生率ランクの推定を行うモデルの構築と評価を行った。

**Key Words:** Machine learning, sewage pipe, Inspection efficiency

## 1. はじめに

現在、我が国の下水道普及率は 2020 年度末で、約 80.1%<sup>1)</sup>、全国下水道管渠の総延長は 2020 年度末で約 490,000km となり、膨大なストックを抱えている。また、2020 年度末において、標準耐用年数 50 年を経過した下水道管渠の延長は約 25,000km (総延長の 5%) であり、10 年後は 82,000km (17%)、20 年後には 19,000km (39%) と今後急速に増加する (図-1)<sup>2)</sup>。下水道管渠は置かれている環境下から交通による荷重やコンクリートの微生物腐食が発生するなど、劣化の要因が時間依存のものにとどまらず、老朽化の進行が他の構造物と比較して早い。さらに、2020 年には下水道管渠の老朽化が起因した道路

陥没事故が 2700 件も発生しており、老朽化は重大な社会問題をも引き起こす<sup>2)</sup>。このような事態を未然に防ぐためにも、点検及び修繕を適切に行う必要がある。現在、

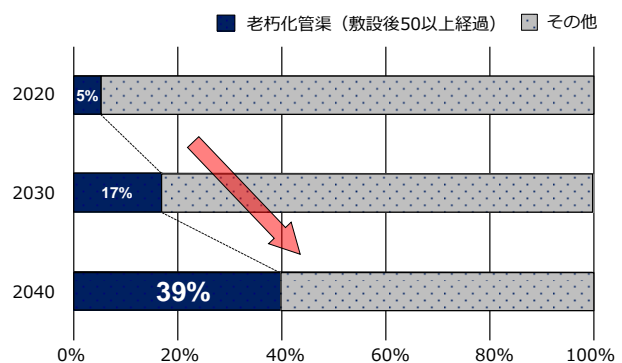


図-1 布設後 50 年以上が経過する下水道管渠の割合<sup>1)</sup>

下水道管渠の点検は目視及びテレビカメラ調査によって行われているが、2018年度の調査延長は目視とテレビカメラによるものとを合わせてわずか 6,686km (総延長の 1.4%)<sup>3)</sup> であり、このペースでは全下水道管渠を一巡調査するのに 72 年もかかり、適切な維持管理が行えない。加えて、我が国では、地方自治体の財政難や少子高齢化、人口減少により、財源・人材・技術の三つの不足が問題となっている。このような状況下で、膨大なストックに対して、均等に調査を行い、さらに修繕を行うことは困難である。

したがって、ストックマネジメント計画を策定し、少ない現地調査結果をもとに、現地調査を実施していない下水道管渠を含めた全管渠の各スパンについて緊急度の推定を行い、優先付けを行った上で、施設の点検・調査、修繕・改築を実施し、下水道管渠における維持管理の最適化を図る必要がある。

本研究では、下水道管渠における不良発生率ランク推定を行うための機械学習モデルの構築を行う。より具体的には、テレビカメラ調査による点検の結果を使用し、「管齢」や「土被り」、「周辺の環境情報」など、調査を行わず取得可能なデータから下水道管渠のうちコンクリート管における各スパン毎の不良発生率ランクを推定するモデルの構築を行う。本研究では、機械学習のうち、アンサンブル学習モデルとして代表的な XGBoost、及び Random Forest の 2 種類のモデルの構築と評価を行った。

## 2. 既往研究の整理と本研究の位置づけ

本章では、本研究の目的に沿い、下水道管渠を対象とした健全度の予測並びに優先度に関するスクリーニングを行っている研究、点検結果を活用し健全度の推定を試みている研究について整理を行い、それを踏まえた上で本研究の新規性について述べる。

### (1) 既往研究の整理

2011年9月、国土交通省国土総合研究所より下水道事業へのアセットマネジメント導入支援を目的とし、管渠の劣化データで構成される国総研 DB が公開された。松宮ら<sup>4)</sup>は、国総研 DB を活用した健全度予測式の推定や改築事業量予測を定量的に行っている。しかしながら、提案された健全度予測式は説明変数として敷設年数のみを使用しており、管種や管径、取り付け管の本数などの差による健全度の違いを考慮することは出来ていない。また、自治体全体の改築事業量予測などには適しているが、具体的にどこにどの程度の損傷を有した下水道管渠が存在するかといった細かい単位での予測をすることは不可能である。

藤生ら<sup>5,6)</sup>は国総研 DB を使用し、DB から取得可能な質的データ及び量的データを差分関数に変換することで、情報量低減を防ぎ、これに対して線形判別分析を適用することで、個々の下水道管渠スパンに対して、緊急度 I, II から構成される劣化の大きい集合である 0 群と緊急度 III, IV から構成される劣化度の小さい集合である 1 群の判別を行っている。

これに対して、筆者ら<sup>8)</sup>は国総研 DB を使用し、深層学習手法の 1 つである、1 次元畳み込みニューラルネットワーク (ID-CNN) による緊急度分類モデルの構築を行っており、分類性能を向上させるためには、下水道管渠が埋設されている環境を再現できるような追加の変数が必要であると結論付けている。

貝戸ら<sup>9)</sup>は、下水道管渠に対する目視点検データを用いて、管渠の劣化過程を混合マルコフ劣化ハザードモデルで表現し、劣化速度の異質性に関するベンチマーク分析を行い、個々の管渠の異質性パラメータを推定し、推定値と管渠の位置情報からデュアル・カーネル密度分析を援用し、劣化速度の空間的劣化異質性分布を評価した。さらに、空間的劣化異質性分布に基づいて、相対的に劣化速度が速い管渠が密集する重点管理区域をスクリーニングするとともに、シミュレーションを通して管渠の改築更新施策を提案した。

### (2) 本研究の位置づけ

本研究の主目的は、テレビカメラ調査等によって取得される管渠内部の損傷状況を説明変数として使用せず、管齢や管径、土被りなど調査を行わず取得可能なデータからスパン単位での不良発生率に関する 4 クラス分類のモデルの構築を行うことである。前節で挙げた既往研究<sup>5,6)</sup>で予測が試みられている緊急度は、本研究で予測を試みる、不良発生率ランクのほか、下水道管のたるみに関するランクと下水道管の腐食に関するランクを総合して評価が行われている。このことから、緊急度の予測を行うためには、複雑な要因が絡み合っているものと考えられる。本研究では、緊急度の評価に用いられる 3 つのランクのうち、不良発生率に着目して予測モデルの構築を行う。また、既往研究で用いられている国総研 DB は下水道管渠の位置が不明であり、管が埋設されている状況を考慮する変数の考慮が困難であった。本研究では、位置情報が付属されている点検結果を用いることで、管の埋設状況に関する変数の考慮を行った。下水道管渠の埋設状況を考慮し、不良発生率のランクを予測している研究は行われておらず、この点が本研究の新規性である。加えて、下水道管渠が埋設構造物であることから、大半の管渠に対して、点検を実施することが出来ていない自治体が存在する。本研究では、点検を必要とせず取得可能な変数により緊急度の分類を行うモデルを作成する。

したがって、財政的に点検を実施することが困難な自治体に対しての技術展開へのハードルが低く、本研究の有用な点であると言える。

### 3. 使用データ及び変数

#### (1) 使用データの概要及び不良発生率ランクについて

本研究では、A 市で実施された下水道管渠の点検結果を用いて、不良発生率ランクの推定を行うモデルを構築する。本データはビデオカメラ調査によって技術者により、健全性の評価が行われたデータであり、点検対象の下水道管居における位置情報と健全性指標が収録されている。その他にも、管齢や管路勾配など管路の諸元情報も存在する。管路の置かれている環境情報は、収録されている位置情報を用いて、GIS により国土数値情報<sup>10)</sup>に存在する環境データと空間結合を行うことで取得を行う。また、本研究で推定を試みる不良発生率ランクは、表-1 に示す通り、管渠における不良の程度によって、A（重度）、B（中度）、C（軽度）、D（健全）の4ランクに診断が行われる。表-1における診断項目のうち、赤文字で示している項目は、発見時に清掃を行い除去できなかった場合にのみ、考慮されるため、診断には主に管の破損やクラックといったコンクリートの損傷が重要されている。

#### 2) 使用変数について

本研究で使用する変数は表-2に示す。本研究では、管の諸元情報、気象条件、土壌・地形条件を変数として考慮した。諸元情報においては、管齢や関係、管渠機能などの設備情報が含まれている。

次に、気象条件については、国土数値情報より、年間降水量と年最深積雪深を取得した。管路には降水や積雪で生じる雨水が流入するため、流入量によって劣化に差が生じると考えられる。また、管路は埋設されているため外気の気温等の影響は小さいと判断し、本研究では、日射量・気温は変数として考慮しないこととした。

最後に、土壌については、埋設されている環境を表現するため、土壌・表面地質区分を用いた、また、管路周辺の人間活動の違いによって流入物が異なり、劣化の度合いがことなると仮定し、代替変数として用途地域区分を使用した。また、管路の置かれている位置や周辺の勾配によって、雨水の流入量が異なると考えられるため、表現するため、管底高さと平均傾斜角を用いた。

#### (3) 不均衡データの整形

本研究で使用するデータの内訳を図-1に示す。本研究

表-1 不良発生率ランクと診断項目

診断項目	ランク(スパン全体で評価)				判定の基準
	重度	中度	軽度	健全	
管の破損	A	B	C	D	A:不良発生率が高い B:不良発生率が中位 C:不良発生率が低い D:不良なし
管のクラック					
管の継手ズレ					
浸入水					
取り付け感の突き出し					
油脂の付着					
樹木根侵入					
モルタル付着					

で用いるデータは不良発生率ランク A のデータ数が非常に少量であり、データ数に大きな偏りが存在する。機械学習モデルの構築において、不均衡なデータを学習に使用する場合に、データ数の多いクラスに分類を行うことで見かけ上の精度が高くなるため、分類において重要な特徴を学習するのではなく、見かけ上の精度が向上するような学習が行われることが懸念される。そのため、本研究では、最もデータ数の少ない緊急度Iのデータ数に、他のクラスのデータ数を合わせるようにデータをランダムサンプリングすることでデータの整形を行った。また、サンプリングの方法によって、分類性能に差が生じる可能性があるため、本研究では 100 回ランダムサンプリングを行い、各データセットに対してモデルの構築を行った。

### 4. 機械学習による不良発生率ランク分類モデル

#### (1) 使用する機械学習モデル

本研究では、緊急度の分類モデルにアンサンブル学習モデルとして代表的な手法である XGBoost<sup>11)</sup>、Random Forest<sup>12)</sup>の2種類のモデルを使用した。

アンサンブル学習とは、複数の弱学習機を組み合わせることでより精度の高いモデルを構築する手法であり、代表的な手法にバギングとブースティングがある。本研究で使用する Random Forest は決定木とバギングを組み合わせた手法であり、XGBoost はブースティングを組み合わせた手法である。本研究では、機械学習のライブラリである Scikit-learn を使用し、前述した各モデルの実装を行った。また、機械学習モデルにおいては、学習に先立って設定を行うハイパーパラメータが存在する。本研究では、バギング及びブースティングに用いる決定木の個数 (n\_estimators)、決定木の深さの最大値 (max\_depth) を考慮した。ハイパーパラメータの調整には、設定した n\_estimators と max\_depth を総当りで検討し、最もモデルの予測精度が高くなるパラメータを決定するグリッドサ

表-2 本研究で使用する変数

項目	説明
管齢(年)	建設してから点検までの期間
管の長さ(m)	上流側マンホールと下流側マンホールの距離
管径(mm)	管の内径
取付管の数(本)	本管に接続されている取付管の本数
管路勾配(%)	管路の勾配
管渠機能	合流式幹線管渠or合流式枝線管渠or汚水幹線管渠or汚水式枝線管渠のいずれか
土壌区分	管路の位置における土壌の区分
表面地質区分	管路の位置における表面地質の区分
年間降水量(mm)	管路の属する1kmメッシュにおける年間降水量
年最深積雪深(m)	管路の属する1kmメッシュにおける年最深積雪深
用途地域区分	管路の位置における用途地域区分
平均傾斜角(度)	管路の属する1kmメッシュにおける平均傾斜角
平均管底高(m)	上流と下流における管底高(東京湾平均海面からの高さ)の平均

一ちを採用した。ハイパーパラメータはそれぞれ、 $n\_estimators=10, 20, 30, 50, 60, 70, 80, 100, 200, 300, 400$  の 11 パターン、 $max\_depth=2, 3, 4, 5, 6, 7, 8, 9, 10$  の 9 パターンに設定した。グリッドサーチの結果、XGBoostは $n\_estimator=9$ 、 $max\_depth=70$ 、RandomForestは $n\_estimator=200$ 、 $max\_depth=2$ の時、最も予測の性能が高くなり、これらを予測モデルのハイパーパラメータとして採用した。

### 5. 分類モデルの評価

本研究では、各機械学習手法に対して、第4章で述べた100回のランダムサンプリングによって作成したデータセットに対して、k分割交差検証法を適用し、モデルの評価を行う。k分割交差検証法では、図-3に示す通り、作成したデータセットをk群に分割し、分割されたデータセットのうち、k-1群を学習、残りの1群をモデルの検証に使用する。この手順をk回繰り返し、各イテレーションにおける評価値の平均値をモデルの評価値とする。本研究では分割数をk=5とした。表-4に本研究で定義する混合行列を示す。本研究では、各モデルにおける分類性能の評価指標として各クラスの再現率(Recall)の平均値であるマクロ平均再現率(Macro-Average Recall)、各クラスの適合率(Precision)の平均値であるマクロ平均適合率再現率(Macro-Average Precision)、各クラスのF値(F-measure)の平均値であるマクロ平均F値(Macro-Average F-measure)の三つの指標を用いる。各クラスにおける再現率、適合率、F値は其々、表-3に示す混合行列によって定義される値より、以下の式(1)から式(3)によって算出される。適合率は分類モデルが対象クラスに分類した数のうち、どれだけ正しく分類されているかを示す指標で

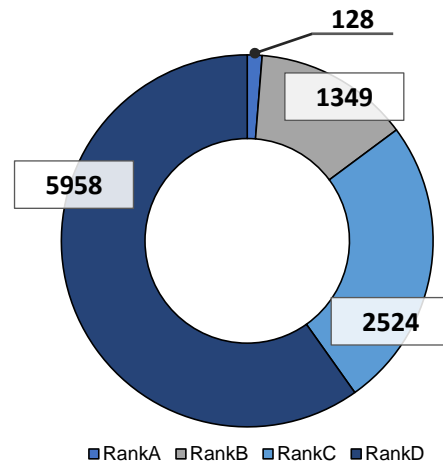


図-2 不良発生率ランク別データ数の内訳

ある。再現率は、対象クラスのうち、実際に分類モデルが正しく分類出来ているかを示す指標である。F値は、再現率と適合率のバランスを示す指標であり、両値の調和平均をとったものである。

$$Recall = \frac{TP}{TP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$F - measure = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{3}$$

各モデルにおける、k分割交差検証の結果を図-4に示す。なお、各分類モデルにおける性能指標はランダムサンプリングによって作成した100セットの均衡データに対して、k分割交差検証法を適用した結果の平均値であ

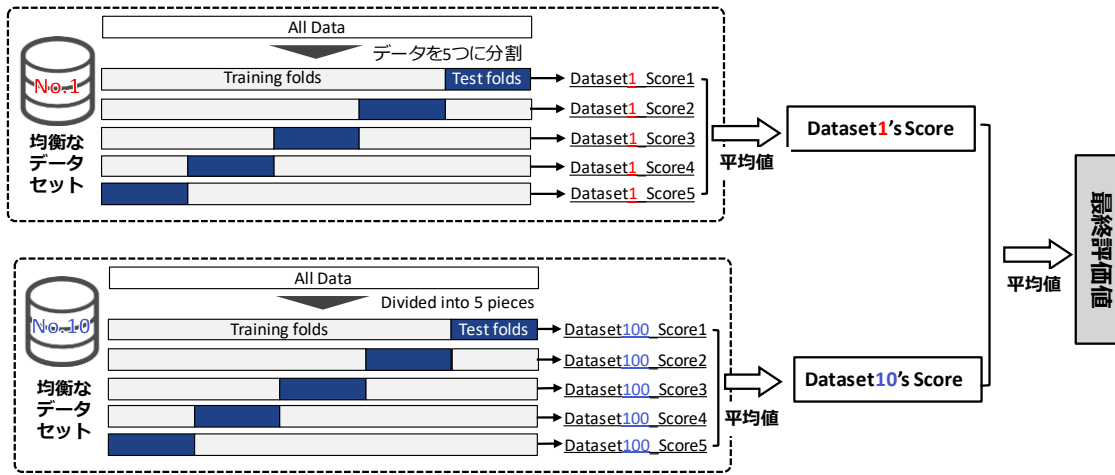


図-3 k 分割交差検証によるモデル評価の概要

表-3 混合行列

		予測クラス	
		該当クラス	その他
正解クラス	該当クラス	TP (True Positive)	FP (False Positive)
	その他	FN (False Negative)	TN (True Negative)

る。各モデルにおけるマクロ平均再現率・マクロ平均適合率・マクロ平均F値はいずれも 0.45 程度であり、結果として、僅かではあるが、XGBoost による分類モデルの性能が良好であった。しかしながら、マクロ平均適合率が 0.45 程度であることは半分程度分類結果が外れていることを示している。高い分類性能が得られなかった原因としては、不良の発生を説明する変数とデータ数の不足が考えられる。本研究では、外力として雨水によるものしか考慮されておらず、交通荷重などは考慮されていない。また、データ数の不足について、本研究では、最もデータ数の少ない RankA のデータ数に合わせるような、ダウンサンプリングを実施したため、モデルの学習に使用したデータ数が非常に少量であった。データの不均衡を除去する際に少ないデータを増加させるアップサンプリングなどの手法も存在するため、こちらも検討する必要がある。

## 6. 本研究のまとめ及び今後の課題

本研究では、今後、老朽化した下水道管渠の割合が加速度的に増加する状況に対して、適切な維持管理を実現するためのアプローチとして、下水道管居の点検結果を用いて、管渠における不良発生率ランクの推定を行う手法

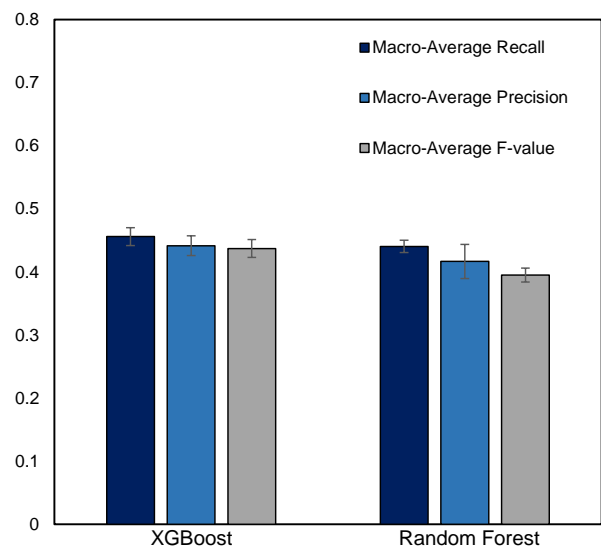


図-4 各モデルにおける分類性能指標

の構築と評価を行った。不良発生率ランクの予測手法には機械学習のうち、アンサンブル学習の代表的な手法である、XGBoost, Random Forest を採用した。また、入力データにおいては、データの性質上、最も損傷の程度が大きい不良発生ランク A のデータが非常に少量であるため、データの偏りが存在し、分類モデルを適切に構築することが難しいという問題が存在した。本研究では、ダウンサンプリングを行うことでこの問題を解決した。ダウンサンプリングによって整形を施したデータセットに対して、k 分割交差検証法を適用し、分類モデルの評価を行った結果、各モデルにおけるマクロ平均再現率・マクロ平均適合率・マクロ平均F値はどれも 0.45 程度であり、結果として、僅かではあるが、XGBoost による分類モデルの性能が最も良好であった。しかしながら、マクロ平均適合率が 0.45 程度であることは半分程度分類結果が外れていることを示している。高い分類性能が得られなかった原因としては、不良の発生を説明する変数とデータ

数の不足が考えられる。本研究では、外力として雨水によるものしか考慮されておらず、交通荷重などは考慮されていない。交通荷重による影響は ETC2.0 プローブデータを用いることで、可能になると考えられ、これは今後の課題である。また、データ数の不足について、本研究では、最もデータ数の少ない RankA のデータ数に合わせるような、ダウンサンプリングを実施したため、モデルの学習に使用したデータ数が非常に少量であった。データの不均衡を除去する際に少ないデータを増加させるアップサンプリングなどの手法も存在するため、こちらも検討する必要がある。

#### 参考文献

- 1) 公益社団法人日本下水道協会 HP：都道府県別の下水処理人口普及率， [https://www.jswa.jp/wp2/wp-content/uploads/2021/09/R03\\_penetration-rate.pdf](https://www.jswa.jp/wp2/wp-content/uploads/2021/09/R03_penetration-rate.pdf), 2023年2月15日閲覧。
- 2) 国土交通省 HP：下水道の維持管理， [https://www.mlit.go.jp/mizukokudo/sewerage/crd\\_sewerage\\_tk\\_000135.html](https://www.mlit.go.jp/mizukokudo/sewerage/crd_sewerage_tk_000135.html), 2023年2月15日閲覧。
- 3) 日本下水道協会：平成 29 年度版下水道統計，電子ファイル「H29\_02-2 管路施設(各種作業).xlsx」, 2023年2月15日閲覧。
- 4) 松宮洋介, 吉田敏章, 深谷渉, 福田康雄：下水道管渠におけるストックマネジメント導入に関する検討調査，「国土交通省国土総合研究所平成 22 年度年報」国土交通省国土総合研究所下水道研究部下水道研究室
- 5) 藤生和也：布設年度と管齢の多重共線性を解消する方法の研究，下水道協会誌，Vol.56, No.682, pp.118-126, 2019.
- 6) 藤生和也, 佐藤克己, 森田弘昭：下水管渠劣化に関し量的データを差分関数に変換して線形判別分析に使用する手法の提案，下水道協会誌，Vol.57, No.696, pp.72-82, 2020.
- 7) 藤生和也, 佐藤克己, 森田弘昭：下水管渠劣化に関し質的及び量的データを差分関数に変換して線形判別分析に使用する手法の提案，下水道協会誌，Vol.58, No.699, pp.84-94, 2021.
- 8) 諏訪太紀, 藤生慎, 森崎裕磨, 福岡知隆, 吉倉麻衣：1D-CNN を用いた下水道管渠における緊急度分類モデルに関する基礎的研究，AI・データサイエンス論文集，3 巻 J2 号，pp.954-961, 2022.
- 9) 貝戸清之, 篠崎秀太, 鎌田敏郎, 前川波奈江, 山中明彦：下水道管渠の空間的劣化異質性に着目した重点管理区域スクリーニングと改築更新施策，土木学会論文集 F4 (建設マネジメント), Vol. 77, No.1, pp.115-134, 2021.
- 10) 国土交通省：国土数値情報ダウンロードサイト， <https://nlftp.mlit.go.jp/index.html>，2022年2月20日閲覧。
- 11) Tianqi, C. and Carlos, G. : XGBoost: A Scalable Tree Boosting System, In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16)*, pp. 785-794, 2016.
- 12) Leo, B.: Random forests, *Machine learning*, Vol.45, No.1, pp. 5-32, 2001.

## ANALYSIS ON ESTIMATION OF DEFECT INCIDENCE RANKS IN SEWAGE PIPES USING MACHINE LEARNING

Taiki SUWA, Makoto FUJII, Yuma MORISAKI, Tomotaka FUKUOKA and Mai Yoshikura

Currently, Japan has a huge stock of sewer culverts with a total length of approximately 490,000 km. In addition, the percentage of sewer culverts with a standard service life of 50 years will increase at an accelerating rate in the future. On the other hand, the financial difficulties of local governments, the declining birthrate, the aging population, and the shrinking population have resulted in a shortage of financial resources, human resources, and technology. Under these circumstances, it is difficult to conduct equal surveys of the vast number of sewer culverts. Therefore, it is necessary to formulate a stock management plan, estimate the soundness of sewage pipes that have not been surveyed in the field, prioritize them, and implement inspection, investigation, repair, and reconstruction of facilities based on the results of few field surveys, in order to optimize the maintenance and management of sewage pipes and culverts. In this study, as a more efficient maintenance management method, we developed and evaluated a model to estimate the defect incidence ranks of sewage pipes utilizing the inspection results .