Condition modelling of railway drainage pipes

Aljafari, Nour; Burrow, Michael; Ghataora, Gurmel; Eskandari Torbaghan, Mehran; Raja, Jamil

DOI:
10.1061/(ASCE)IS.1943-555X.0000708

Condition Modelling of Railway Drainage Pipes

Nour Aljafari [1], Michael Burrow [2], Ph.D., Gurmel Ghataora [3], Ph.D., Mehran Eskandari Torbaghan [4], Ph.D., Jamil Raja [5], Dr.Eng.

[1]Ph.D. candidate, School of Engineering, College of Engineering and Physical Sciences, University of Birmingham, Edgbaston, B15 2TT, UK (Corresponding author). ORCID: https://orcid.org/0000-0002-3498-7363. Email: N.Aljafari@pgr.bham.ac.uk

[2]Reader in Infrastructure Asset Management, School of Engineering, College of Engineering and Physical Sciences, University of Birmingham, Edgbaston, B15 2TT, UK. Email: M.P.Burrow@bham.ac.uk

[3] Honorary Senior Research Fellow, School of Engineering, College of Engineering and Physical Sciences, University of Birmingham, Edgbaston, B15 2TT, UK. Email: G.S.Ghataora@bham.ac.uk

[4]Lecturer in Infrastructure Asset Management, School of Engineering, College of Engineering and Physical Sciences, University of Birmingham, Edgbaston, B15 2TT, UK. ORCID: https://orcid.org/0000-0001-6069-6108. Email: M.Eskandaritorbaghan@bham.ac.uk

[5]Principal Engineer, Network Rail, Elder Gate, Milton Keynes, MK9 1EN, UK. Email: Jamil.Raja@networkrail.co.uk

Abstract

Condition of drainage asset systems can have substantial impact on the structural and operational integrity of railway tracks. It is therefore important to ensure that the various components of the drainage system are well-maintained. To this end, decision makers in the railway industry have been moving towards predictive, risk-informed drainage asset management. The approach aims to optimise the allocation of the limited time and financial resources for maintenance works. To achieve this more research is required to develop predictive condition models for railway drainage assets.

This paper describes the development of data-driven condition prediction models using drainage pipe asset records. The models were tested for both structural and service condition prediction. Nine input factors were considered in the prediction models. Significance of the factors was evaluated using Connection Weight Analysis. Four Machine Learning (ML) algorithms, namely, Neural Networks, Decision Trees, Bagged Trees, and K-Nearest Neighbour, were compared based on their condition prediction performance for pipe drainage assets. The models were developed and tested using field data collected from the UK owner of rail assets, Network Rail. The results demonstrate that Bagged Trees performed best on a balanced dataset with 87% overall accuracy for structural condition prediction and 72% accuracy for service condition prediction. It was found that pipe length, previous condition, years since previous condition and maintenance are the most significant factors in predicting condition.

Introduction and Background

Although the impact of poor drainage on railway track is well known, only a few researchers, e.g. Usman et al. (2017) and Sañudo et al. (2019), have focused on understanding the failure modes of drainage assets and their impact on a track. According to the UK railway drainage standards, performance of drainage system describes the physical ability of the system to carry flow from rainfall and runoff from adjacent areas and groundwater as to not disturb the stable operation of other rail infrastructure. Further, the drainage system should prevent any pollutants from leaving the system in uncontrolled manner. Poor performance of a drainage system causes a myriad of problems including ballast deterioration, subgrade failure (Usman et al., 2015), loss of track geometry degradation, corrosion of track elements, earthwork failures, electric system failures and even derailments (Sañudo et al., 2019). Poor drainage performance is not only caused by incidental defects and collapses of drainage assets, but it also occurs because of natural deterioration experienced by the drainage assets (Ana and Bauwens, 2010). Level of performance for a drainage asset is usually reflected through a condition scoring system, where asset inspectors allocate a condition score depending on the observed condition of the asset. Asset management is a complex process consisting of periodic inspections, routine and emergency maintenance, enhancement, and renewal activities with a direct impact on reliability and safety of railway operation (Fecarotti and Andrews, 2017). These activities need to be arranged or strategised efficiently such that to maintain a given level of service and safety within budget constraints (Fecarotti and Andrews, 2017, Jovanonic and Guler, 2006). Jovanonic and Guler (2006) stated that for efficient management of railway assets, work on two items is crucial: 1) improvement of performance monitoring, 2) having reliable condition assessment and prediction leading to optimised maintenance and rehabilitation planning and resource allocation. For the first, Kovacevic et al. (2016) developed a methodology to asses railway infrastructure using Ground Penetrating Radar (GPR), seismic refraction and drones to evaluate track performance features including ballast fouling, soil water content, slope geometry and drainage condition. While this is important to move away from subjective assessments of railway assets, there is still a gap in predictive condition modelling for railway drainage assets. Wu et al. (2021) utilised the statistical Markov chains for assessment of rail drainage asset service condition at cohort level. However, it is prudent to address structural condition in addition to service condition. Further, through development of a two stage framework for railway track geometry maintenance optimisation, Fecarotti and Andrews (2017) emphasised the importance of individual-asset condition assessment prior to any network-level strategy evaluation.

Currently, there is shift from corrective or reactive maintenance strategies towards proactive and condition-based strategies. Predictive condition models are a step in this direction. Consilvio et al. (2016) developed a risk-based predictive maintenance framework using petri nets for rail assets and utilised deterioration curves for condition assessment as a basis for the point in time when asset requires maintenance. Asset condition evaluation is, therefore, a main part for asset risk-assessment. A study undertaken by Xu and Sinha (2019) incorporated asset performance modelling as an integral part of a risk framework for water mains. Within the railway context, Papathanasiou and Adey (2021) proposed the use of event tree for risk assessment of switches, track sections, and bridges. The study accounted for temperature and traffic as hazards but did not account for flood damage. Oslakovic et al. (2013) addressed climate change impact on railway infrastructure, through identifying incidental failures due to weather events like storms, snow, temperatures and the failure modes like short circuits. However, the study did not account for the condition of the drainage assets, which is a potential cause of the problem. Wang et al. (2021) studied the hazard of floods, earthquakes, and typhoons on railways, using seismic and fluvial flood hazard maps to quantify risk in terms of expected annual damage at a network scale. Koks et al. (2019) also performed a similar study to estimate exposure and risk imposed by natural disasters on a global level. Neither of these studies looked at risk on an individual asset level. Therefore, a drainage asset condition prediction model forms a basis for a risk assessment in railway infrastructure.

Furthermore, with recent development in digital technologies, models which incorporate the use of 'Big Data' for railway asset condition assessment , such as Machine Learning (ML), are indispensable as decision support tools for asset managers (Thaduri et al., 2015, and Mcmahon et al., 2020). Within railway, Consilvio et al. (2020) used ML cluster analysis for a data-driven condition classification of railway earthworks. ML was used as part of a decision support system for strategic asset management using petri nets. The study achieved 22% improvement in distinction between condition classes within a cluster. The application, however, was limited to a two-dimensional cluster analysis utilising track geometry top and alignment standard deviation and soil moisture index for clustering earthworks. Furthermore, Bukhsh et al. (2018) used Random Forests, an ML method, for prediction of railway crossings maintenance needs. It involved a binary classification to determine whether a maintenance intervention was warranted, given the exploratory variables: age, condition state and historic maintenance records of problems. The study achieved 87% accuracy rate and 92% F-score in prediction of maintenance need. Thereon, it would be pertinent to see an extension of the work for prediction of type of maintenance need. Kalathas and Papoutsidakis (2021) used Decision Trees, another ML algorithm, for improving maintenance of rolling stock of traction/braking system. The study achieved 80% accuracy in prediction of the rolling stock maintenance need as either control maintenance or urgent repair. The considered predictors were kilometres of travel, kilometres travelled when malfunction appeared, and total annual malfunctions of equipment. Alawad et al. (2020) used Decision Trees for safety analysis of railway stations. The study achieved 89% accuracy in prediction of accident type using historic accident records containing passenger traits and details of travel timings. Ferreño et al. (2021) also utilised multiple ML algorithms for determination of dynamic stiffness of rail pads. The regression problem used experimental in-service condition of rail pads as predictors, including temperature, frequency, axle load and toe load. The result show that Gradient Boosting ML achieved R^2 of 0.0995. Arshad and Ahmed (2021) achieved 95% overall accuracy using ML for prediction of train delay minutes before start of journey based on historic delays and weather data. Furthermore, Neural Networks, an ML method, was used by Mittal and Rao (2017) for detecting track defects, mapping of switches and signals, and monitoring track health using digital images. The study demonstrates the incorporation of ML as an asset management decision support tool which considers the trade-off between financial limitation and maintenance criticality through detailed evaluation of performance. For e.g., the ability of ML to detect all sun kink defects and switches is crucial as failure to do so could cause derailment. Therefore, the model had to show high retrieval ability, which was achieved at 100% recall of sun kinks (although with very limited dataset) and 93% recall for switches. On the other hand, for loose ballast defects, it was more pertinent to avoid false alarms as they warrant maintenance dispatch, which was achieved at 95% precision classification rate.

The above literature demonstrates the wide applicability of ML methods for various assessment and condition predictions of railway assets, in addition to their seamless integration as part of an asset management decision support tool. However, ML methods have not been used for railway drainage assets. There is a substantial volume of drainage inventory inspection records available in the UK, which could be used for ML condition modelling of individual drainage assets. Therefore, research on predictive condition models could pave the way for 'Big Data' integration in the field of railway drainage asset management.

To this end, this paper proposed a data-driven railway asset structural and service condition prediction model which was verified using drainage pipe asset records. The study targets individual rail assets' condition prediction to help asset managers identify candidate assets that are in critical condition for inspection and maintenance prioritisation, and hence help with better allocation of limited resources within asset management strategies.

Thereon, the following section reviews the existing condition prediction models for urban pipe assets (sewers, drainage and water mains), followed by a description of the adopted methodology for this study within context of rail assets using ML models (Methodology Section). A case study in which four ML condition prediction algorithms were trialled is reported in Case Study Section. The results and comparisons between the four ML methods are presented in Results and Comparisons Section, before drawing the conclusions in the Conclusion Section.

Overview of Existing Condition Models for Drainage Systems

Infrastructure condition models can be divided into three categories: physical, statistical and ML models (Yang, 2004). All three models have been used to model storm water/wastewater pipe condition, focusing mainly on urban/road drainage systems. The applications of the three types of models in literature are reviewed in this Section, mainly Empirical model, Markov models and ML models.

Physical empirical condition models (such as linear and exponential models) are based on physical mechanisms that govern pipe deterioration (Tran et al., 2007). They require relatively large amount of specific type of data to fully capture the deterioration behaviour, but such a data is often unavailable (Ana and Bauwens, 2010). Moreover, failure mechanisms of pipes are very complex as they are the result of interaction between numerous factors with randomly occurring damage propagation (Rajani and Kleiner, 2001, Madanat et al., 1997). An example of these empirical model is the ExtCorr linear model developed by Konig (2005), which gives an estimate of external corrosion of concrete pipes due to environmental factors. The validation of the model was not reported in the article and hence it is not possible to comment on its accuracy. Furthermore, Vollersten and König (2005) stated that in such a model there is high level of uncertainty associated with the 'external conditions'.

Statistical models on the other hand allow for randomness in behaviour of deteriorating pipes. They assume a stochastic behaviour where predictions are based on past distribution of occurrences. The applications of various prediction techniques and models gathered from the literature are summarised in Table 1. The commonly used statistical pipe condition model is the Markov chains model (Wu et al., 2021). Markov model assumes the current state of the asset fully captures all the information regarding the factors that influence future condition and hence, future condition depends solely on current condition (Ana and Bauwens, 2010, Meegoda et al., 2004). Wu et al. (2021) uses Markov Chains for simulations of transition probabilities of railway drainage assets. Moghtadernejad et al. (2021) also uses Markov chains for estimation of deterioration curves of railway supporting structures like retaining walls. Markov model assumes no maintenance will take place so the asset condition cannot improve but can only deteriorate or maintain the current condition. There are two main types of Markov models; Homogeneous Markov chains which are time-independent and non-homogeneous Markov chains in which transition probability from one condition to the other changes with the age of an asset (Ana and Bauwens, 2010). Cohort survival model which is used by Baur and Herz (2002) for sewer condition forecasting is based on statistical regressions methods utilizing survival probability of pipes (Xu and Sinha, 2021). The method estimates the duration that an asset remains in a certain condition until it moves to the next condition, which is presented as a survival curve (Baur and Herz, 2002). Xu and Sinha (2021) used Weibull proportional hazards model (WPHM) as a survival analysis on pipe break records to predict pipes' mean time to failure. Xu and Sinha (2020), Xu and Sinha (2021) described the main issue with survival analyses for pipes is the missing data points in historical records, i.e., left truncation statistical problem. To address this, Xu and Sinha (2021) have used Artificial Neural Networks for imputation of missing data which significantly reduced error from 14% to 2%.

ML models, a subsection of Artificial Intelligence (AI), are data-driven models, which are meant to complement or replace knowledge-driven models when describing a physical phenomenon, such as deterioration of an asset (Solomatine and Ostfeld, 2008). ML simulates the acquisition of knowledge similar to human brain learning which enables continuous improvement of model performance (Liu et al., 2019). They learn the hidden patterns behind the observed data, which enables them to make predictions (Tran, 2007). ML methods are suitable for identifying complex non-linear relationships between input and output data (Lee et al., 2021). They are tolerant of imprecise, subjective, and limited noisy data and can handle both real values and categorical/ordinal input (Tran et al., 2007, Ellis et al., 2008). Unlike statistical methods, ML methods make no assumptions about the data distribution or the required explanatory variables (Asnaashari et al., 2013, Tran, 2007, Solomatine and Ostfeld, 2008). There is a wide range of successful applications of ML methods, mainly using Artificial Neural Networks (ANN), Support Vector Machine (SVM), K-Nearest Neighbours (KNN) and Ensemble Classifiers which include Decision Trees (DT), Bagged Trees (BT), Random Forests (RF), Gradient Boosted Trees (GBT) and AdaBoost. Table 2 summarises these applications and their measures of performance. These studies highlight the suitability and success of ML algorithms as condition prediction models. This study introduces the use of ML for condition prediction of rail drainage assets. Thereon, this study provides a comprehensive approach to develop data-driven prediction models for rail drainage assets.

Parameters Affecting Pipe Condition

Davies et al. (2001a) highlights that there is no single parameter, or even small group of parameters, which stand out as having a particularly strong influence on pipe condition. Rather, it is a process resulting from the interaction of many factors. These factors are categorised into four groups according to Ana and Bauwens (2010), Davies et al. (2001a), Rajani and Kleiner (2001), as follows:

- physical factors: pipe age, pipe shape, pipe size, pipe depth, pipe length, pipe material, pipe slope, pipe type, joint type and material;
- environmental factors: groundwater level infiltration/exfiltration, presence of trees, traffic and surface loadings and soil condition including soil type, pH, density, resistivity, aeration;
- operational factors: sediment level, sewage characteristics of flow, maintenance, and repair strategies. water pressure, surge pressures, summer and winter air and water temperatures, wheel loads, vehicle impact factor and frost load factor; and
- installation factors: installation method, standard of workmanship, laying condition, load factor, coefficient of horizontal stress at rest, coefficient of sliding friction.

Factors considered in statistical and ML applications are summarised in Table 1. Although previous condition and maintenance parameters are missing from existing condition models, Morcous et al. (2002) highlights their influence on infrastructure condition. Therefore, in this study both factors are taken into consideration in the analysis (Inputs Section).

Methodology

Following the result of literature review process, this study was structured to trial the suitability of the following ML methods for predicting the condition of rail drainage assets:

- NN as a popular form of deep learning in recognising patterns between input and outputs (Dawood et al., 2018),
- DT as a basic form of Tree-like models,
- BT as an ensemble classifier, and
- KNN as a simple and powerful form of non-parametric model for large datasets (Parvin et al., 2008).

Additionally, this study addresses some gaps in previous applications of ML to produce robust models with interpretable results. Models which measure performance using statistical tests and metrics (e.g., Coefficient of determination, Mean Square Error) often evaluate the overall performance of a model (Caradot et al., 2018). Since pipe condition assessment is based on a multi-class ranking system in which some classes are of higher criticality, it becomes essential to independently examine the models' performance in each class. Evaluations and comparisons of models based on individual class predictions were undertaken using a set of appropriate measures as discussed in Performance Evaluation section. The study also addresses the issue of black-box or lack of interpretability associated with ML through a significance analysis of input parameters (discussed in Significance of Input Factors Section). Furthermore, it tackles the issue of class-imbalance in Issue of Class Imbalance section. This is often observed in asset networks, where few pipes would be observed in critical condition which can lead to biased predictions (Caradot et al., 2018). To this end, the following steps were followed:

1. Identify potential ML algorithms for condition prediction of rail drainage assets (Description of Classifiers).
2. Define measures for performance comparisons of algorithms (Performance Evaluation).
3. Construct case study data from rail asset database with a focus on pipe assets. (Case Study).
4. Apply data balancing techniques on case study data to tackle issue of class imbalance (Issue of Class Imbalance).
5. Perform case study runs to examine the proposed algorithms (Description of Runs).
6. Perform comparisons using predefined performance measures (Results and Comparisons).
7. Perform significance analysis of input factors (Significance of Input Factors and Discussion of Input Significance).

Description of Classifiers

The following subsection describes the chosen ML algorithms for modelling railway drainage asset condition.

- Neural Networks (NN)

  A basic neural network consists of an input layer, a hidden layer and an output layer (Svozil et al., 1997). Feed Forward Back Propagation technique is used to train the Network, in which the training input data is fed into the network iteratively at different orders to allow for 'Back-Propagation' calibration of assigned weights based on the performance error observed during iteration (Svozil et al., 1997). Studies show that one hidden layer is good enough for approximating most non-linear functions like pipe condition (Tran, 2007, Asnaashari et al., 2013, Svozil et al., 1997, Xu and Sinha, 2021). The number of hidden neurons in the hidden layer requires parameter tuning. For this study 20 hidden neurons were found optimal. Fig. 1 shows the structure of NN.

- Decision Trees (DT)

  They are classification algorithms graphically structured in the form of tree with a root node, decision nodes and terminal nodes known as leaf nodes (Quinlan, 1990). The classification decision is governed by if-then rules which are learned through the training process. Training involves recursive partitioning of the data based on the input variables until the leaf node (class) is reached (Quinlan, 1990). The maximum number of decision splits for the decision tree training was set at 100. Fig. 2 shows the structure of DT.

- Bagged Trees (BT)

  Bagged trees or bootstrap aggregated trees are ensembles of multiple decision trees. The method aims to improve the performance of decision trees by reducing the variance observed in an individual decision tree through bootstrapping the training data into 'N' samples which are then used to train 'N' Trees/Learners (Hastie et al., 2009, Miller et al., 2016). After training, the classification prediction is the averaged class of the 'N' decision trees (Miller et al., 2016). Multiple 'N' values were trialled, N=30 Learners was found enough to achieve an acceptable outcome. Fig. 3 shows the structure of BT.

- K-Nearest Neighbours (KNN)

  It stores all available data as feature vectors and measures how far off the test vector is from each of the stored vectors using Euclidean distance (Zhang, 2016). The test vector is then classified based on the most voted classification of 'K' of its nearest neighbours (Zhang, 2016, Ahmed et al., 2021). After trials of different 'K' values, K value of three nearest neighbours achieved acceptable results. Fig. 4 shows a representation of KNN.

Performance Evaluation

There are two main testing outputs of ML classification algorithms, confusion matrix and Receiver Operator Characteristic curve (ROC), which are described in the following Sections. Derived from these are five measures to be used as bases for performance comparison of individual classes, namely: Overall Accuracy, Recall, Precision, F-score and Area Under the Curve (AUC). For regression problems, most ML applications (Table 2) evaluate overall model performance using Mean Squared Error (MSE) and Root Mean Square Error (RMSE), as in Lee et al. (2018), Wang et al. (2020), Mazari and Rodriguez (2016), Fathi et al. (2019), Ferreño et al. (2021), and Asnaashari et al. (2013). For classification problems, overall accuracy is predominantly used for performance evaluation (Table 2), as in Wu et al. (2015), Mohammadi et al. (2020), Liu et al. (2018), Asadi et al. (2019), and Alawad et al. (2020). However, in multi-class problems it is crucial to obtain a reflection of the model's classification performance for each individual class (Grandini et al., 2020). It is especially important for a model to prove reliability in prediction of critical condition class in order to be accepted to inform asset management decisions. The studies which addressed prediction reliability in the reported parameters are as follows:

- Allah Bukhsh et al. (2019); where F-score was used in addition to overall accuracy,
- Caradot et al. (2018), and Harvey and McBean (2014); where Recall was used,
- Xu and Sinha (2021), and Rokstad and Ugarelli (2015); where AUC was used.
- Kalathas and Papoutsidakis (2021), and Mittal and Rao (2017); where recall and precision were used.

Thereon, the following subsections describes the five adopted measures of performance and their derivations.

Confusion Matrix

A confusion matrix is a visual summary of the trained model's performance. Table 3 shows the configuration of the confusion matrix used in this study. It is customised based on the five possible outcomes for asset condition prediction. Overall Accuracy, Precision and Recall are calculated from the confusion matrix, using Equations (1), (2), and (3), respectively.

Overall accuracy provides a broad overview of the model performance (1). On the other hand, precision and Recall provide insight into individual class performance by dissecting the confusion matrix into prediction rows and target columns (Davis and Goadrich, 2006). Precision is concerned with the exactness or relevance of predictions to an individual class (Raghavan et al., 1989). Meanwhile, Recall is concerned with the class retrieval ability of the model (Raghavan et al., 1989). In condition prediction problems, there is usually a trade-off between precision and Recall (Buckland and Gey, 1994). F-score is useful to reflect the average of both parameters (4).

$$Accuracy = \frac{\sum_{p=1}^{5} TP_p}{\sum_{p=1}^{5} TP_p + \sum_{t=1}^{5}\sum_{p=1}^{5} FP_{p,t}} \quad (1)$$

$$Precision_{class=c} = \frac{TP_c}{TP_c + \sum_{t=1}^{5} FP_{c,t}} \quad (2)$$

$$Recall_{class=c} = \frac{TP_c}{TP_c + \sum_{p=1}^{5} FN_{p,c}} \quad (3)$$

$$F-score_{class=c} = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (4)$$

ROC Curve

ROC shows the True Positive Rate (i.e., Recall) against the False Positive Rate (FPR, i.e., false alarms). FPR calculation from confusion matrix is shown in Equation (5). Fig. 5 shows the ROC curve from one of the NN simulations in this study. As a rule of thumb, the highest point above the diagonal ROC, would be expected to have the least randomness in the classifications (Harvey and McBean, 2014). Thereon a larger AUC indicates higher performance (low randomness). AUC is one of the adopted measures of comparison in this study.

$$FPR_{class=c} = \frac{\sum_{t=1}^{5} FP_{c,t}}{\sum_{p\neq c} TP_p + \sum_{t\neq c}\sum_{p=1}^{5} FP_{p,t}} \quad (5)$$

Significance of Input Factors

In order to tackle the issue of the Black-Box associated with ML, i.e. lack of interpretability of internal work mechanism and identification of causal relationships, significance of factors was investigated in this study (Adadi and Berrada, 2018, Ana and Bauwens, 2010, Allah Bukhsh et al., 2019). Olden et al. (2004) developed a method called Connection Weight Analysis (CWA). This approach achieves Interpretable ML (IML) by analysing individual model components (Molnar et al., 2020). It uses the generated weight to compare the relative significance of the input parameters (Olden et al., 2004). Tran et al. (2009) further developed CWA to make it applicable to multiclass output prediction. Table 4 lists the steps of CWA for an example network of 3 Hidden Neurons (HN) as described by Tran et al. (2009). CWA was used in this study to compare the significance of the input parameters towards the prediction, as demonstrated in the case study (Results and Comparisons Section).

Issue of Class Imbalance

Often in multi-class real training datasets, one or more output classes are represented with more samples than the other classes, creating 'data imbalance' (Guo and Viktor, 2004). The imbalance can have detrimental impact on performance of model (Buda et al., 2018) it causes the trained model to favour the over-represented classes in predictions, while, the minority class may be the critical class in asset condition datasets. In such case, the overall accuracy measure ceases to be a suitable indicator of performance. Imbalance creates the illusion of high performance when in reality the model fails to retrieve the critical-condition observations (Guo and Viktor, 2004).

From a performance analysis point of view, Recall, precision, and AUC are reflective measures of individual class performance and are therefore considered in addition to overall accuracy. However, from a performance enhancement point of view, class imbalance requires attention. Previous pipe condition modelling studies have not yet addressed this problem in the datasets. There are two data-level methods suggested to fix class imbalance, namely: under sampling and over-sampling (Buda et al., 2018). Under-sampling cuts down the size of majority-class data to match the size of the minority class data (Buda et al., 2018). This is an unreasonable option when the difference in representation is substantial which will leave very little data for training the models. Two common over-sampling methods are replication and Synthetic Minority Over-sampling Technique (SMOTe) (Buda et al., 2018). Using replication, as the name implies, the underrepresented classes' samples are replicated repeatedly to reach the size of the over-represented class. Replication can make the model prone to overfitting as it reduces the decision region (Chawla et al., 2002). Whereas SMOTe is an advanced method that avoids over-fitting (Buda et al., 2018, Chawla et al., 2002). It augments the dataset by creating a statistical space between a pre-specified number of neighbouring feature points of the minority class and interpolates new synthetic data points. Both methods can be applied using MATLAB R2020a (Michio, 2020). In this study, the impacts of both the aforementioned over-sampling methods on performance are compared to that of the unbalanced dataset.

Case Study

Data Source

The data used for case study was obtained from the UK Network Rail's (NR) drainage asset database 'DAMBUSTER'. The database records 50% of the rail drainage assets that NR manages in the UK. The retrieved records are structured into 11 individual datasets, covering 11 routes from the main geographical regions in the UK, namely: Wales, Scotland, Wessex, Sussex, Western, Anglia, Kent, East Midlands (EM), London North Eastern (LNE) and London North Western South coast (LNWS), London North Western North coast (LNWN).

The parameters recorded during routine inspections describe the physical structure of the assets, surrounding area, flow, and the condition score. Both service and structural conditions are assessed on a scale of '1' to '5'. Table 5 lists the description of the condition scores according to drainage asset inspection standards by NR. It shows that conditions '3' and '4' are critical and condition '5' shows failure.

Since the focus of the study reported here is on prediction of pipe condition, only asset records identified as pipes were extracted to form the models' training datasets. The records examined spanned from 2010 to 2020; period during which many assets have had multiple inspection entries with a range of conditions. Thus, a temporal dimension to the model inputs has been provided. The parameters extracted from the records were length, type of effluent, topography, material, size, shape, service condition, structural condition, and date of inspection.

Datasets

In order to construct the datasets for analysis, the records obtained were filtered based on:

- Length: only used records of pipes that are equal or under 100m long. Due to the subjective nature of survey inspections, the inspector would survey consecutive pipes of the same observed condition as one continuous length. Therefore, the 100m cap was introduced on recommendation by NR expert on the common pipe lengths.
- Condition state: discarded records with condition class '0', indicating the pipes were inaccessible for survey.

After filtering, two following datasets were extracted from each route database: one was used for service condition prediction (SRVC) and the other for structural condition prediction (STRC). The datasets from each individual route were combined into one large dataset of all routes. Fig. 6 and Fig. 7 show the sample size contribution of each route towards the collective dataset for STRC and SRVC, respectively. Fig. 8 and Fig. 9 show the distribution of samples across the five condition classes for STRC and SRVC, respectively. The data imbalance (discussed in Issue of Class Imbalance Section) is reflected in the distinct unequal split of samples across the five classes. Conditions 1 and 2 are over-represented while critical conditions 3, 4 and 5 are significantly under-represented in both service and structural datasets.

Inputs

After filtering, the six following basic inputs were directly extracted from database; 'Effluent', 'Topography', 'size', 'material' and 'shape'. These are all categorical variables except for length being a scale parameter, as shown in Table 6. Slope, tree count and soil type were considered as significant factors in previous literature (Tran et al., 2006, Tran et al., 2009, Baik et al., 2006, Wirahadikusumah et al., 2001, Davies et al., 2001a, Angarita et al., 2017, Micevski et al., 2002, Rokstad and Ugarelli, 2015). However, data for these factors could not be sourced for this study. Two additional parameters, 'Previous Condition' and 'Years since previous recorded condition' were derived from the data by matching asset ID and used as input.

The temporal inspection data was used as proxy for age of an asset, if age was unknown, by treating the previous condition as an age reference point. Thus, the two factors 'previous condition' and 'years since previous condition' intrinsically account for the change in deterioration rate with age and initial condition. Finally, 'Maintenance' parameter was derived by comparing the derived 'Previous Condition' with the current condition for each record. The observed inconsistency was that for some records, the previous condition was worse than the current condition. In these cases, it was assumed that a 'significant' maintenance intervention has taken place between the two inspections which led to improvement in condition. Another explanation could be the subjectivity in inspector's judgment. Furthermore, this assumption was not to eliminate the possibility that maintenance had taken place previously, in fact it could have taken place while maintaining the same condition. However, this cannot be confirmed due to the lack of relevant maintenance data in the UK. Therefore, for the purpose of this work the assumption was only made to explain discrepancies. 'Maintenance' was added as a dichotomous variable indicating whether maintenance was assumed to have taken place. A total of nine inputs were extracted, which are described in Table 6.

Description of Runs

A total of 12 training runs were performed, one for each combination of classifier (out of four), type of prediction (service or structural) and data balance (unbalanced, balanced with replication, balanced with SMOTe). A comparison of the results is discussed in Results and Comparisons Section. The datasets for NN were split into 60% training, 15% validation and 25% testing following the methodology adopted by Tran (2007), Liu et al. (2018), and Caradot et al. (2018), Moghtadernejad et al. (2021). The validation dataset is required as part of NN for 'Early-Stopping' of training to avoid over-fitting the model while updating the neural weights (Svozil et al., 1997). In DT, BT and KNN the validation dataset is the same one used for testing; therefore, datasets were split into 75% training and 25% hold-out as a test dataset, similar to the ranges used by Fathi et al. (2019), Asnaashari et al. (2013), Assaad and El-adaway (2020), Harvey and McBean (2014), and Xu and Sinha (2021). All the runs were performed using the Neural Net Pattern Recognition and Classification Learner toolboxes on MATLAB R2020a.

Results and Comparisons

For each run, the overall accuracy, precision, Recall and AUC for each class were obtained from the confusion matrix and ROC (Performance Evaluation Section). The following subsections describe performance comparisons between the classifiers and the data-balancing techniques based on the five measures.

Accuracy Comparison

Fig. 10 and Fig. 11 show the comparison of accuracy across the runs for STRC and SRVC, respectively. Structural prediction models generally performed better than service models. The highest achieved accuracy for structural condition was attained by BT, using the balanced dataset, which was 87% while for service condition it was 72%.

Before balancing, all classifiers were seen to perform at about the same accuracy. After balancing, NN and DT exhibited similar behaviour across the runs in which the accuracy drops with introduction of balanced datasets. As the data becomes evenly distributed among classes, each class has an equal contribution towards the calculation of overall accuracy. Therefore, post-balancing accuracy was considered a better representative of the actual prediction ability of the model. The reduced accuracy showed the masked underperformance of NN and DT pre-balancing. Alternatively, KNN and BT showed a similar pattern of accuracy improvement post-balancing. BT showed highest performance both before and after balancing. Hence, it reflects the sophistication of ensemble classifiers as a result of collective voting and reduced variance. Wu et al. (2015) observed superior performance of ensemble classifiers for pipe classifications over single models for pipe defect classification. Perhaps the insignificance of the observed change in accuracy of BT after application of SMOTe implies high resilience of the classifier when subjected to unbalanced datasets.

The high performance of KNN may be explained with it being a ‘lazy classifier’ (Perrizo et al., 2002). While ‘Lazy’ classifiers do not entail a training process, they use a richer hypothesis space compared to ‘eager classifiers’ which make predictions using the trained generalised single-hypothesis model (Shreemali et al., 2021). The inconsistency in post-balancing behaviour between NN/DT and KNN/BT indicates that NN and DT are more sensitive to class imbalance in the training dataset.

Recall Comparison

Recall gives insight into the performance at an individual-class level. Fig. 12 and Fig. 13 demonstrate the change in Recall rates at a class-level for STRC and SRVC, respectively. Results show that for unbalanced runs, Recall rates of classes 3, 4 and 5 are significantly lower than the rates for classes 1 and 2. Thus, highlighting the unreliability of overall accuracy as a performance measure when class imbalance is present. From an asset management point of view, high Recall of pipes in critical condition is crucial. Since the critical classes are the minority group 3, 4 and 5, the imbalance-induced low Recall is emphasised as a deficiency in the model. The highest Recall for STRC, prior to balancing, is in class 1, which makes up to 67% of the data distribution (Fig. 8). Similarly, for SRVC, the highest Recall is in classes 1 and 2, with a smaller gap between the two because the class distribution is more even (Fig. 9).

After balancing, the recall rates for classes 3, 4 and 5 increase drastically. The increase is larger when using replication. However, bearing in mind the identical nature of generated samples using replication compared to the synthetic new data generated with SMOTe, the model is therefore less likely to over-fit with SMOTe . Chawla et al. (2002) demonstrated this conclusion through a DT application by visually comparing the decision regions created with both replication and SMOTe. The synthetic examples cause the classifier to create larger and less specific decision regions, allowing better generalisation of model (Chawla et al., 2002). Hence, SMOTe is likely to produce a more realistic and reliable model than replication.

Conversely, Recall for class 1 reduces after balancing. Although, KNN displays a smaller reduction in post-balancing Recall for condition 1 than BT, both classifiers demonstrate the same net level of post-balancing Recall for classes 1 and 2. KNN and BT's consistently high Recall rates across the classes suggest they are the best performing classifiers.

Precision Comparison

Comparison of precision for STRC and SVRC, respectively shown in Fig. 14 and Fig. 15, indicate that post-balancing precision reduced significantly across the five classes for NN and DT. KNN and BT experienced minor drops and some major improvements, specifically with KNN. The drastic precision improvement in condition 5 for STRC paralleled with a precision drop for SRVC is a discrepancy which may have been caused by the lack of test samples in STRC 5 pre-balancing. Consequently, no frame of reference existed for the change in precision of STRC 5 post-balancing.

The observed reduction in precision and improvement in Recall (Recall Comparison Section) for critical classes 3 and 4, reflects a trade-off dilemma between the two measures. From a risk point of view, the ability of the model to identify more of the critical pipes (high risk of failure) bears more advantage than the disadvantage of false alarms. Whereas, from a financial viewpoint, false alarms incur higher costs associated with unnecessary inspections. Therefore, the models should be enhanced to achieve a balance of high precision and high Recall. This may be best achieved with BT and KNN. The high overall accuracy achieved with these two models further highlights their acceptable performance and appropriateness.

F-score Comparison

F-score provides a good representation of the model performance by combining both precision and Recall (Sokolova and Lapalme, 2009). Fig. 16 and Fig. 17 show F-score for STRC and SRVC conditions, respectively. Across all classifiers, balancing has resulted in a slight reduction in F-score for classes 1 and 2. Whereas it has increased the F-score for classes 3, 4 and 5. It can be therefore concluded that balancing increases individual class predictive performance. The highest performance is observed for KNN and BT. for STRC, F-scores are 78%, 87% and 95% for classes 3, 4 and 5, respectively. For SRVC, F-scores for the three critical classes are 59%, 67% and 71%, respectively.

AUC Comparison

AUC for each class for STRC is shown in Fig. 18 while for SRVC, it is shown in Fig. 19. In STRC models, AUC improved across all classes after balancing. The improvement indicates that balancing reduces the randomness in classification for all classes. This is especially interesting for classes 1 and 2 where the Recall rate reduced after balancing (Recall Comparison Section) while AUC improves. This could possibly indicate that there was some level of randomness in classification with the unbalanced set. Rokstad and Ugarelli (2015) specifies that AUC less and equal to 0.5 is a random classification. Harvey and McBean (2014) proposes 0.8 as threshold for an 'excellent classifier'. In light of this, comparing AUC for the different classifiers after balancing shows KNN and BT to exhibit the highest performance, especially for critical classes 3, 4 and 5.

Discussion of Input Significance

CWA (described in Significance of Input factors Section) was applied on the NN training weights in order to analyse the significance of each of the nine input factors towards the prediction ability. Although NN did not show the highest performance, the model still showed an AUC greater than 0.75 which indicate the predictions by NN are not arbitrary. Therefore a significance analysis of input parameters based on NN Connection weights is valid. Olden et al. (2004) performed a comparison between multiple methods to quantify input significance of NN. These methods included CWA, Garson's Algorithm, Partial Derivatives, Input perturbation, Sensitivity Analysis, Stepwise Addition and Elimination (Olden et al., 2004). It was found that CWA was the only method that consistently identified the correct ranked importance of all predictor variables (Olden et al., 2004). Fig. 20 shows a comparison between the significance of the inputs for both service and structural models. The following six basic inputs; length, effluent type, topography, pipe size, material and shape are evidently more pertinent to structural condition prediction than service condition prediction. This indicates that for service condition, other factors need to be considered like soil type, pipe slope and tree count. Table 7 summarises the observed order of significance using CWA. The rest of the section discusses the findings in light of the literature.

Significance of Input Factors for Serviceability Prediction

Tran et al. (2008) used the Wald test Z-statistic to evaluate the significance of input factors towards service condition prediction using an ordinal regression model. Age, slope, location, climatic condition, Thornthwaite Moisture Index (TMI), and structural condition were found to be significant for serviceability prediction while pipe size, depth, soil type and tree count were found insignificant. Tran et al. (2008) explained that tree count and soil type are inherently reflected through Location and TMI. Similarly, depth and size are correlated factors and are reflected through structural condition, when used as an input. Current study shows low significance of location (presented as topography in case of railway) towards service condition. This indicates that a simple cutting/embankment description is not enough. More nuance is required regarding degree of slope and type of cover. There is a link between topography and slope of earthwork as well as type of cover. Higher slope increases intensity of surface erosion due to surface runoff (Cheng et al., 2017). Eco-hydrologic simulations by Bieger et al. (2015) showed that increase in forested area, reduces surface runoff and erosion. Whereas, deforested hillslopes are more susceptible to erosion (Alavez-Vargas et al., 2021). Moreover, material impacts pipe rate of siltation due to material roughness, thus affecting serviceability condition. None of the other significant factors for serviceability prediction used in this study were identified in the literature. Effluent type and shape of pipe is expected to affect degree of sedimentation for service condition. Length could be significant for service condition when considering flow pressure, which in turn affects level of sedimentation. Longer pipe runs have an increased hydraulic head loss, thus reducing flow velocity and increasing sedimentation.

Significance of Input Factors for Structural Condition Prediction

For structural predictions, comparisons were drawn with Baik et al. (2006), Davies et al. (2001a), Tran et al. (2006), Tran et al. (2009). Tran et al. (2009) applies CWA on the NN model for structural condition of storm water pipes. Tran et al. (2006) used univariate analysis for significance analysis with probabilistic NN while it used stepwise method with the discriminant model.

(Davies et al., 2001b) studied significance of the factors for sewer pipe condition through a stepwise multivariate logistic regression. Baik et al. (2006) identified significance of variables based on the parameters' estimates of the ordered probit model for sewer pipe condition prediction. Table 8 shows a summary of the significance analysis for the factors used in the literature.

Pipe size, length and material are the factors that were used both in the literature and in this study as shown in Table 8. Size of a pipe seems to be significant across all the previous studies (Micevski et al., 2002, Baik et al., 2006, Davies et al., 2001a). Despite the consensus around the significance, there are contradicting views around the nature of impact the pipe size has on pipe condition. Micevski et al. (2002) found that smaller pipes experience greater deterioration due to underestimation of design requirements in terms of loadings and cover. Angarita et al. (2017) made a similar observation, particularly for pipes smaller than 500 mm. Baik et al. (2006) however, explained that larger pipes are more likely to have higher rates of deterioration. Davies et al. (2001a) provided two contradicting views: one stated that larger sewers were more structurally sound, the other provided evidence that risk of structural failure is higher for larger pipes due to installation difficulties.

Pipe length, which was considered by Baik et al. (2006) and Davies et al. (2001a) showed high significance for condition prediction, similar to the current analysis. With regards to the nature of impact of pipe length on pipe condition, Baik et al. (2006) suggests that shorter pipe section runs induce faster condition deterioration due to increased number of connections. Similarly, Davies et al. (2001b) claimed that sewage pipes with length more than 1.5 m are less likely to be in failing condition, as shorter individual sewer pipe runs indicate greater number of joints per unit length of sewer. This was linked with infiltration of soils and sedimentation being more likely through joints/connections. On the other hand, Davies et al. (2001a) described that the higher length to diameter ration causes increased structural bending stress and hence, worse structural condition. The two opposite views on nature of impact on hydraulic and structural conditions make it difficult to make a recommendation on the optimal pipe length. However, they indicate that there is trade-off between the two during design process.

Conversely, pipe material was found insignificant by both Baik et al. (2006) and Davies et al. (2001a). However, Micevski et al. (2002) and Wirahadikusumah et al. (2001) both used material as a grouping criteria when applying the Markov chains for storm water-pipe prediction and showed that it was a statistically significant factor in pipe condition. Angarita et al. (2017) also showed that material is a significant factor. The findings of this study, however, confirm the significance of material as an input factor. The type of pipe material affects the experienced rate of structural corrosion.

Effluent type was found to be significant in this study. Davies et al. (2001a)' includes it as 'sewer purpose/use', with the following three categories: foul water, surface water and combined. It was shown to be among the top 10 significant factors, from the 18 tested factors. Angarita et al. (2017) also confirmed its significance through a linear regression study of explanatory variables.

Topography was not used in any other studies, except for Davies et al. (2001a) and Tran et al. (2009). Both showed location to be a significant factor. These studies considered storm water and wastewater systems in urban and suburban areas, in which location depended on the use of the area, mainly reflecting the impact of traffic load and type of cover on structural condition of pipes. In the scope of railway drainage, 'topography' could be viewed as a derivative of 'location'. Pipes in a cutting are more likely to experience tree-root penetration compared to embankments. Thus, explaining the observed significance of the parameter towards structural condition in this study.

Shape was found to be of high significance in this study, possibly because it influences the load distribution or because it is indicative of other significant factors that are unaccounted for, like installation and connection workmanship.

Factors that were found significant in other studies but are not included in the current study are slope, soil type and tree count. Slope was found to be significant in Probabilistic Neural Networks (PNN) (Tran et al., 2006) and ordered probit model used with Markov chains (Baik et al., 2006). Baik et al. (2006) explained that steeper slopes are associated with higher flows and lower stability, leading to increased rate of deterioration and consequently condition reduction.

Soil type has conflicting views around it. Tran et al. (2009) identified soil type as significant towards NN but insignificant towards the ordered probit model. Tran et al. (2006) also identified soil type as insignificant towards both PNN and discriminant model. Whereas, Davies et al. (2001a) listed two soil-related factors in the top ten most significant factors in pipe deterioration; soil fracture potential and soil corrosivity. Micevski et al. (2002) also confirmed the significance of soil type. Similarly, Wirahadikusumah et al. (2001) stressed the importance of using soil type in grouping cohorts when using the Markov chains for pipe condition modelling.

Tree count only showed significance towards NN model in Tran et al. (2009) but not towards the ordered probit model, discriminant model or PNN (Tran et al., 2006, Tran et al., 2009). It is an interesting finding since tree root interference is a major cause of structural damage in pipes (Davies et al., 2001a, Rokstad and Ugarelli, 2015). Angarita et al. (2017) also reflected the high impact of tree roots through the use of 'cover type' as a significant explanatory factor in pipe condition models. The issue is even more prevalent in railway drainage, considering the substantial portion of earthworks using green cover for stabilisation. Tree count is therefore a parameter worth sourcing.

Age is a matter of controversy, although it is a central factor in any Markov chain model (Baik et al., 2006, Micevski et al., 2002, Wirahadikusumah et al., 2001, Tran et al., 2010, Le Gat, 2008, Kannapiran et al., 2008), it was determined insignificant in Tran et al. (2006), Tran et al. (2009) and Davies et al. (2001a). Baik et al. (2006) also noted that age parameter becomes less significant when pipe is in worse condition. Tran (2007) suggested that age is more a reference point for monitoring and that pipe condition is affected by a combination of many factors and damaging events which are independent of pipe age.

Based on the literature review, it seems that tree count, pipe slope, and soil type as well as age, despite the contradicting views, are parameters which should be investigate for future development of the model.

Hydraulic or serviceability condition as an input parameter for structural condition prediction was found significant by Tran et al. (2006) and Tran et al. (2009). Davies et al. (2001a) also reflected on the significance of hydraulic condition, although indirectly, through the use of 'debris presence' which observed to be a significant factor (Table 8). Micevski et al. (2002) on the other hand found hydraulic condition to have insignificant influence on structural condition. However, they still explain a need to consider both structural and service condition for maintenance strategies.

The importance of hydraulic condition and structural condition as inputs for structural prediction and service prediction, respectively, can be explained with the bidirectional correlation between the two factors. On the one hand, more structural damage (cracks) allows for more intrusion of material that hinders the flow of water in the pipe. Furthermore, deteriorating hydraulic condition may cause frequent surcharges due to blockages and intrusion which, in turn, cause a need for rehabilitation or replacement (Micevski et al., 2002). Despite their importance, the reason for not including these factors as inputs is that the primary objective of the study was to predict the structural and service condition of drainage, therefore the developed model treats them as output rather than input parameters.

Finally, with respect to the derived parameters, Previous condition shows highest significance towards both the structural and service model. 'Years' is the second-highest significant factor for STRC. It shows lower significance towards SRVC. Maintenance shows high significance towards both models. It shows higher significance towards service prediction which is a rational result since regular maintenance mainly entails serviceability-related works, like rodding and jetting to improve hydraulic behaviour of pipes.

Conclusion

In this paper a comparison between the performance of NN, DT, BT and KNN for condition prediction of railway drainage pipes was conducted using nine input factors. Data was obtained from the UK NR drainage and lineside asset database. The results show that for structural condition prediction BT performs the best on balanced dataset with 87% overall accuracy. On unbalanced dataset NN, BT, DT showed similar performance within 70%-80% accuracy range. Similarly, for service condition, BT performs the best with 72% accuracy on balanced dataset.

The input factors (Shape, Material, Length, Size, Effluent and Topography, Previous condition, Years since previous condition and Maintenance) were found to be more relevant to structural condition prediction. This was confirmed with the CWA which also showed that length, effluent type, topography, pipe size, and shape were not significant in predicting service condition. The derived parameters 'previous condition', 'maintenance' and 'years' show the highest significance towards prediction of both service and structural conditions.

The impact of data balancing was analysed by investigating Recall and precision of individual condition classes. Balancing improved the Recall rate of condition classes 3, 4 and 5 while reducing the Recall rate of classes 1 and 2. Precision tended to drop across all the classes for all the classifiers except for KNN, for which it showed an improvement. The F-score is used as a mean measure of Recall and precision. When balancing using SMOTe, F-score has improved significantly for critical classes, but it reduces slightly for classes 1 and 2. When improving the model, high F-score and high AUC should be aimed for the critical class as they characterize a reliable classifier performance.

Contributions:

This study has had the following contributions to the knowledge in the field of railway asset management: 1) It addresses predictive railway drainage condition assessment at an asset-level, making use of extensive historic inspection records in data-driven ML models while addressing some of the problems traditionally faced when using ML methods, e.g. data imbalance and black box. 2) It provides a high-performance predictive decision-support tool for the railway drainage asset managers to identify locations of critical assets using four ML algorithms. Thus, allowing optimisation of time and money allocations for the maintenance and inspection works. 3) It paves the way for data-driven risk-assessment for quantification of criticality of individual assets. 4) Further, it provides a detailed descriptive analysis of parameter importance for condition prediction. Hence, making recommendations to railway drainage asset managers about parameters to be included in the inspection surveys. 5) Finally, this work forms the stepping-stone for the future of efficient data-driven predictive asset management as opposed to reactive asset management. Adoption of ML algorithms allows for seamless integration of 'Big Data' and IoT into railway drainage asset management.

Limitations

The following are some points of consideration for future extension of this study:

1) In this study the only maintenance record involved was the assumption that maintenance took place when condition was improved. However, maintenance could have taken place at other points in time where condition did not necessarily improve, or improvement was not reported. Therefore, investigating the significance of maintenance intervention on pipe's condition remains a prospect for further research. This requires accurate historical data regarding maintenance intervention.
2) Handling dataset imbalance is normally concerned with the target output parameters which was the scope of this work. Further extension of the work could be on investigating the imbalance in individual input parameters. It could enhance prediction performance by expanding the dataset and exposing the classifier to more variations in input vectors. Higher variation and representation of different input parameters could result in a better analysis of significance of the input parameters.
3) CWA was the chosen method to investigate input significance based on NN connection weights as an attempt to tackle the black-box issue associated with NN. Other methods like stepwise elimination/ addition analysis or sensitivity analysis could be applied to other models to investigate parameter significance.

Data availability statement

All data, models or code generated or used during the study was provided by a third party. Direct requests for these materials may be made to the provider as indicated in the acknowledgments.

Acknowledgement

The authors would like to thank Network Rail for sponsoring this research and in particular Ms Mona Sihota Technical Head – Drainage and Off-Track (Network Rail).

References

ADADI, A. & BERRADA, M. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access,* 6**,** 52138-52160.

AHMED, M., KHALEF, R., ALI, G. & EL-ADAWAY, I. 2021. Evaluating Deterioration of Tunnels Using Computational Machine Learning Algorithms. *Journal of Construction Engineering and Management,* 147**,** 1-15.

ALAVEZ-VARGAS, M., BIRKEL, C., CORONA, A. & BREÑA-NARANJO, J. A. 2021. Land cover change induced sediment transport behaviour in a large tropical Mexican catchment. *Hydrological Sciences Journal,* 66**,** 1069-1082.

ALAWAD, H., KAEWUNRUEN, S. & AN, M. 2020. Learning From Accidents: Machine Learning for Safety at Railway Stations. *IEEE Access,* 8, 633-648.

ALLAH BUKHSH, Z., SAEED, A., STIPANOVIC, I. & DOREE, A. G. 2019. Predictive maintenance using tree-based classification techniques: A case of railway switches. *Transportation Research Part C: Emerging Technologies,* 101, 35-54.

ALTARABSHEH, A., KANDIL, A. & VENTRESCA, M. 2016. Multi-Objective Optimization Algorithm for Sewer Network Rehabilitation Using Life Cycle Cost Analysis and Semi-Markov Deterioration Models. *Construction Research Congress 2016: Old and New Construction Technologies Converge in Historic San Juan*, 2089-2099.

ANA, E. V. & BAUWENS, W. 2010. Modeling the structural deterioration of urban drainage pipes: the state-of-the-art in statistical methods. *Urban Water Journal,* 7, 47-59.

ANGARITA, H., NIÑO, P., VARGAS, D., HERNÁNDEZ, N. & TORRES, A. 2017. Identifying explanatory variables of structural state for optimum asset management of urban drainage networks: a pilot study for the city of Bogota. *Ingeniería e Investigación,* 37, 6-16.

ARSHAD, M. & AHMED, M. 2021. Train Delay Estimation in Indian Railways by Including Weather Factors Through Machine Learning Techniques. *Recent Advances in Computer Science and Communications* 14, 1300-1307.

ASADI, P., GINDY, M. & ALVAREZ, M. 2019. A Machine Learning Based Approach for Automatic Rebar Detection and Quantification of Deterioration in Concrete Bridge Deck Ground Penetrating Radar B-scan Images. *KSCE Journal of Civil Engineering,* 23, 2618-2627.

ASNAASHARI, A., MCBEAN, E. A., GHARABAGHI, B. & TUTT, D. 2013. Forecasting watermain failure using artificial neural network modelling. *Canadian Water Resources Journal,* 38, 24-33.

ASSAAD, R. & EL-ADAWAY, I. H. 2020. Bridge Infrastructure Asset Management System: Comparative Computational Machine Learning Approach for Evaluating and Predicting Deck Deterioration Conditions. *Journal of Infrastructure Systems,* 26, 04020032.

BAIK, H. S., JEONG, H. S. & ABRAHAM, D. M. 2006. Estimating transition probabilities in Markov chain-based deterioration models for management of wastewater systems. *Journal of Water Resources Planning and Management-Asce,* 132, 15-24.

BAUR, R. & HERZ, R. 2002. Selective inspection planning with ageing forecast for sewer types. *Water Science and Technology,* 46, 389-396.

BIEGER, K., HÖRMANN, G. & FOHRER, N. 2015. The impact of land use change in the Xiangxi Catchment (China) on water balance and sediment transport. *Regional Environmental Change,* 15, 485-498.

BUCKLAND, M. & GEY, F. 1994. The relationship between Recall and Precision. *Journal of the American Society for Information Science,* 45, 12-19.

BUDA, M., MAKI, A. & MAZUROWSKI, M. A. 2018. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks,* 106, 249-259.

BUKHSH, Z., SAEED, A. & STIPANOVIC, I. 2018. A machine learning approach for maintenance prediction of railway assets. *Transport Research Arena,* 2018.

CARADOT, N., RIECHEL, M., FESNEAU, M., HERNANDEZ, N., TORRES, A., SONNENBERG, H., ECKERT, E., LENGEMANN, N., WASCHNEWSKI, J. & ROUAULT, P. 2018. Practical benchmarking of statistical and machine learning models for predicting the condition of sewer pipes in Berlin, Germany. *Journal of Hydroinformatics,* 20, 1131-1147.

CHAWLA, N. V., BOWYER, K. W., HALL, L. O. & KEGELMEYER, W. P. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research,* 16, 321-357.

CHENG, N. N., HE, H., YANG, S., LU, Y. J. & JING, Z. W. 2017. Impacts of topography on sediment discharge in Loess Plateau, China. *Quaternary International,* 440.

CONSILVIO, A., FEBBRARO, A. D. & SACCO, N. Stochastic scheduling approach for predictive risk-based railway maintenance. 2016 IEEE International Conference on Intelligent Rail Transportation (ICIRT), 23-25 Aug. 2016 2016. 197-203.

CONSILVIO, A., SOLÍS-HERNÁNDEZ, J., JIMÉNEZ-REDONDO, N., SANETTI, P., PAPA, F. & MINGOLARRA-GARAIZAR, I. 2020. On Applying Machine Learning and Simulative Approaches to Railway Asset Management: The Earthworks and Track Circuits Case Studies. 12**,** 2544.

DAVIES, J. P., CLARKE, B. A., WHITER, J. T. & CUNNINGHAM, R. J. 2001a. Factors influencing the structural deterioration and collapse of rigid sewer pipes. *Urban Water,* 3**,** 73-89.

DAVIES, J. P., CLARKE, B. A., WHITER, J. T., CUNNINGHAM, R. J. & LEIDI, A. 2001b. The structural condition of rigid sewer pipes: a statistical investigation. *Urban Water,* 3**,** 277-286.

DAVIS, J. & GOADRICH, M. 2006. The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd international conference on Machine learning.* Pittsburgh, Pennsylvania, USA: Association for Computing Machinery.

DAWOOD, T., ZHU, Z. & ZAYED, T. 2018. Computer Vision–Based Model for Moisture Marks Detection and Recognition in Subway Networks. *Journal of Computing in Civil Engineering,* 32**,** 04017079.

ELLIS, E., YU, H.-S., MCDOWELL, G., DAWSON, A. & THOM, N. 2008. *Advances in transportation geotechnics,* Boca Raton, Boca Raton : CRC Press.

FATHI, A., MAZARI, M., SAGHAFI, M., HOSSEINI, A. & KUMAR, S. 2019. Parametric Study of Pavement Deterioration Using Machine Learning Algorithms. *Airfield and Highway Pavements 2019.*

FECAROTTI, C. & ANDREWS, J. 2017. Optimising strategy selection for the management of railway assets.

FERREÑO, D., SAINZ-AJA, J. A., CARRASCAL, I. A., CUARTAS, M., POMBO, J., CASADO, J. A. & DIEGO, S. 2021. Prediction of mechanical properties of rail pads under in-service conditions through machine learning algorithms. *Advances in Engineering Software,* 151**,** 102927.

GRANDINI, M., BAGLI, E. & VISANI, G. 2020. Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*.

GUO, H. & VIKTOR, H. L. 2004. Learning from imbalanced data sets with boosting and data generation: the DataBoost-IM approach. *SIGKDD Explor. Newsl.,* 6**,** 30–39.

HARVEY, R. R. & MCBEAN, E. A. 2014. Predicting the structural condition of individual sanitary sewer pipes with random forests. *Canadian Journal of Civil Engineering,* 41**,** 294-303.

HASTIE, T., TIBSHIRANI, R. & FRIEDMAN, J. 2009. Random Forests. *In:* HASTIE, T., TIBSHIRANI, R. & FRIEDMAN, J. (eds.) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* New York, NY: Springer New York.

JOVANONIC, S. & GULER, H. Railway infrastructure assets management by using appropriate systems. 7th Int. Congress on Advances in Civil Engineering, Yildiz Technical Univ., Istanbul, Turkey, 2006.

KALATHAS, I. & PAPOUTSIDAKIS, M. 2021. Predictive Maintenance Using Machine Learning and Data Mining: A Pioneer Method Implemented to Greek Railways. *Designs,* 5**,** 5.

KANNAPIRAN, A., CHANAN, A., SINGH, G., TAMBOSIS, P., JEYAKUMARAN, J. & KANDASAMY, J. 2008. Strategic asset management planning of stormwater drainage systems. *Water Practice and Technology,* 3**,** wpt2008065.

KLEINER, Y., SADIQ, R. & RAJANI, B. 2006. Modelling the deterioration of buried infrastructure as a fuzzy Markov process. *Journal of Water Supply: Research and Technology - Aqua,* 55**,** 67-80.

KOKS, E. E., ROZENBERG, J., ZORN, C., TARIVERDI, M., VOUSDOUKAS, M., FRASER, S. A., HALL, J. W. & HALLEGATTE, S. 2019. A global multi-hazard risk analysis of road and railway infrastructure assets. *Nature Communications,* 10**,** 2677.

KONIG, A. 2005. CARE-S WP2 External corrosion model description. *SINTEF Technology and Society, Trondheim, Norway*.

KOVACEVIC, M., GAVIN, K., OSLAKOVIC, I. S. & BACIC, M. 2016. A new methodology for assessment of railway infrastructure condition. *Transportation research procedia,* 14, 1930-1939.

LE GAT, Y. 2008. Modelling the deterioration process of drainage pipelines. *Urban Water Journal,* 5, 97-106.

LEE, J., PARK, C. Y., BAEK, S., HAN, S. H. & YUN, S. 2021. Risk-Based Prioritization of Sewer Pipe Inspection from Infrastructure Asset Management Perspective. *Sustainability,* 13, 7213.

LEE, J. S., HWANG, S. H., CHOI, I. Y. & KIM, I. K. 2018. Prediction of Track Deterioration Using Maintenance Data and Machine Learning Schemes. *Journal of Transportation Engineering, Part A: Systems,* 144, 04018045.

LIU, K., WANG, M., CAO, Y., ZHU, W. & YANG, G. 2018. Susceptibility of existing and planned Chinese railway system subjected to rainfall-induced multi-hazards. *Transportation Research Part A: Policy and Practice,* 117, 214-226.

LIU, Y., MA, X., LI, Y., TIE, Y., ZHANG, Y. & GAO, J. 2019. Water Pipeline Leakage Detection Based on Machine Learning and Wireless Sensor Networks. *Sensors,* 19, 5086.

MADANAT, S., KARLAFTIS, M. G. & MCCARTHY, P. S. 1997. Probabilistic infrastructure deterioration models with panel data. *Journal of Infrastructure Systems,* 3, 4.

MAZARI, M. & RODRIGUEZ, D. D. 2016. Prediction of pavement roughness using a hybrid gene expression programming-neural network technique. *Journal of Traffic and Transportation Engineering (English Edition),* 3, 448-455.

MCMAHON, P., ZHANG, T. & DWIGHT, R. 2020. Requirements for Big Data Adoption for Railway Asset Management. *IEEE Access,* 8, 15543-15564.

MEEGODA, J. N., JULIANO, T. M., AYOOLA, M. G. & DHAR, S. 2004. A Methodology to Predict the Remaining Service Life of CSCPs. *International Conference on Case Histories in Geotechnical Engineering.*

MICEVSKI, T., KUCZERA, G. & COOMBES, P. 2002. Markov Model for Storm Water Pipe Deterioration. *Journal of Infrastructure Systems,* 8, 49-56.

MICHIO. 2020. *michio (2020). Oversampling Imbalanced Data: SMOTE related algorithms (https://github.com/minoue-xx/Oversampling-Imbalanced-Data/releases/tag/1.0.1), GitHub. Retrieved October 9, 2020.* [Online]. [Accessed].

MILLER, P. J., LUBKE, G. H., MCARTOR, D. B. & BERGEMAN, C. S. 2016. Finding structure in data using multivariate tree boosting. *Psychological methods,* 21, 583-602.

MITTAL, S. & RAO, D. 2017. Vision based railway track monitoring using deep learning. *arXiv preprint arXiv:.06423*.

MOGHTADERNEJAD, S., HUBER, G., HACKL, J. & ADEY, B. T. 2021. Data-driven estimation of deterioration curves: a railway supporting structures case study. *Infrastructure Asset Management*, 1-15.

MOHAMMADI, M. M., NAJAFI, M., SALEHABADI, N., SERAJIANTEHRANI, R. & KAUSHAL, V. 2020. Predicting Condition of Sanitary Sewer Pipes with Gradient Boosting Tree. *Pipelines 2020.*

MOLNAR, C., CASALICCHIO, G. & BISCHL, B. Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges. *In:* KOPRINSKA, I., KAMP, M., APPICE, A., LOGLISCI, C., ANTONIE, L., ZIMMERMANN, A., GUIDOTTI, R., ÖZGÖBEK, Ö., RIBEIRO, R. P., GAVALDÀ, R., GAMA, J., ADILOVA, L., KRISHNAMURTHY, Y., FERREIRA, P. M., MALERBA, D., MEDEIROS, I., CECI, M., MANCO, G., MASCIARI, E., RAS, Z. W., CHRISTEN, P., NTOUTSI, E., SCHUBERT, E., ZIMEK, A., MONREALE, A., BIECEK, P., RINZIVILLO, S., KILLE, B., LOMMATZSCH, A.

& GULLA, J. A., eds. ECML PKDD 2020 Workshops, 2020// 2020 Cham. Springer International Publishing, 417-431.

MORCOUS, G., RIVARD, H. & HANNA, A. 2002. Case-based reasoning system for modeling infrastructure deterioration. *Journal of Computer Civil Engineering,* 16**,** 104.

OLDEN, J. D., JOY, M. K. & DEATH, R. G. 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling,* 178**,** 389-397.

OSLAKOVIC, I. S., TER MAAT, H., HARTMANN, A. & DEWULF, G. Risk assessment of climate change impacts on railway infrastructure. EPOC 2013 Conference, 2013.

PAPATHANASIOU, N. & ADEY, B. T. 2021. Making comparable risk estimates for railway assets of different types. 8**,** 53-74.

PARVIN, H., ALIZADEH, H. & MINAEI, B. 2008. MKNN: Modified k-nearest neighbor. *Lecture Notes in Engineering and Computer Science,* 2173.

PERRIZO, W., DING, Q. & DENTON, A. Lazy Classifiers Using P-trees. CAINE, 2002. Citeseer, 176-179.

PIRYONESI, S. M. & EL-DIRABY, T. E. 2021. Using Machine Learning to Examine Impact of Type of Performance Indicator on Flexible Pavement Deterioration Modeling. *Journal of Infrastructure Systems,* 27**,** 04021005.

QUINLAN, J. R. 1990. Decision trees and decision-making. *IEEE Transactions on Systems, Man, and Cybernetics,* 20**,** 339-346.

RAGHAVAN, V., BOLLMANN, P. & JUNG, G. S. 1989. A critical investigation of recall and precision as measures of retrieval system performance. *ACM Trans. Inf. Syst.,* 7**,** 205–229.

RAJANI, B. & KLEINER, Y. 2001. Comprehensive review of structural deterioration of water mains: physically based models. *Urban Water,* 3**,** 151-164.

ROKSTAD, M. M. & UGARELLI, R. M. 2015. Evaluating the role of deterioration models for condition assessment of sewers. *Journal of Hydroinformatics,* 17**,** 789-804.

SAÑUDO, R., MIRANDA, M., GARCÍA, C. & GARCÍA-SANCHEZ, D. 2019. Drainage in railways. *Construction and Building Materials,* 210**,** 391-412.

SHREEMALI, J., MALVIYA, L., PALIWAL, P., CHAKRABARTI, P., PODDAR, S., JINDAL, B. & CHAUBISA, H. 2021. Comparing performance of multiple classifiers for regression and classification machine learning problems using structured datasets. *Materials Today: Proceedings*.

SOKOLOVA, M. & LAPALME, G. 2009. A systematic analysis of performance measures for classification tasks. *Information Processing & Management,* 45**,** 427-437.

SOLOMATINE, D. P. & OSTFELD, A. 2008. Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics,* 10**,** 3-22.

SVOZIL, D., KVASNICKA, V. & POSPICHAL, J. Í. 1997. Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems,* 39**,** 43-62.

THADURI, A., GALAR, D. & KUMAR, U. 2015. Railway assets: A potential domain for big data analytics. *Procedia Computer Science,* 53**,** 457-467.

TRAN, NG, A. W. M., PERERA, B. J. C., BURN, S. & DAVIS, P. 2006. Application of probabilistic neural networks in modelling structural deterioration of stormwater pipes. *Urban Water Journal,* 3**,** 175-184.

TRAN, PERERA, B. J. C. & NG, A. W. M. 2009. Comparison of Structural Deterioration Models for Stormwater Drainage Pipes. *Computer-Aided Civil and Infrastructure Engineering,* 24**,** 145-156.

TRAN, D. H., NG, A. W. M., MCMANUS, K. J. & BURN, S. 2008. Prediction models for serviceability deterioration of stormwater pipes. *Structure and Infrastructure Engineering,* 4**,** 287-295.

TRAN, D. H., NG, A. W. M. & PERERA, B. J. C. 2007. Neural networks deterioration models for serviceability condition of buried stormwater pipes. *Engineering Applications of Artificial Intelligence,* 20**,** 1144-1151.

TRAN, H. D. 2007. *INVESTIGATION OF DETERIORATION MODELS FOR STORMWATER PIPE SYSTEMS* Doctor of Philosophy Degree, Victoria University.

TRAN, H. D., PERERA, B. J. C. & NG, A. W. M. 2010. Markov and Neural Network Models for Prediction of Structural Deterioration of Storm-Water Pipe Assets. *Journal of Infrastructure Systems,* 16**,** 167-171.

UGARELLI, R. M., SELSETH, I., LE GAT, Y., ROSTUM, J. & KROGH, A. H. 2013. Wastewater pipes in Oslo: from condition monitoring to rehabilitation planning. *Water Practice and Technology,* 8**,** 487-494.

USMAN, K., BURROW, M. & GHATAORA, G. 2015. Railway track subgrade failure mechanisms using a fault chart approach. *Procedia Engineering,* 125**,** 547-555.

USMAN, K., GHATAORA, G., BURROW, M., TORBAGHAN, M., WEHB, M. & MUSGRAVE, P. 2017. Fault Tree For Poor Drainage Mechanisms of Railway Ballasted Track. *Railway Engineering*.

VOLLERSTEN, J. & KÖNIG, A. 2005. WP2 Report D6: Model testing and evaluation, Computer Aided REhabilitation of Sewer networks (Care-S). *SINTEF Technology and Society, Trondheim, Norway*.

WANG, G., BURROW, M. & GHATAORA, G. 2020. Study of the Factors Affecting Road Roughness Measurement Using Smartphones. *Journal of Infrastructure Systems,* 26**,** 04020020.

WANG, Q., LIU, K., WANG, M. & KOKS, E. E. 2021. A River Flood and Earthquake Risk Assessment of Railway Assets along the Belt and Road. *International Journal of Disaster Risk Science,* 12**,** 553-567.

WIRAHADIKUSUMAH, R., ABRAHAM, D. & ISELEY, T. 2001. Challenging Issues in Modeling Deterioration of Combined Sewers. *Journal of Infrastructure Systems,* 7**,** 77-84.

WU, W., LIU, Z. & HE, Y. 2015. Classification of defects with ensemble methods in the automated visual inspection of sewer pipes. *Pattern Analysis and Applications,* 18**,** 263-276.

WU, Y., TAIT, S., NICHOLS, A. & RAJA, J. 2021. Simulation of Railway Drainage Asset Service Condition Degradation in the UK Using a Markov Chain–Based Approach. *Journal of Infrastructure Systems,* 27**,** 04021023.

XU, H. & SINHA, S. K. 2019. A Framework for Statistical Analysis of Water Pipeline Field Performance Data. *Pipelines.*

XU, H. & SINHA, S. K. 2020. Applying Survival Analysis to Pipeline Data: Gaps and Challenges. *Pipelines***,** 148-158.

XU, H. & SINHA, S. K. 2021. Modeling Pipe Break Data Using Survival Analysis with Machine Learning Imputation Methods. *Journal of Performance of Constructed Facilities,* 35**,** 04021071.

XU, H., SINHA, S. K. & VISHWAKARMA, A. 2020. Development of a Fuzzy Inference Performance Rating System for Water Pipelines Using a Comprehensive List of Input Variables. *Pipelines.*

YANG, J. 2004. Road crack condition performance modeling using recurrent Markov chains and artificial neural networks.

ZHANG, Z. 2016. Introduction to machine learning: k-nearest neighbors. *Annals of translational medicine,* 4**,** 218-218.

List of Tables:

Table 1. Summary of existing pipe condition models

Utilised method	Study	Considered asset	Pipe size (diameter)	Pipe material	Pipe shape	Pipe lining	Pipe protection	Location	Route	Asset subclass[a]	Type of subgrade	Presence of trees	Traffic load	Construction period	Type of effluent	Serviceability condition	Corrosion -related properties	Age	Cover depth	Backfill material	Groundwater level	Length	Gradient	Land use	Maintenance	Sojourn time	Thornthwaite Moisture Index (TMI)
Homogeneous Markov Chains	Wu et al. (2021)	Rail drainage assets (channel, chamber, culvert, pipe)	X	X	X			X	X	X					X												
	Micevski et al. (2002)	Storm water pipes	X	X							X					X	X										
	Meegoda et al. (2004)	Urban Drainage Culverts	X					X									X	X									
	Wirahadikusumah et al. (2001)	Sewers		X															X	X	X						
	Altarabsheh et al. (2016)	Sewers																X								X	
	Baik et al. (2006)	Sewers	X	X														X				X	X				
	Tran et al. (2008), Tran et al. (2010)	Stormwater pipes	X					X										X	X				X				
Non-homogenous	Le Gat (2008)	Urban drainage pipes	X											X	X												

Markov chains	Kleiner et al. (2006)	Water mains	X	X														X									
Deterministic Linear model	Vollersten and König (2005)	Concrete sewer pipe												X			X		X		X						
Cohort Survival model	Baur and Herz (2002)	Sewer pipes	X	X	X			X		X				X	X								X				
	Xu and Sinha (2021)	Water mains	X			X	X							X			X					X					
Fuzzy method	Xu et al. (2020)	Water mains	X	X		X					X		X				X	X	X		X						
Ensemble Classifier	Mohammadi et al. (2020)	Sewer pipes	X	X													X	X	X		X	X	X				
Neural networks	Tran et al. (2008), Tran et al. (2007), Tran et al. (2006), Tran et al. (2009), Tran et al. (2010)	Stormwater pipes									X	X				X											X

[a]For chambers and channels

Table 2. Applications of ML for various infrastructure assets

Study	ML	Application	Data split	Measure of performance
Assaad and El-adaway (2020)	ANN KNN	Prediction of bridge deck condition	80% training/validation 20% testing	Accuracy= 91%
Asadi et al. (2019)	AdaBoost	Quantification of deterioration in concrete bridge deck	unspecified	Precision Vs. Recall trade-off curve, described as “very good results”
Piryonesi and El-Diraby (2021)	Gradient Boosted Trees (GBT)	Prediction of International Roughness Index (IRI) and Pavement condition index for pavement deterioration	90% training 10% testing	Accuracy=88%
Lee et al. (2018)	ANN SVM	Prediction of Track deterioration as Track Quality Index	85% training 15% test	R^2 =0.86 MSE=0.03
Fathi et al. (2019)	ANN RF	Pavement deterioration prediction of Alligator Deterioration Index (ADI)	75% training 25% testing	R^2 =0.79 RMSE=14.43
Caradot et al. (2018)	RF	Sewer pipe condition	60%training 40% testing	Recall of pipes in bad condition=67%
Mazari and Rodriguez (2016)	RF	Pavement deterioration prediction (IRI)	unspecified	R^2= 0.994and RMSE= 0.049
Wang et al. (2020)	NN DT RF	Prediction of road roughness (IRI)	10-fold cross validation	R^2 =0.92 NN R^2 =0.9 DT R^2 =0.93 RF
Allah Bukhsh et al. (2019)	DT RF GBT	Predict maintenance needs of railway switches	70% training 30% testing	Accuracy= 84% F-score=86%
Moghtadernejad et al. (2021)	NN KNN RF	Conversion of condition from old scheme to new rating scheme for railway supporting structures	66% training 33% testing	F-score= 0.5-0.8
Asnaashari et al. (2013)	NN	Forecasting watermain failure rate	50% training 25% validation 25% testing	R^2=0.94
Ahmed et al. (2021)	KNN NN SVM RF	Prediction of tunnel condition	80% training/validation 20% testing	Accuracy= KNN 80.12% RF 85.38% NN 56.14% SVM 56.73%
Rokstad and Ugarelli (2015)	RF	Structural condition of sewers	Unspecified	ROC Area under curve=60%
Harvey and McBean (2014)	RF	Pipe condition prediction	70% training 30% testing	Recall=82%
Mohammadi et al. (2020)	GBT	Prediction of sewer pipe condition	80% training 20% testing	Accuracy=87%
Wu et al. (2015)	RF	Pipe defect classifications	Four-fold cross validation	Accuracy=80%
Xu and Sinha (2021)	ANN	Imputation of pipe failure records for pipe failure occurrence (classification) and year of failure (regression)	70% training 30% testing	AUC=0.88 MSE=29.79
Tran et al. (2008), Tran et al. (2007), Tran et al. (2006), Tran et al. (2009), Tran et al. (2010)	NN	Storm water pipe condition	60% training 15% validation 25% testing	Accuracy=81%
Ferreño et al. (2021)	Multiple Linear	Prediction of rail pads stiffness	80% training 20% testing	R^2= MLR 0.452 KNN 0.628

	Regression (MLR) KNN Regression Tree (RT) RF GBT ANN Support Vector Machine (SVM)			RT 0.923 RF 0.965 GBT 0.995 ANN 0.990 SVM 0.060
Alawad et al. (2020)	DT	Safety analysis of railway stations	80% training 20% testing	Accuracy= 88.7%
Kalathas and Papoutsidakis (2021)	DT RT	Maintenance prediction of railway traction/braking system	Not specified	Accuracy= 80% Weighted Precision= 0.8 Weighted Recall= 0.8
Mittal and Rao (2017)	NN	Detection of track defects, switches and signals and monitor track health	80% training 20% testing	Defect detection: Precision= 95% Recall= 25% Signal detection: Accuracy= 93% Precision= 99% Switch Detection: Precision= 94% Recall= 83%
Arshad and Ahmed (2021)	LR GBT DT RF	Prediction of train delay due to weather impact	Not available	Accuracy= LR 90% GBT 92% DT= 94% RF 96%
Liu et al. (2018)	RF	Quantify susceptibility of railway to rainfall-induced hazards	60% testing 40% validation	Accuracy=95%

Table 3. Confusion matrix used in this study

		Target class				
	p t	**1**	**2**	**3**	**4**	**5**
Predicted class	**1**	**TP_1**	**$FP_{1,2}$ ($FN_{1,2}$)**	**$FP_{1,3}$ ($FN_{1,3}$)**	**$FP_{1,4}$ ($FN_{1,4}$)**	**$FP_{1,5}$ ($FN_{1,5}$)**
	2	**$FP_{2,1}$ ($FN_{2,1}$)**	**TP_2**	**$FP_{2,3}$ ($FN_{2,3}$)**	**$FP_{2,4}$ ($FN_{2,4}$)**	**$FP_{2,5}$ ($FN_{2,5}$)**
	3	**$FP_{3,1}$ ($FN_{3,1}$)**	**$FP_{3,2}$ ($FN_{3,2}$)**	**TP_3**	**$FP_{3,4}$ ($FN_{3,4}$)**	**$FP_{3,5}$ ($FN_{3,5}$)**
	4	**$FP_{4,1}$ ($FN_{4,1}$)**	**$FP_{4,2}$ ($FN_{4,2}$)**	**$FP_{4,3}$ ($FN_{4,3}$)**	**TP_4**	**$FP_{4,5}$ ($FN_{4,5}$)**
	5	**$FP_{5,1}$ ($FN_{5,1}$)**	**$FP_{5,2}$ ($FN_{5,2}$)**	**$FP_{5,3}$ ($FN_{5,3}$)**	**$FP_{5,4}$ ($FN_{5,4}$)**	**TP_5**

Note: $FP_{p,t}$: False Positive (e.g. $FP_{1,2}$ means false positive with respect to predicted class 1, actual class is 2)
$FN_{p,t}$: False Negative (e.g. $FN_{1,2}$ means false negative with respect to target class 2, misclassified as 1)
TP: True Positive

Table 4. Common steps of CWA

Step	Factor X_j	HN_1	HN_2	HN_3

1	Connection weights between the input factor X_j and HN from NN code	A_1	A_2	A_3
2	Connection weights between HN and the output neuron C from NN code	B_1	B_2	B_3
3	Local Significance measure $Z_{j,c}$ of input factor X_j to output neuron C	$Z_{j,c} = \sum_{i=1}^{3} A_i * B_i$		
4	Overall significance measure OZ_j of input factor X_j to NN model	$OZ_j = \frac{1}{3}\sum_{c=1}^{3} abs(Z_{j,c})$		

Source: Reprinted with permission from Tran et al. (2009)

811 **Table 5.** Description of condition scores

Condition score	Structural description	Service description
0	Not inspected	Not inspected
1	No defects	Clear
2	Superficial defects	Superficial deposits
3	Minor defects	Performance slightly reduced
4	Major defects	Performance severely reduced
5	Not fit for purpose/unsafe	Blocked

Source: Adapted from Network Rail unpublished drainage inspection standards, 2018

812 **Table 6.** Input data into the developed model

Input	Description	Possible values
Length	Scale	(0 m-100 m)
Effluent	Categorical (1-3)	1; Surface water 2; Combined 3; Foul water
Topography	Categorical (1-3)	1; At Grade 2; Cutting 3; Embankment
Size	Categorical (1-4)	Ranges between (<150 mm->451 mm)
Material	Categorical (1-11)	Steel, uPVC, Concrete, Cast Iron, Vitrified Clay, Pitch Fibre, Spun Iron, Asbestos Cement, lined Brick, Unlined Brick, other.
Shape	Categorical (1-4)	1; Circle 2; Rectangle 3; Square 4; egg
Previous Recorded condition	Ordinal (1-5)	1, 2, 3, 4, 5
Years since previous recorded condition	Scale	Number of years since the previous recorded condition (0 if not matched)
Maintenance	Dichotomous	1; If condition improved since last recorded condition 0; If else

813 **Table 7.** Order of significance of input factors

Order of significance	STRC	SRVC
1	Previous condition	Previous condition
2	Length	Maintenance
3	Years since previous condition	Years since previous condition
4	Shape	Shape

5	Maintenance	Material
6	Size	Length
7	Topography	Size
8	Effluent	Effluent
9	Material	Topography

Table 8. Significance of input factors towards STRC in the literature compared with the findings of the current study

Literature	Pipe Size	Pipe Length	Pipe Material	Effluent type	Location	Pipe Shape	Slope	Soil type	Tree count	Age	Depth	Groundwater table	Climatic condition	Road traffic	Hydraulic condition
(Baik et al., 2006)	x	x	-				x			x					
(Davies et al., 2001a)	x	x		x	x										x
(Tran et al., 2006)	-				-		x	-	-	-	x		-		x
(Tran et al., 2009)	x	-			x		-	x	x	-					
(Micevski et al., 2002)	x		x					x					x		-
(Wirahadikusumah et al., 2001)			x					x			x	x			
(Angarita et al., 2017)	x		x	x	x										
(Ugarelli et al., 2013)	x			x				x	x	x				x	
Mohammadi et al. (2020)	x	x	x				x	-		x	x	x			
Current study from CWA	x	x	-	x	x	x									

Note: x indicates the factor has been considered and was found significant
- indicates the factor has been considered and was found insignificant

List of Figures

Figure 1

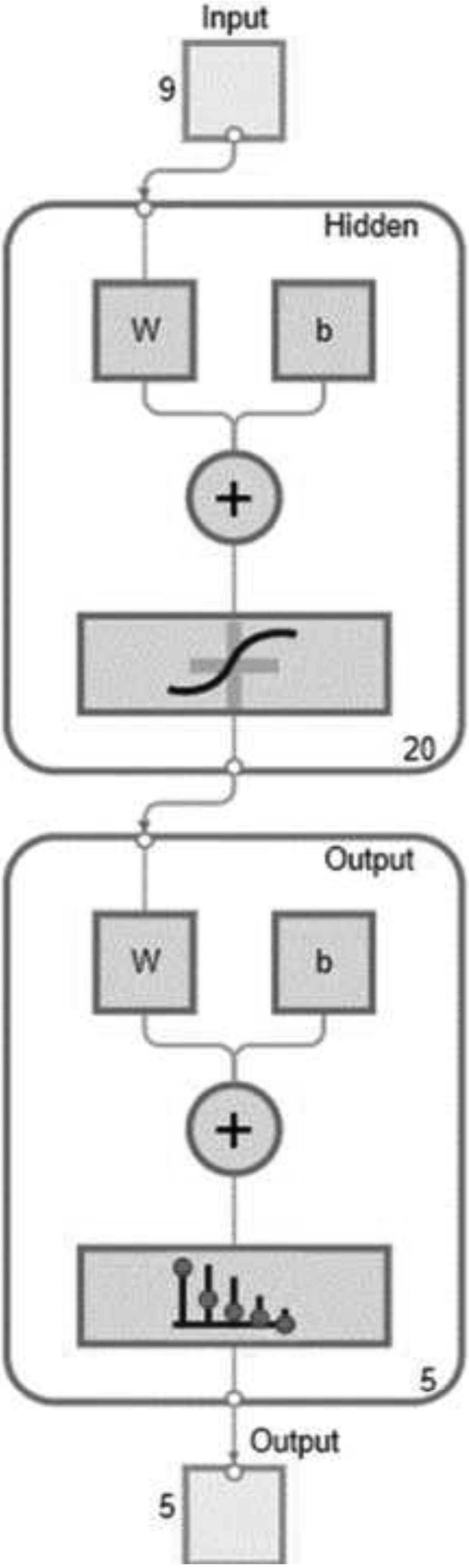


Figure 2

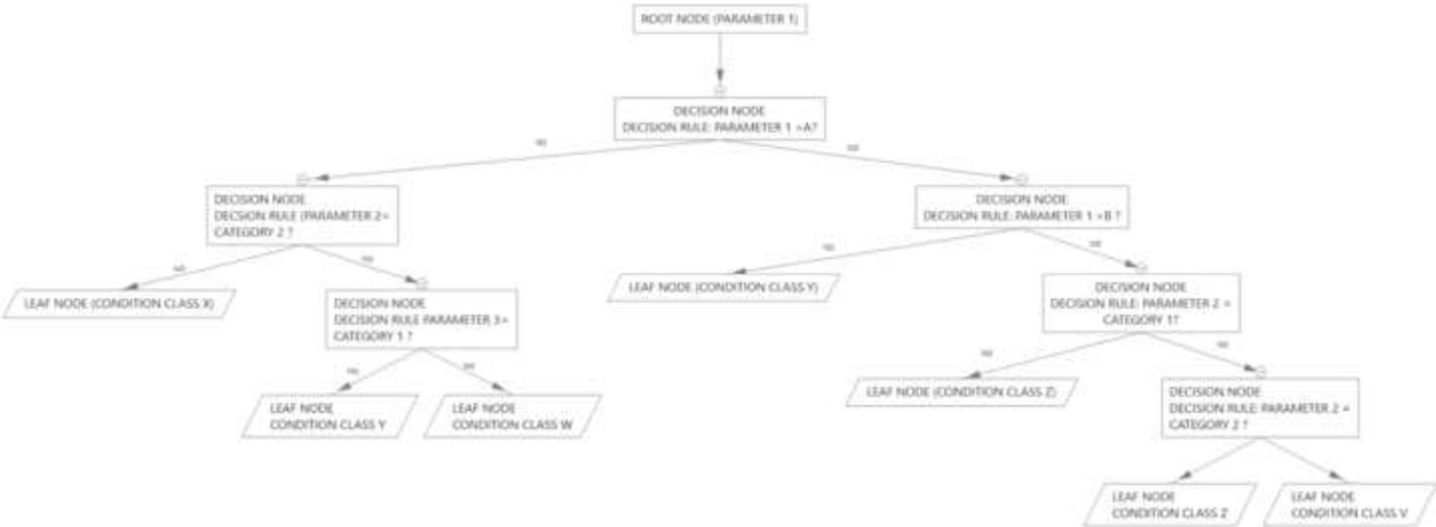

ROOT NODE (PARAMETER 1)
DECISION NODE
DECISION RULE: PARAMETER 1 =A?
DECISION NODE
DECSION RULE (PARAMETER 2=
CATEGORY 2 ?
DECISION NODE
DECISION RULE: PARAMETER 1 =B ?
LEAF NODE (CONDITION CLASS X)
DECISION NODE
DECISION RULE PARAMETER 3=
CATEGORY 1 ?
LEAF NODE (CONDITION CLASS Y)
DECISION NODE
DECISION RULE: PARAMETER 2 =
CATEGORY 1?
LEAF NODE
CONDITION CLASS Y
LEAF NODE
CONDITION CLASS W
LEAF NODE (CONDITION CLASS Z)
DECISION NODE
DECISION RULE: PARAMETER 2 =
CATEGORY 2 ?
LEAF NODE
CONDITION CLASS Z
LEAF NODE
CONDITION CLASS V

Figure 3

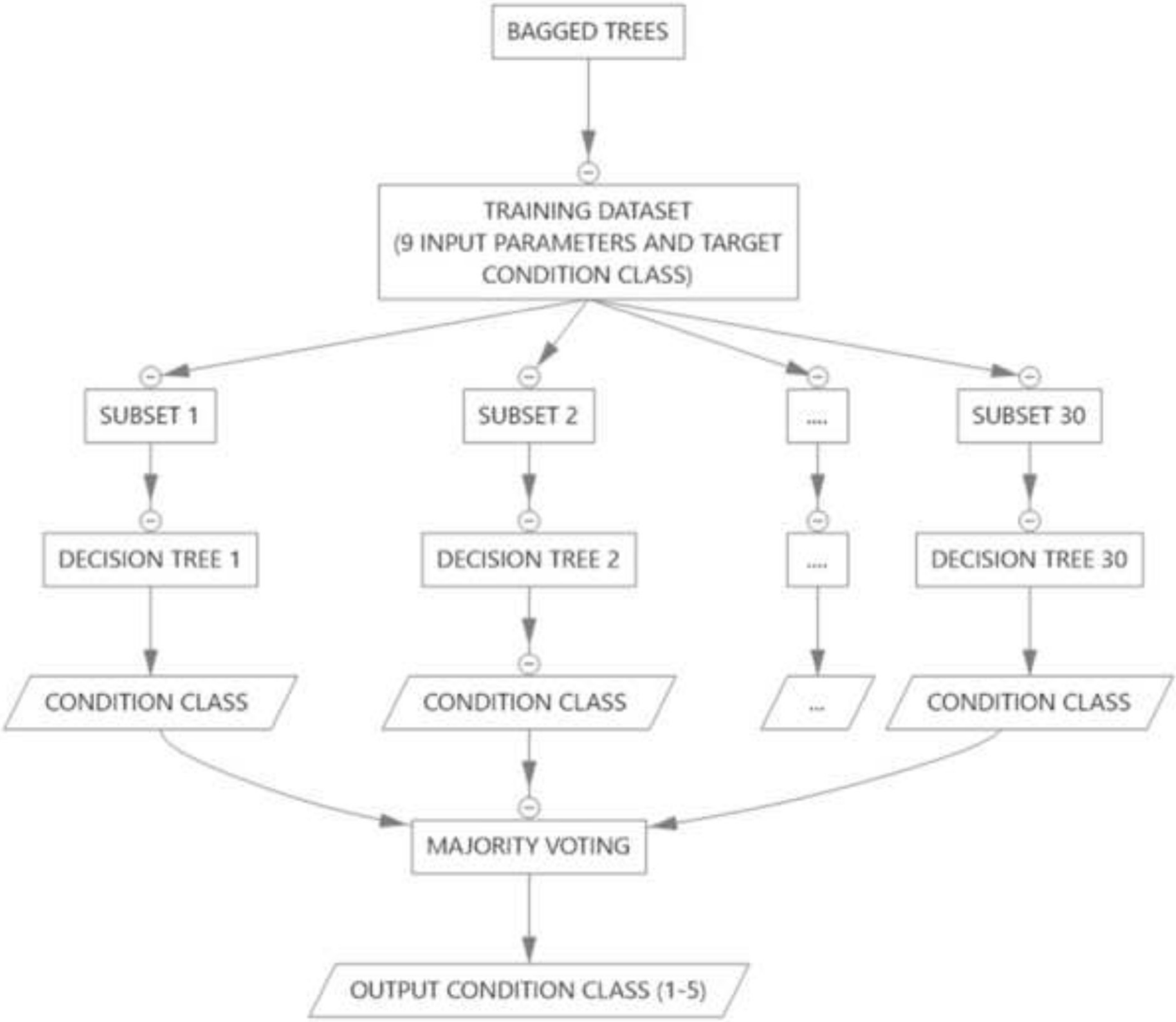

BAGGED TREES
TRAINING DATASET
(9 INPUT PARAMETERS AND TARGET
CONDITION CLASS)
SUBSET 1
SUBSET 2
....
SUBSET 30
DECISION TREE 1
DECISION TREE 2
....
DECISION TREE 30
CONDITION CLASS
CONDITION CLASS
...
CONDITION CLASS
MAJORITY VOTING
OUTPUT CONDITION CLASS (1-5)

Figure 4

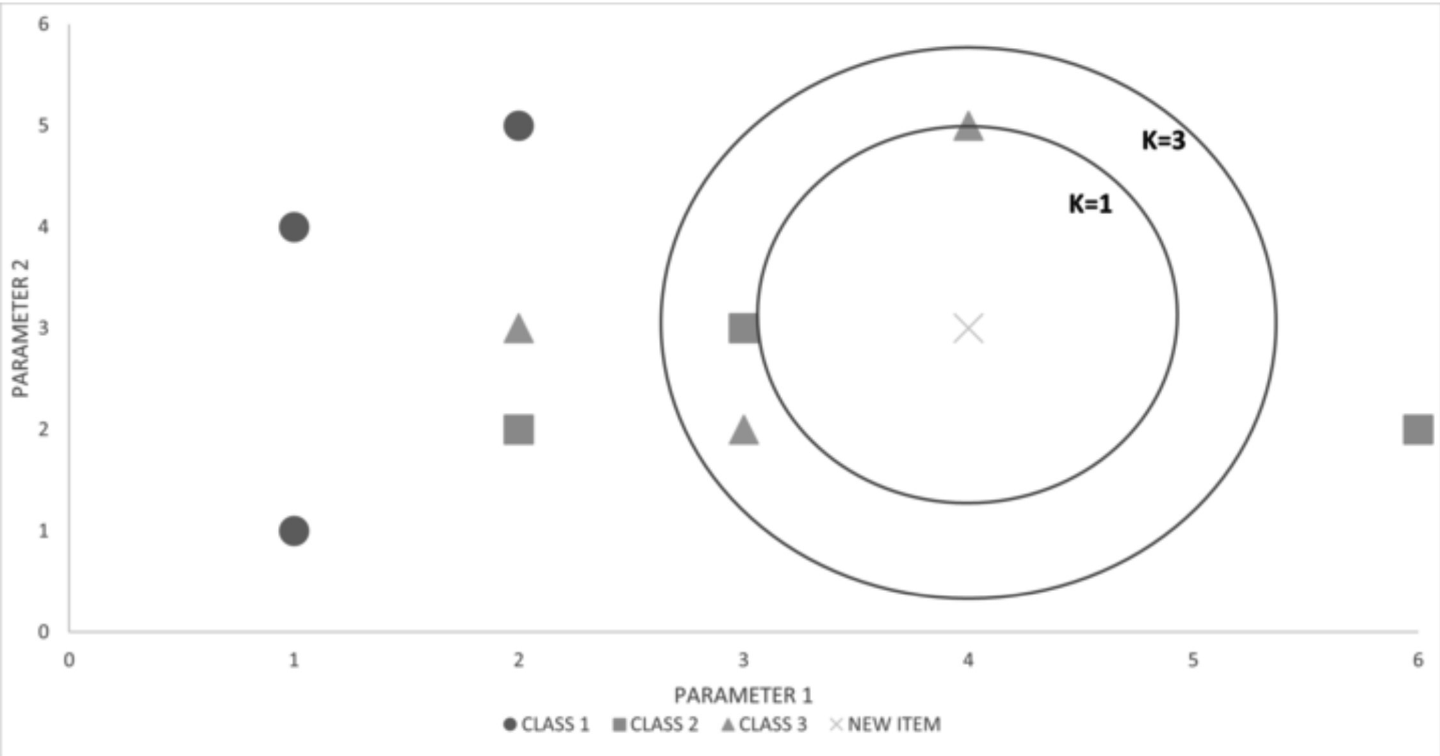

6
5
4
3
2
1
0
PARAMETER 2
K=3
K=1
0
1
2
3
4
5
6
PARAMETER 1
CLASS 1
CLASS 2
CLASS 3
NEW ITEM

Figure 5

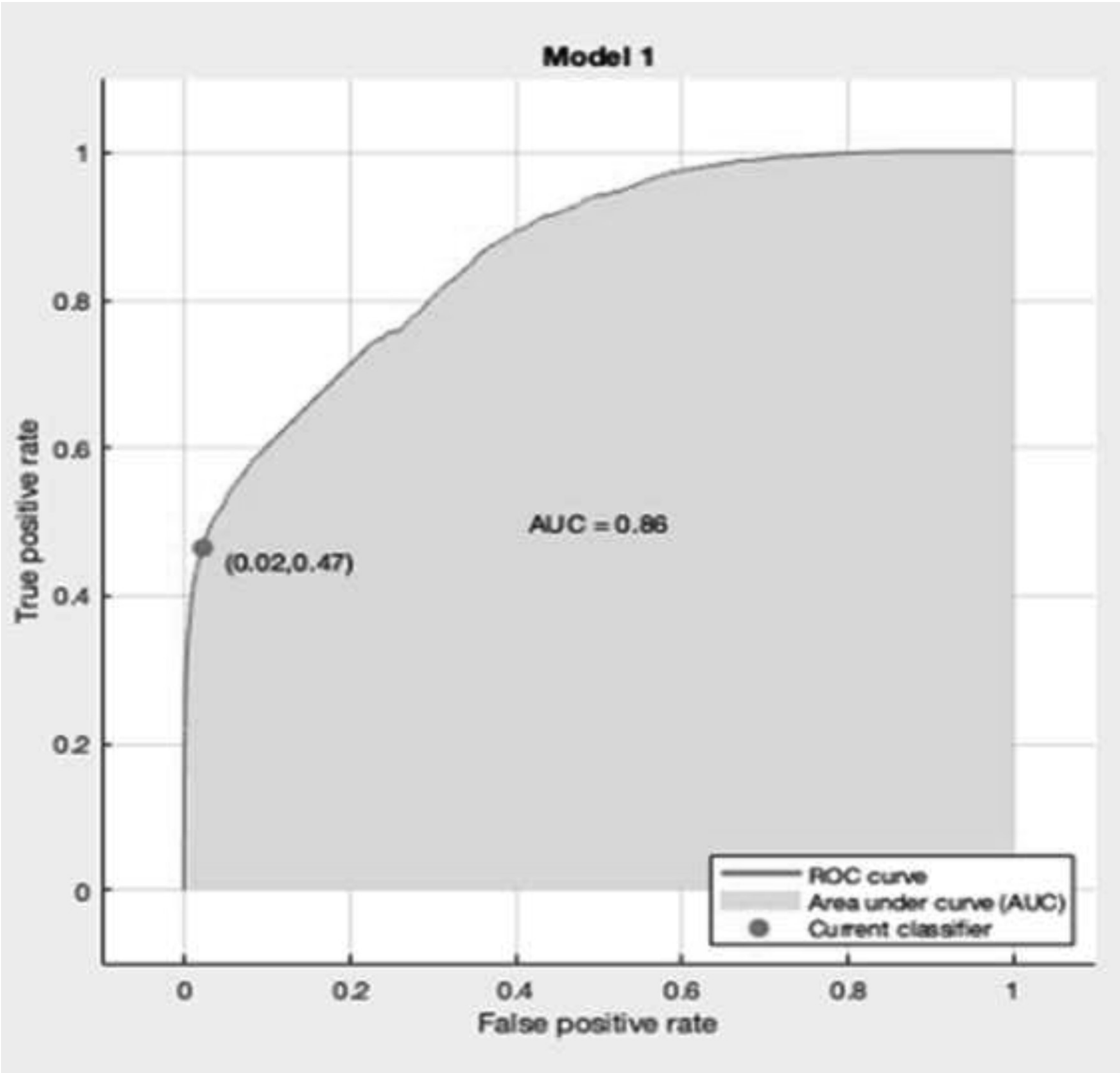

Model 1
True positive rate
False positive rate
0
0.2
0.4
0.6
0.8
1
AUC = 0.86
(0.02,0.47)
ROC curve
Area under curve (AUC)
Current classifier

Figure 6

Click here to access/download;Figure;FIG 6.tif

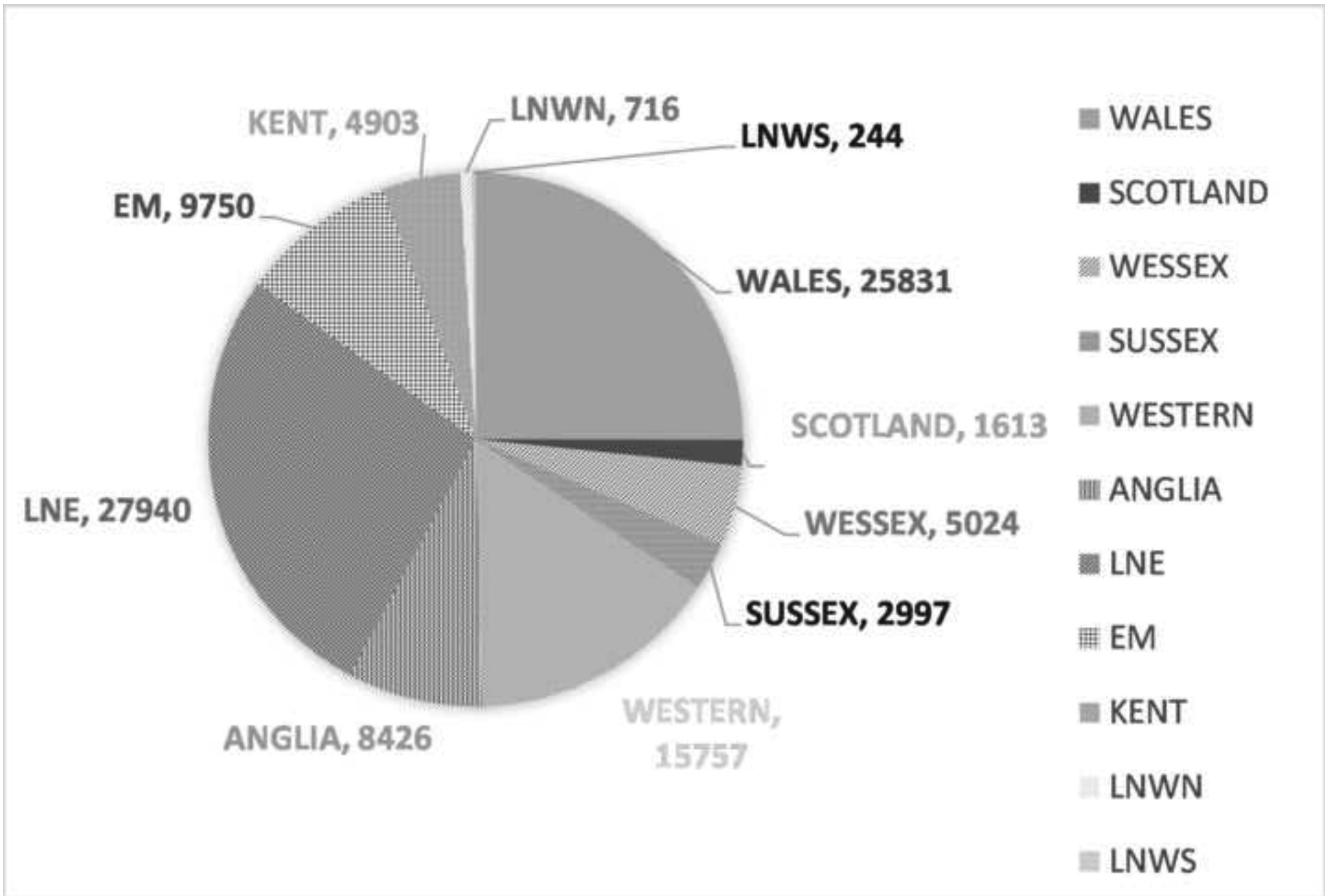


Figure 7 Click here to access/download;Figure;FIG 7.tif

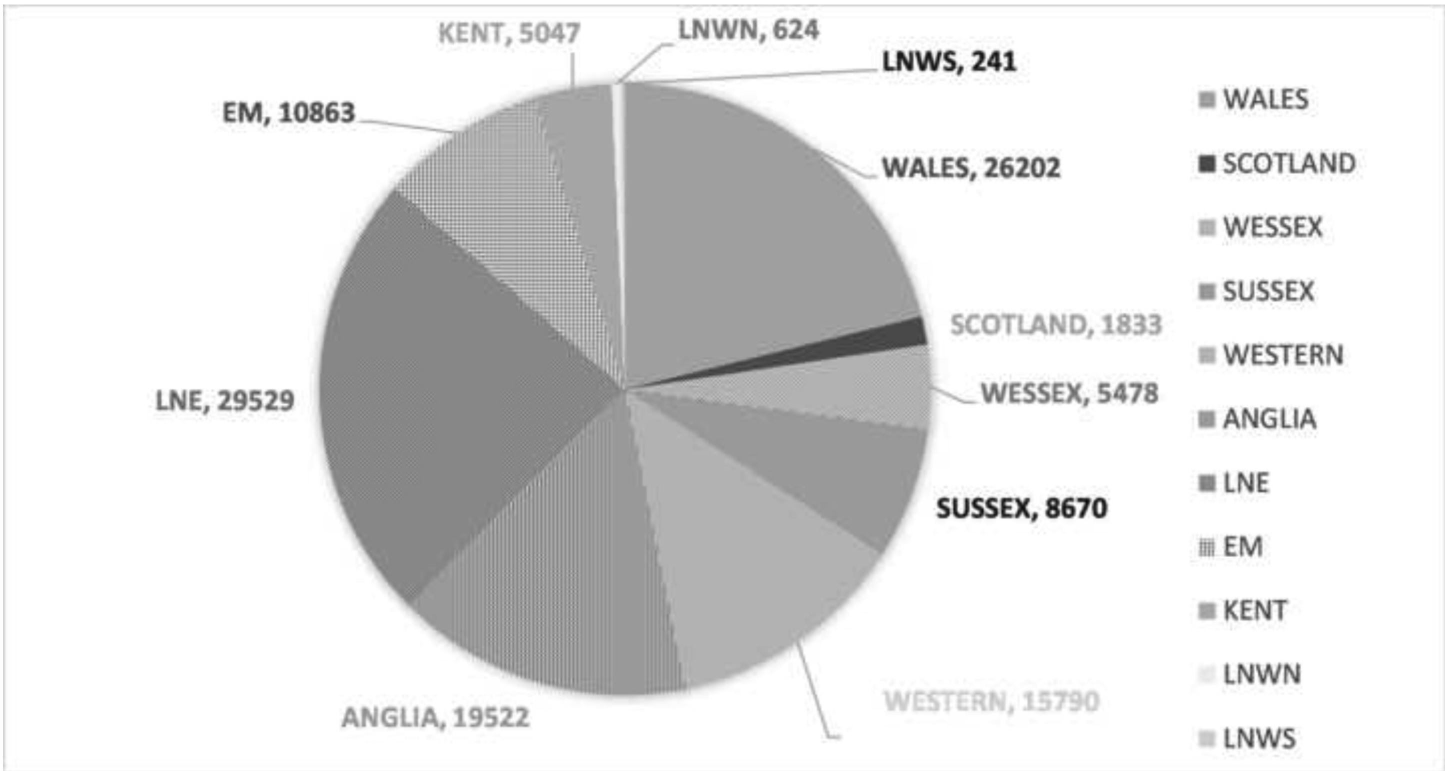


Figure 8

Click here to access/download;Figure;FIG 8.tif

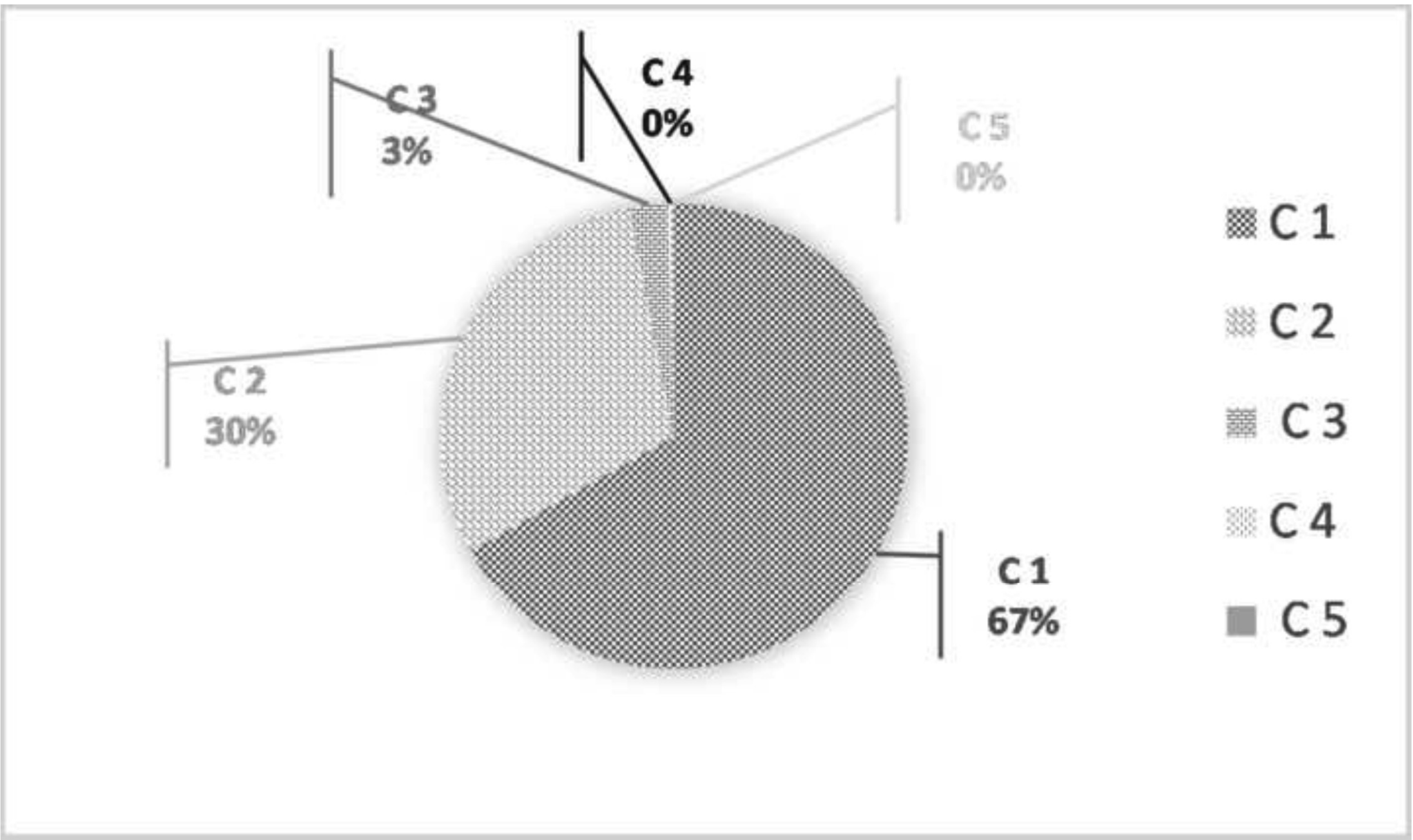


Figure 9

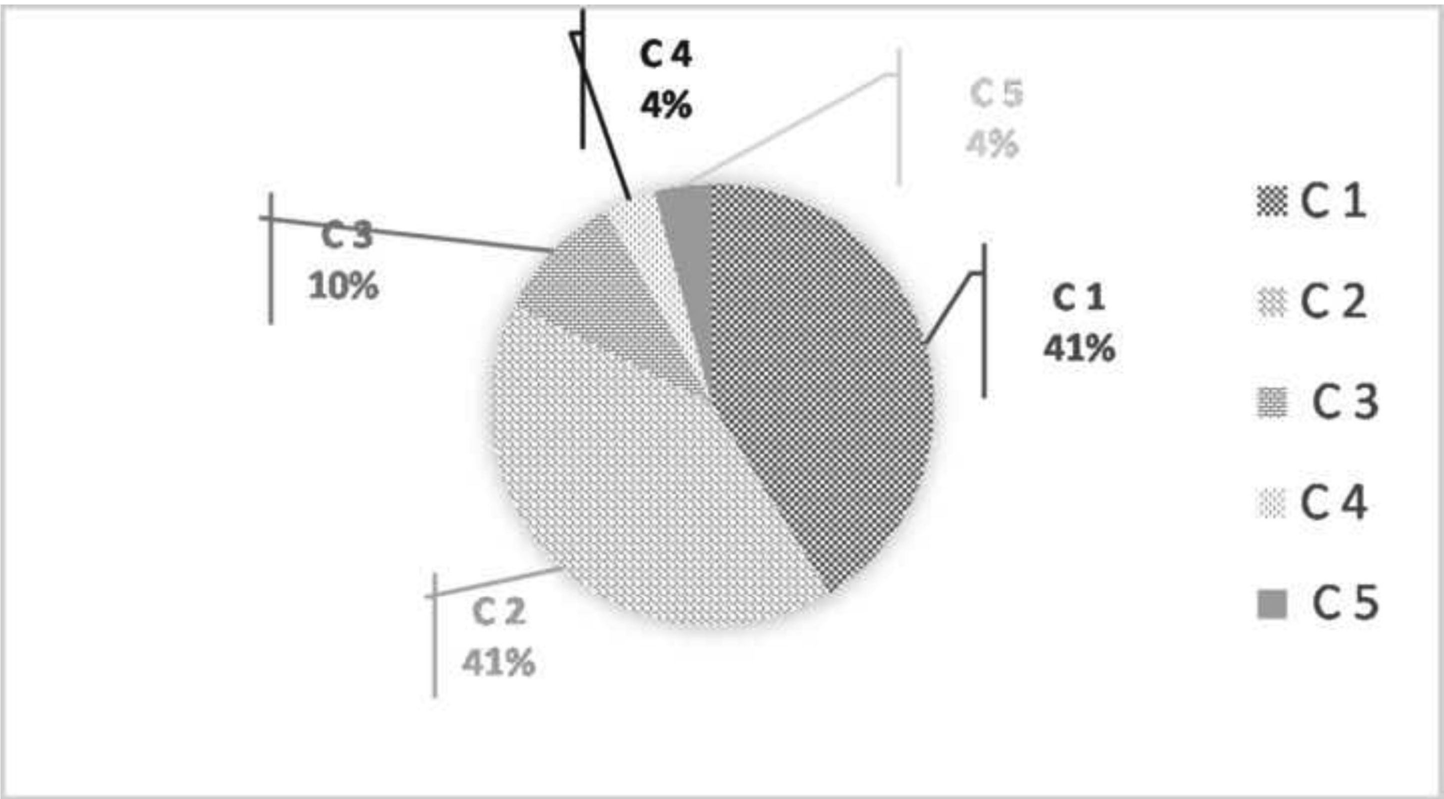
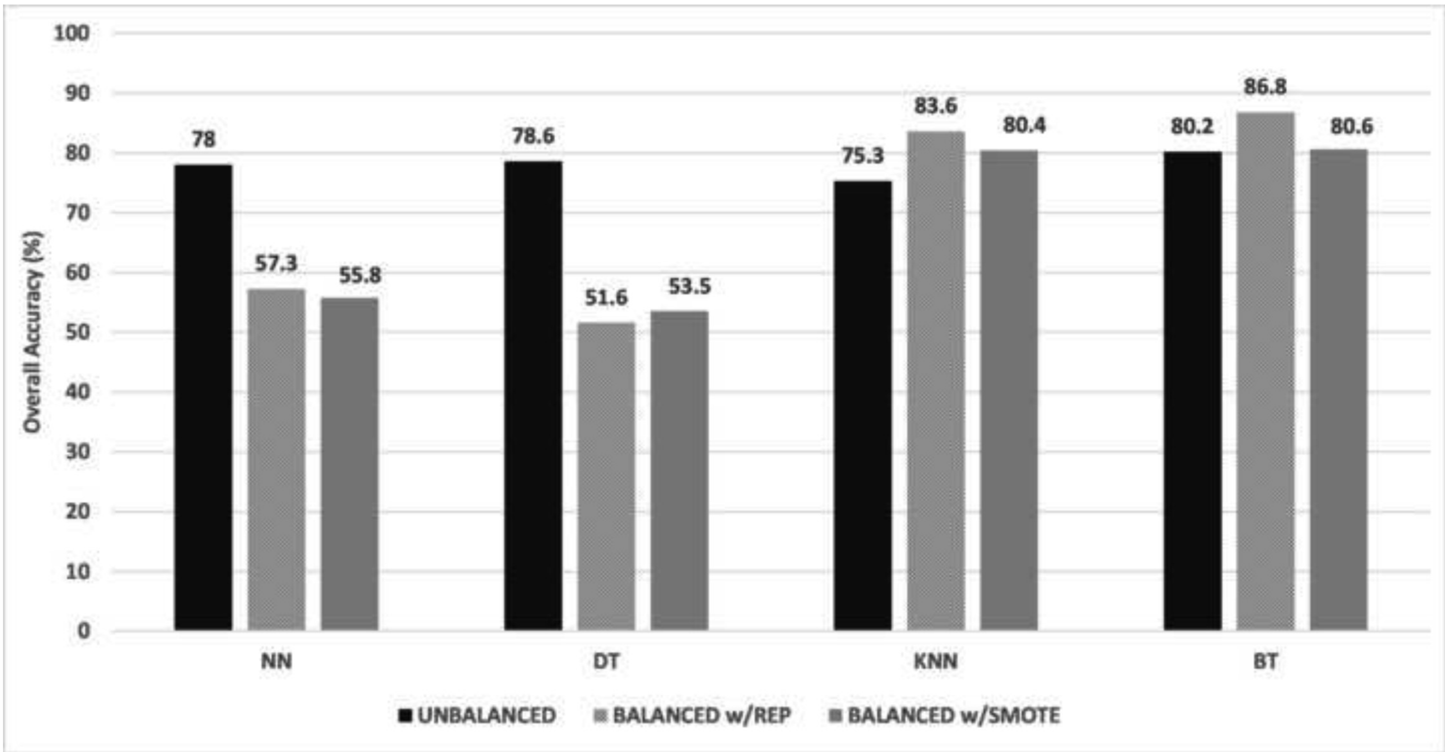


Figure 10

Click here to access/download;Figure;FIG 10.tif



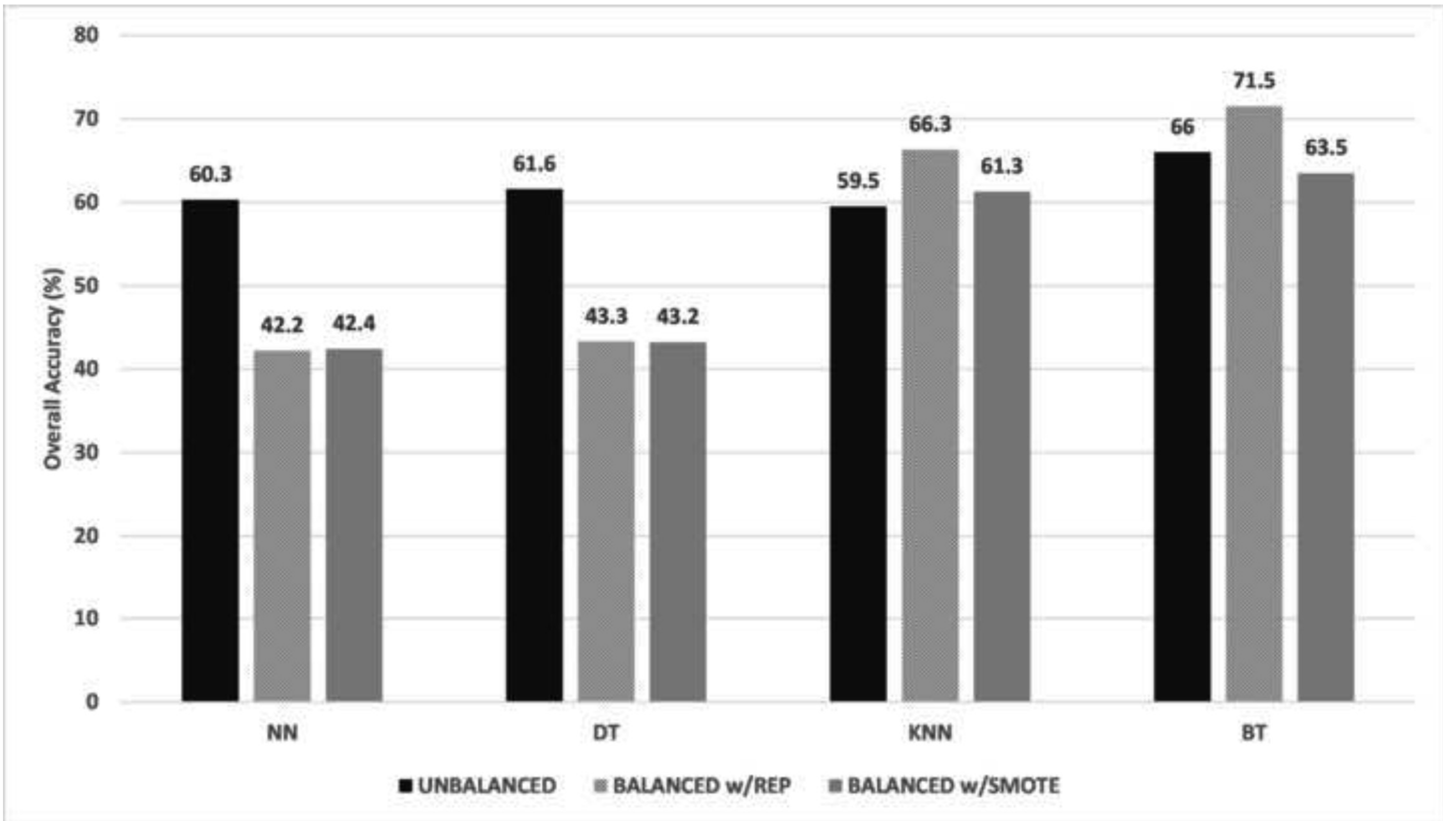

80
70
60
50
40
30
20
10
0
Overall Accuracy (%)
60.3
42.2
42.4
61.6
43.3
43.2
59.5
66.3
61.3
66
71.5
63.5
NN
DT
KNN
BT
UNBALANCED
BALANCED w/REP
BALANCED w/SMOTE

Figure 12

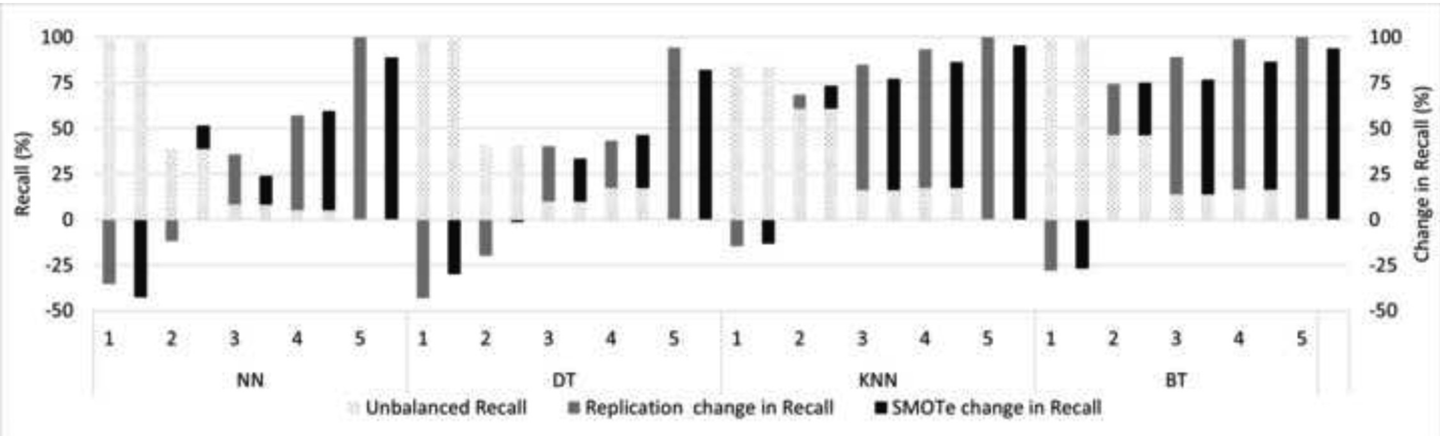
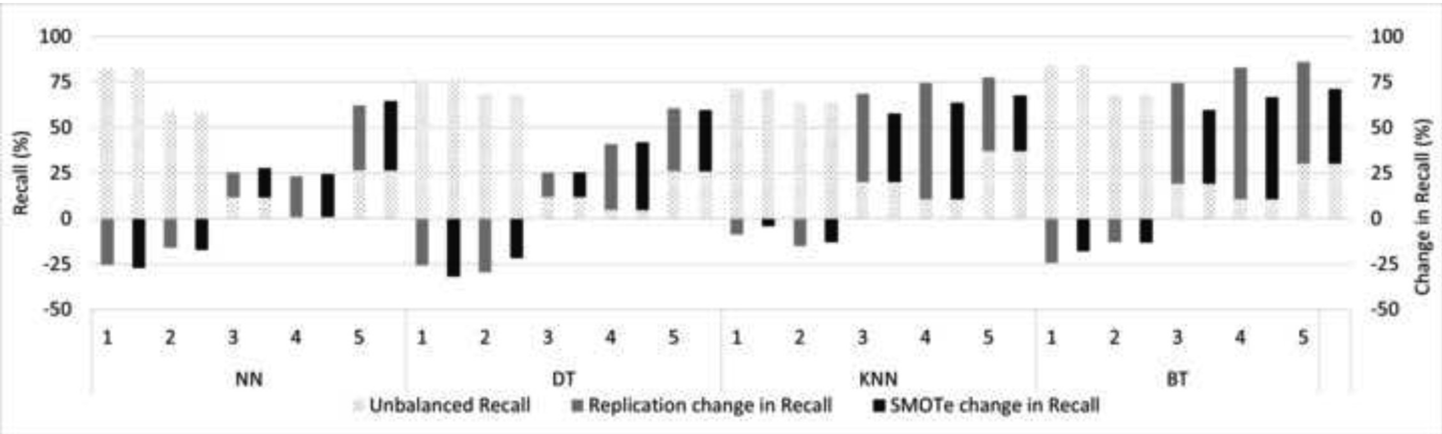
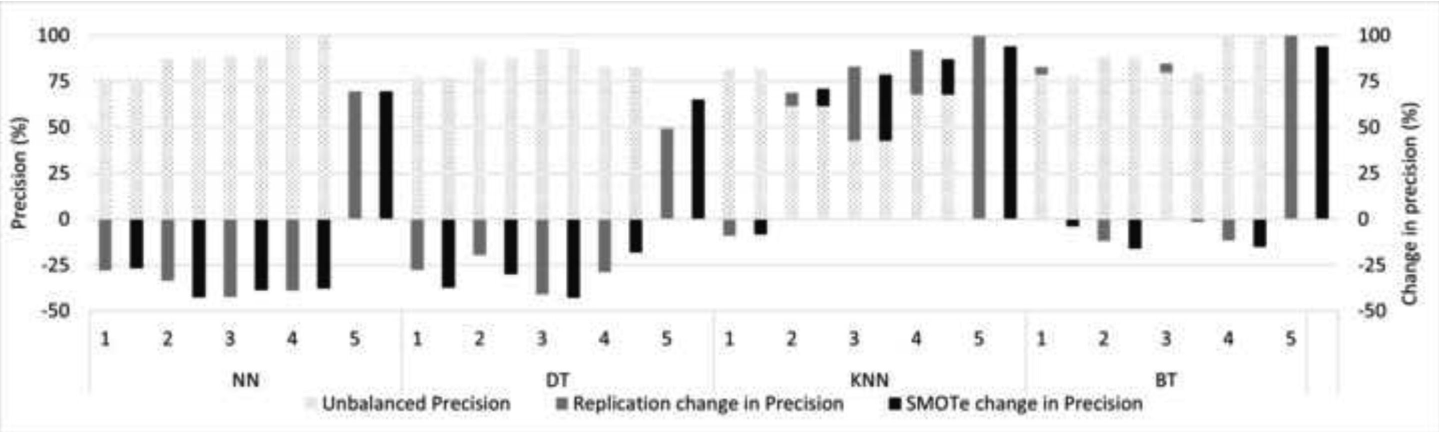

100
75
50
25
0
-25
-50
Recall (%)
Change in Recall (%)
1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5
NN
DT
KNN
BT
Unbalanced Recall
Replication change in Recall
SMOTe change in Recall

Figure 13



100
75
50
25
0
-25
-50
Recall (%)
Change in Recall (%)
1
2
3
4
5
NN
DT
KNN
BT
Unbalanced Recall
Replication change in Recall
SMOTe change in Recall

Figure 14



100
75
50
25
0
-25
-50
Precision (%)
Change in precision (%)
1
2
3
4
5
NN
DT
KNN
BT
Unbalanced Precision
Replication change in Precision
SMOTe change in Precision

Figure 15

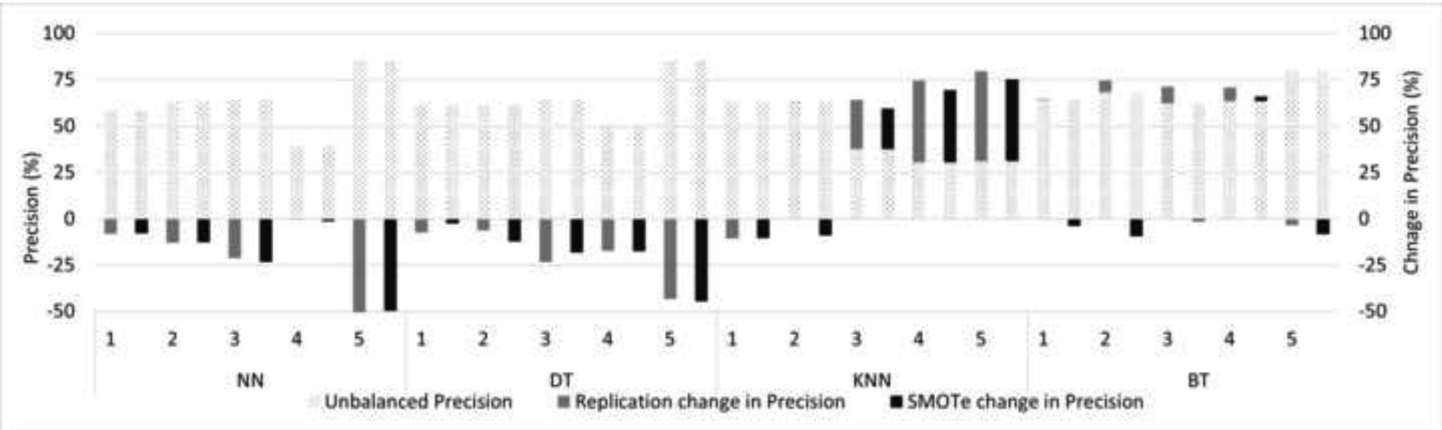

100
75
50
25
0
-25
-50
Precision (%)
Chnage in Precision (%)
1
2
3
4
5
NN
DT
KNN
BT
Unbalanced Precision
Replication change in Precision
SMOTe change in Precision

Figure 16

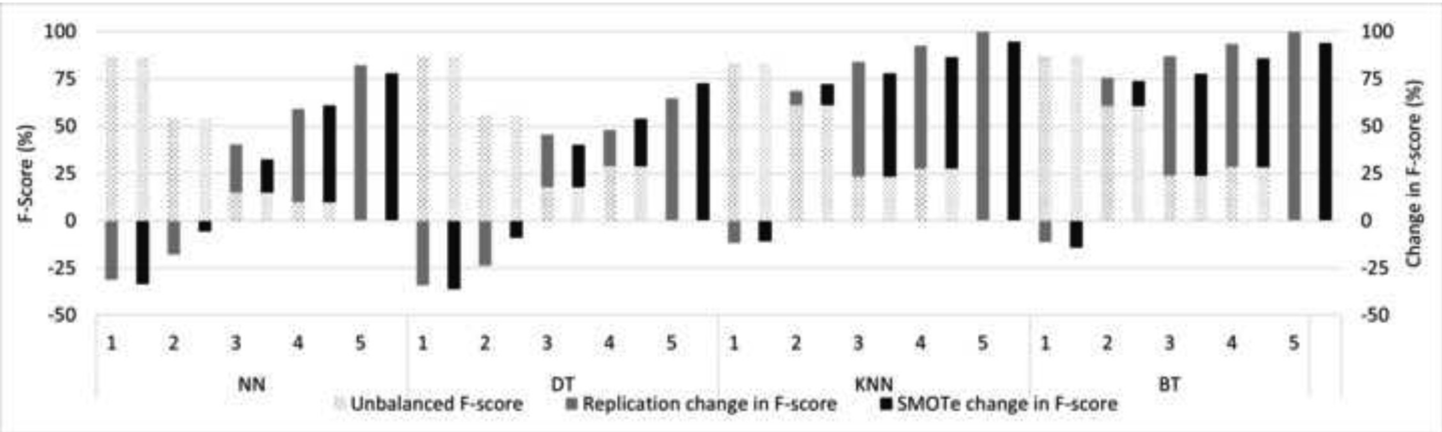

100
75
50
25
0
-25
-50
F-Score (%)
Change in F-score (%)
1
2
3
4
5
NN
DT
KNN
BT
Unbalanced F-score
Replication change in F-score
SMOTe change in F-score

Figure 17 Click here to access/download;Figure;FIG 17.tif

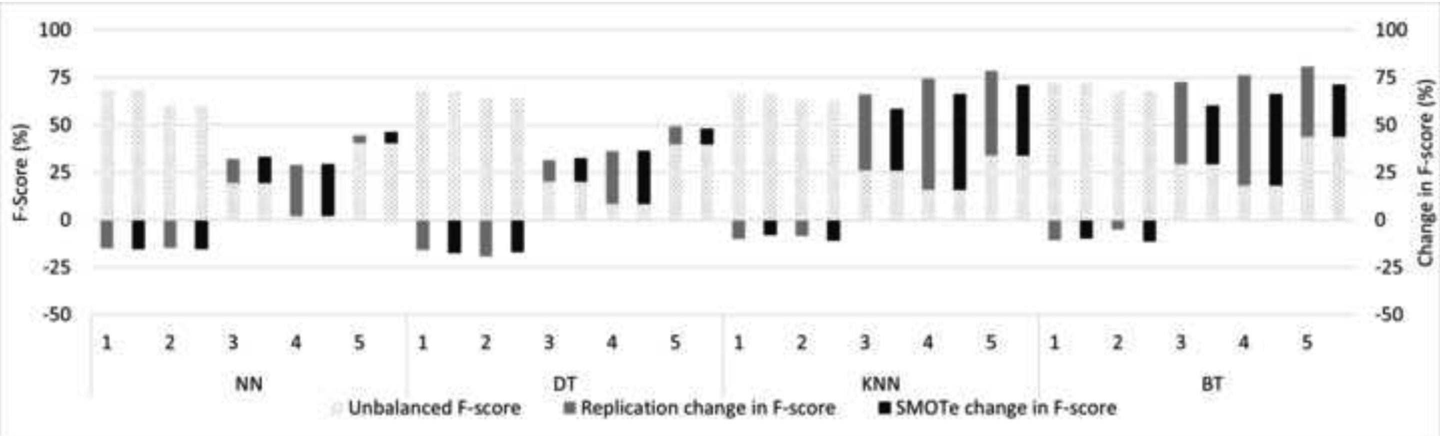


Figure 18

Click here to access/download;Figure;FIG 18.tif

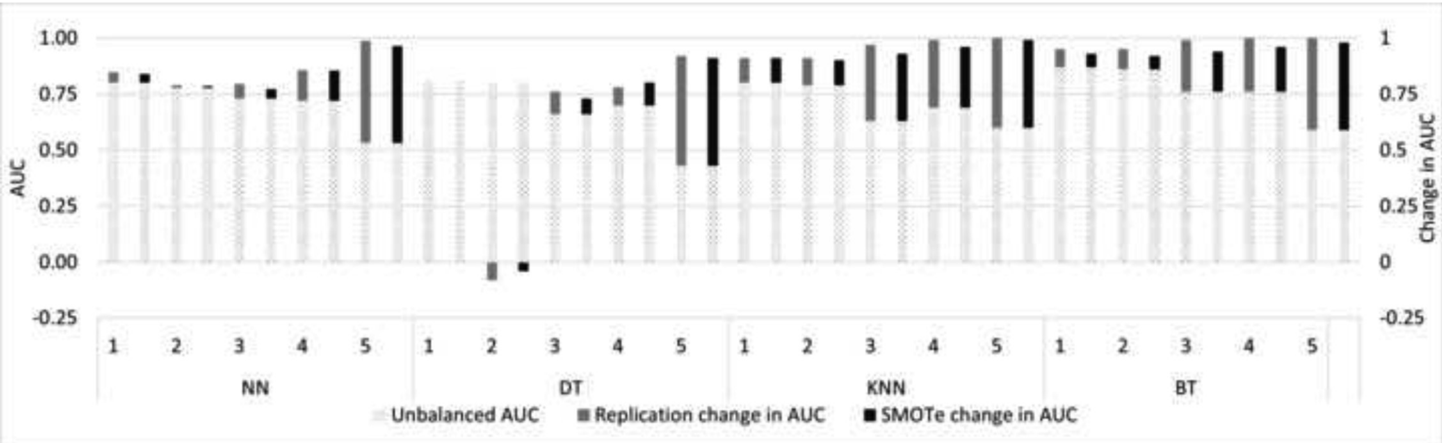
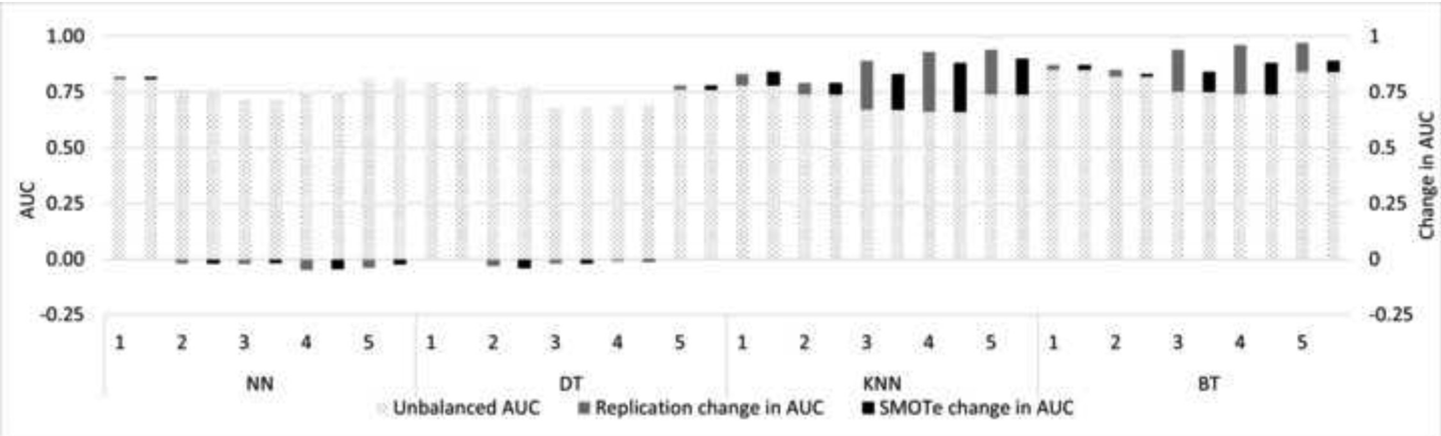
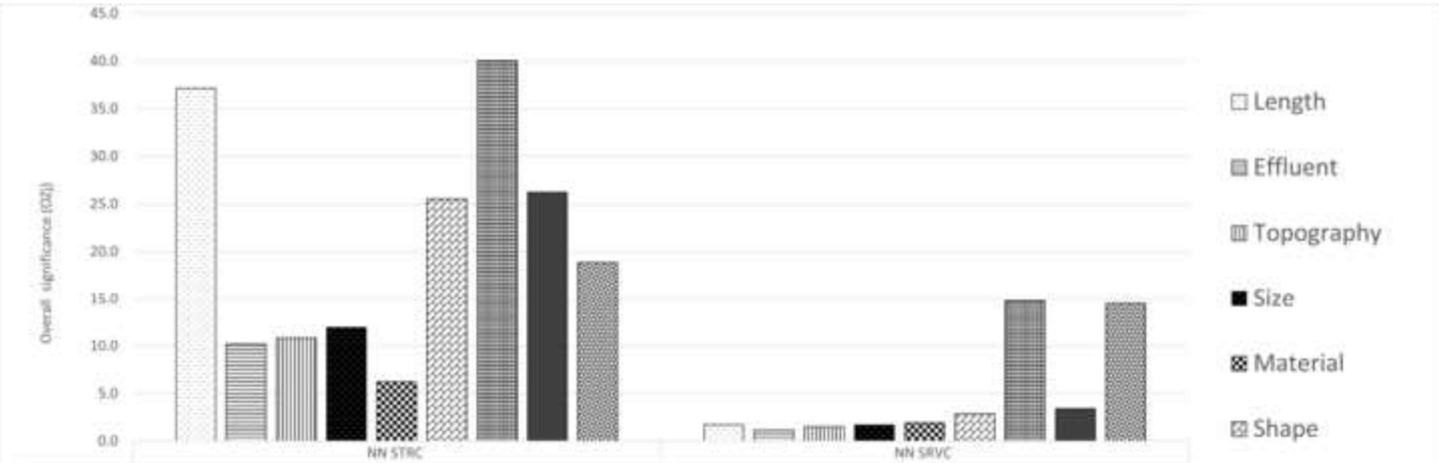


Figure 19



1.00
0.75
0.50
0.25
0.00
-0.25
AUC
1
0.75
0.5
0.25
0
-0.25
Change in AUC
1 2 3 4 5
NN
DT
KNN
BT
Unbalanced AUC
Replication change in AUC
SMOTe change in AUC

Figure 20



	NN STRC	NN SRVC
Length	37.1	1.7
Effluent	10.2	1.2
Topography	10.8	1.5
Size	11.9	1.7
Material	6.2	1.9
Shape	25.5	2.9
Previous Cond.	40.0	14.8
Years	26.2	3.4
Maintenance	18.9	14.5

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Dec 20, 2021

This Agreement between university of Birmingham -- nour Aljafari ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	5211520100030
License date	Dec 17, 2021
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Computer-Aided Civil and Infrastructure Engineering
Licensed Content Title	Comparison of Structural Deterioration Models for Stormwater Drainage Pipes
Licensed Content Author	A.W.M. Ng, B. J. C. Perera, D. H. Tran
Licensed Content Date	Dec 18, 2008
Licensed Content Volume	24
Licensed Content Issue	2
Licensed Content Pages	12
Type of use	Journal/Magazine
Requestor type	University/Academic
Is the reuse sponsored by or associated with a pharmaceutical or medical products company?	no

Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No
Circulation	50000 or greater
Title of new article	Condition modelling of drainage pipes
Lead author	Nour Aljafari
Title of targeted journal	Journal of Infrastructure systems
Publisher	ASCE
Expected publication date	May 2022
Portions	Table 2. Example of computing steps for overall significance measure of an input factor
Requestor Location	university of Birmingham 62 lincoln apartments 3 lexington gardens birmingham, B15 2DS United Kingdom Attn: university of Birmingham
Publisher Tax ID	EU826007151
Total	0.00 USD
Terms and Conditions	

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner.**For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS

OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The Creative Commons Attribution Non-Commercial (CC-BY-NC)License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library http://olabout.wiley.com/WileyCDA/Section/id-410895.html

Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.