https://doi.org/10.1016/j.scitotenv.2021.151110

Data-Driven Application of MEMS-based Accelerometers for Leak Detection in Water Distribution Networks

Authors

Author 1: Salman Tariq

Postdoctoral Fellow (Corresponding Author), Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong.

Author 2: Beenish Bakhtawar

Research Assistant, Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong.

Author 3: Tarek Zayed

Professor and Associate Head, Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong.

Citation for this article: Tariq, S., Bakhtawar, B., & Zayed, T. (2021). Data-driven application of MEMS-based accelerometers for leak detection in water distribution networks. *Science of The Total Environment*, 151110. https://doi.org/10.1016/j.scitotenv.2021.151110

Data-Driven Application of MEMS-based Accelerometers for Leak Detection in Water Distribution Networks

Abstract

Water scarcity is a global concern; 68 countries are facing extremely-high to medium-high risk of water stress. In this era of crisis, where water conservation is an absolute necessity, the water distribution networks (WDNs) globally are experiencing significant leaks. These leaks cause tremendous financial loss and unacceptable environmental hazards, thus further aggravating the water scarcity situation. To minimize such damage, the adoption of advanced technologies and methodologies for leak detection in the WDNs is absolutely necessary. In this regard, we have investigated the application of cost-effective MEMS-based accelerometers. Experiments were conducted on real networks (metal and non-metal pipes), over the course of ten months, and the acquired acceleration signals were analyzed using a monitoring algorithm. Monitoring index efficiencies and standard deviations for every leak and no-leak case was extracted. Two individual [KNN and Decision Tree] and two ensembles [Random Forest and Adaboost (Decision Tree)] based machine learning models were developed for the accurate classification of the leak and no-leak cases using extracted features; and separate models were developed for metal and non-metal pipes. Random Forest outperformed the other machine learning models and the overall accuracy reached 100% for metal pipes and 94.93% for non-metal pipes. The machine learning models were further validated using unseen/unlabeled cases and were highly effective in detecting leaks. This study demonstrated the applicability of MEMS-based accelerometers for leak detection and established real network-based machine learning models thereby contributing to the research scarcity in this important area.

Keywords: MEMS; Accelerometers; Leak detection; Real Networks; Machine Learning; Ensembles;

1. Introduction

Recent statistics show that around 27% of the global population has been living in water-scarce areas since the mid-2010 with a predicted escalation to more than 42% in 2050 (UN-WWAP 2018). Additionally, around 50% of the world's population inhabits areas that suffer from water scarcity at least one month a year (Boretti and Rosa 2019). To alleviate the severity of this situation, it is necessary to improve the network capacity and financial operations of existing water distribution networks (WDNs) through effective water management strategies, such as the minimization of non-revenue water (NRW) (El-Zahab and Zayed 2019). USEPA (2010) reported that NRW typically exceeds 30% in most WDNs, and the percentage can even be as high as 50% in older networks (Kanakoudis and Muhammetoglu, 2014). The main sources of NRW include illegal connections, ineffective metering, and leaks (Tariq et al. 2020). However, leaks usually constitute the largest portion; sometimes reaching 70% of the total NRW (Van Zyl and Clayton, 2007), thus, reduction of pipeline leaks alone can effectively enhance the WDN capacity (Hu et al. 2021).

In Hong Kong, freshwater and seawater are distributed through a pipe network of over 8600 kilometers. About half of the pipes were laid around 40 years ago as a part of major urban and town development at that time (Yue and Tang 2011). Subsequent deterioration of the pipes due to aging has resulted in the loss of water through leaks which was estimated at 25% in 2000 in government-operated water mains (Wong 2018). Consequently, the Water Supply Department (WSD) launched an extensive rehabilitation and replacement scheme the same year to improve water conservation in Hong Kong (Yue and Tang 2011). Following the scheme, water loss through leaks was reduced substantially to 15% in 2017, however, significant improvement is still required as 8,512 leaks were reported in 2017 (WSD 2020). In monetary terms, this enormous scale of financial loss from water leaks is costing Hong Kong US$173 million annually (Gupta 2017). Therefore, the Hong Kong government is urgently looking into enhancing its existing leak detection approaches to curtail the damage.

Vibro-acoustic technologies based on piezoelectric transducers are the most commonly used leak detection approach in Hong Kong as in other parts of the world (Bakhtawar and Zayed 2021). However, piezoelectric technology-based leak detection has certain limitations that make these devices less suitable for a congested city urban landscape like Hong Kong (Tariq et al. 2020; Hamilton and Charalambous 2013). These limitations are listed as follows.

- Plastic pipes make up around 35% of the total network in Hong Kong; noise loggers and other piezoelectric transducers are not-so-effective for plastic pipes due to the attenuation of the sound signals and frequently leads to false alarms. Other than the type of pipe material, the efficiency of these technologies greatly varies with pipe diameter and the surrounding environment. In terms of research, the scale of experimentations using noise loggers and piezoelectric accelerometers is typically carried out on the lab and test rig scale, with little validation over real networks, thus, failing to significantly improve the current situation.
- Hydrophones are inconvenient to deploy although effective for plastic pipes.
- The initial cost of deployment for both noise loggers and hydrophones is high.

To overcome these challenges, non-invasive technologies such as using accelerometers can prove helpful. Accelerometers are vibration-based sensors which are easy to deploy and produce high-frequency signals with fewer distortions as compared to other sensors like geophones, hydrophones, and pressure sensors (Almeida et al., 2015; Brennan et al., 2019). In addition, all these prevalent technologies rely on experts who have to make regular visits for data collection and can only detect leaks effectively over a short distance (Ismail et al., 2019). These limitations make long-term and long-range leak detection tedious. Technologies based on Micro-Electro-Magnetic-Sensors can be of great potential to circumvent such limitations.

MEMS-based accelerometers can be more effective for leak detection, combining the benefits of both MEMS and accelerometers for real networks. They offer a three-dimensional sensitivity, better accuracy, affordability, and the option to transmit data wirelessly, proving more desirable for Hong Kong conditions as compared to many other sensing options (Tariq et al., 2020). MEMS accelerometers can provide significant power and communication savings as compared to other piezoelectric transducers. However, challenges related to time synchronization, long-time communication, and implementation for real pipe networks are frequently cited (Shinozuka et al., 2010). Additionally, there is limited work available on modeling approaches effective for vibration/acceleration data collected through MEMS-based accelerometers. For example, Shinozuka et al. (2010) used a PipeTECT wireless sensing system based on MEMS accelerometers and employed frequency domain analysis for detecting water leaks. El-Zahab et al. (2018) employed statistical modeling approaches coupled with the monitoring index methodology developed by Martini et al. (2017). However, current studies are not insightful for the purpose of experimentation on real-life networks and real leaks. Thus, to explore the accuracy and efficiency of such methods, the current study uses MEMS-based accelerometers for real water networks. The current objectives of the research are three-fold: 1) investigating the capability of cost-effective MEMS-based accelerometers in detecting leaks in the WDN; 2) developing the leak thresholds using a monitoring index-based algorithm to identify the leak and no-leak cases in real networks; and 3) establishing accurate machine learning-based leak detection models, thus minimizing the handling of false alarms.

2. Literature Review

Accelerometers are versatile and sensitive sensors used for acceleration measurements. These small-sized devices are non-destructive, easy to deploy, and convenient to use for efficient leak detection in water pipelines. Pipe vibrations caused by leaks have higher acceleration values on the surface as compared to other vibration sources. Based on this principle, accelerometers can detect the constant difference between acceleration values over time so as to differentiate between leak and no-leak situations (Marmarokopos et al., 2018). Various studies have demonstrated the use of accelerometers for detecting water leaks. For example, Hunaidi and Chu (1999) and Gao et al. (2005) investigated the application of piezoelectric accelerometers in small diameter metal and plastic pipes. Martini et al. (2015) developed an algorithm enabling early detection of burst leaks in service pipes. The algorithm was expressed as a ‘monitoring index’ developed using the standard deviations of the leak signals expressed in Equation 1.

$$MI_j = \frac{1}{N}\sum_{k=1}^{N}(\sigma_{j,k}) \dots \dots \dots \dots (1)$$

where $\sigma_{j,k}$ is the kth element of the vector including the N lowest standard deviation values of the jth dataset. Using threshold values, the algorithm serves as a useful tool for the characterization of acceleration data in the leak and no-leak cases. Martini et al. (2017) further compared two mono-axial piezoelectric accelerometers measuring radial and axial accelerations, respectively, with a hydrophone. Small diameter plastic pipes were used for the experimentation under different scenarios to check the performance of the sensors. It was found that the radial accelerometer covered a much narrow frequency bandwidth as compared to the axial accelerometer and the hydrophone. Thus, it is only effective for leak detection over short distances. Both the accelerometers failed to detect distant leaks, even after pre-filtering the data, as a result of the signal attenuation in the plastic pipes. Marmarokopos et al. (2018) also supported this finding by showing the applicability of accelerometers for detecting near-distant leaks in plastic pipes. Their study used a high signal-to-noise ratio piezoelectric accelerometer in a laboratory setting and analyzed the signals in both time and frequency domains. El-Zahab et al. (2016) demonstrated the same algorithm as Martini et al. (2015) on a lab scale using a more sensitive type of accelerometer based on micro-electro-mechanical-system (MEMS) technology. The experiments were performed on both metal and plastic pipes. This study established the accuracy of MEMS accelerometers by varying the distance between leak and accelerometers. Encouraged by the results, El-Zahab et al. (2018) further designed machine learning-based models which provided highly accurate results on lab-scale experiments.

Among the more recent studies, Ismail et al. (2019) compared different vibration sensors on the basis of their sensitivity, accuracy, power consumption, and cost. The study conducted testbed experimentation which indicated the superiority of tri-axial accelerometers for plastic pipes over other sensors. Okosun et al. (2019) used a highly sensitive piezoelectric sensor, integrated with an amplifier to ensure a low signal-to-noise ratio over a wide frequency range. Their sensor system showed the viability for long-term leak detection in plastic pipes. Kampelopoulos et al. (2020a) used adaptive filtering to reduce noise and established thresholds for leak detection parameters such as kurtosis, correlation, power spectral density, and energy. Their system showed particular efficiency in detecting high power leaks up to a distance of 75m.

However, existing studies have some major shortcomings. Firstly, the scale of experiments and the implications of their findings to real networks are limited. Most of the studies use laboratory and testbed settings with artificially simulated leaks to study different parameters that impact on the leak problem in water pipes. For example, Zahab et al. (2016) used a lab setup with simulated leaks to study the effect of pipe material and diameter on the leakage. El-Zahab et al. (2018) used a similar setup to test one leak at a time where the leaks were simulated using valve openings, and claimed high accuracy for the tested algorithms in their ability to detect leaks. Although such studies demonstrate the possible ability of accelerometers to detect leaks in a water distribution system, however, in real pipe networks, there may be more than one leak and their locations are not so obvious to predict in the system. Additionally, the nature of the acceleration data collected under real conditions is expected to be very different because of ambient noise impacts. For the case of accelerometers, the ambient noise effect can be a big problem resulting in high pre-processing efforts. It should be noted that noise signals may not follow the assumptions stated in most correlation models while predicting leaks. In such situations, false alarms pose a threat and may result in high repair costs.

Secondly, the pipe geometries used in the reported experiments are also relatively simple. Most of the experiments used a straight pipe or a pipe with bends to demonstrate the leak problem (Martini et al., 2015; Mostafapour & Davoudi, 2013). However, the real pipe networks in urban environments are not represented by such simplifications. There is also limited information available about the underground conditions and records available concerning repairs and pipe replacements so it is difficult to assess the leak environment accurately. The leak experimentation for real networks is also hindered by site condition factors e-g the shape or size of the valve may not be suitable for accelerometer deployments or the site may be difficult to access. It is also possible that there may not be any consecutive valves available to carry out cross-correlation studies. The dynamics of leak detection in real water distribution networks are, therefore, much more complex as compared to controlled experiments. In the context of this discussion and the highlighted limitations, it becomes necessary to demonstrate accelerometer technology in real water networks. Our current research, thus, uses MEMS-based accelerometers for leak detection experimentation on real networks in Hong Kong WDNs. The demonstrated analyses method extends the MI methodology developed by Martini et al. (2015) on rich real leak and no leak data. MI-based leak thresholds on real leaks are also suggested. Using statistical rationalization and initial inferences, useful inferences about the leak and no leak classification of the data are suggested. Further, machine learning-based classification models are developed and demonstrated for more efficient and accurate leak detection helping decision-makers in timely repairs and pipe rehabilitation planning.

3. Research Methodology

The overall research approach for our study, illustrated in Figure 1, consisted of four distinct phases: 1) research background and conceptualization, 2) data collection on real networks, 3) data analysis, and 4) development of machine learning models. The first phase dealt with (a) leak problem identification in Hong Kong, (b) literature review of accelerometers-based studies, algorithms for handling acceleration signals, and machine learning-based classification techniques, and (c) finalization of the research methods. The second phase lasted over ten months and acceleration signals for both leak and no-leak cases from different pipe types and materials were collected. The third phase occurred in parallel to the second phase. In this phase, we (a) examined the collected acceleration signals, (b) analyzed the collected acceleration signals and developed leak thresholds using a monitoring indexed-based algorithm, and (c) provided the necessary monitoring index efficiencies (leak thresholds) and standard deviations which were used to develop machine learning models in the fourth phase. Performance measurement and validation of the developed models were also conducted in the fourth phase. For developing machine learning models, both individual and ensemble methods were applied. Among the individual methods, Decision Trees and K-Nearest Neighbor (KNN) were applied due to their ability to effectively handle numerical data. Among the ensemble methods, 1) Random Forest which is an ensemble of Decision Tree, and 2) Adaboost to boost Decision Tree were used.

[Insert Figure 1]

4. Leak Thresholds: Procedures and Development

4.1. Data acquisition system

A MEMS accelerometers-based high time-synchronized data acquisition system from the German brand 'Beanair' was used for experiments on real networks. The complete system comprised five triple-axis accelerometers (model AX3D with a sensitivity of ±2g), a gateway, a beanscape software (data acquisition software), and a laptop (for storing data and visualizing signals). The gateway had to be connected to the laptop using an Ethernet cable whereas the accelerometers transferred signals through the gateway wirelessly. The maximum sampling rate of the accelerometers with all three-axis in use was 1,000 samples/second and with one-axis in use was 3,000 samples/second, respectively. However, on-site preliminary investigations revealed that only measurements taken along the radial sensing successfully detected leaks. The measurements along the axial directions in real networks are often impractical as accelerometers are placed on valves not the actual pipes. Therefore, this study only used the radial sensing direction for signals collection with a sampling rate of 3,000 samples/second in the streaming mode. The complete MEMS-based data acquisition system is showed in Figure 2.

[Insert Figure 2]

4.2. Experimental design and protocols

A rigorous experimental campaign was designed to collect acceleration signals, from pipes of all sizes and materials, across various sections of the WDN in Hong Kong. The objective was to collect signals for both leak and no-leak cases to establish leak thresholds. Therefore, signals were collected at the time of leakage and after repair. However, it was necessary to establish leak thresholds before starting experiments for collecting leak signals; no-leak signals during normal network operations confirmed as no-leak sites by the

Hong Kong Water Supply Department (WSD) in advance, were collected for several hours (multiple mid-nights), and analyzed using a monitoring algorithm (details are given later). Signals collected after repair added an additional layer of confirmation for the leak thresholds.

The procedures for taking measurements at the leak site in collaboration with the WSD and a local contractor were as follows: 1) the WSD received alarms of a potential leak and informed the local contractor of the site location; 2) our research team along with the local contractor visited the site the same day at midnight. Mid-night was selected to keep night flow at a minimum and reduce background noise from traffic and other sources.

4.3. Data collection

The experiments were conducted over the course of ten months from October 1st, 2020 to July 31st, 2021. At any particular potential leak location, the accelerometers were placed on the available gate valves or fire hydrants near the leak location. The actual placement of accelerometers in the field is shown in Figure 3. The duration of signal collection varied from 300 seconds to several thousand seconds depending upon the number of site locations to be covered each night, the distance between different site locations, and the site conditions. A snapshot of the acceleration signal is given in Figure 4. Besides signals, notes were taken regarding 1) the distance between the potential leak and the accelerometer, 2) the distance between accelerometers if more than one accelerometer was used, and 3) any nearby noise-generating sources. Later, the leak/no-leak status was confirmed with the WSD. If the WSD stated that it was a false alarm, the case was closed and signals were stored as no-leak signals. If the WSD confirmed it as a true leak, our research team and the local contractor again visited the site at midnight on the same day of repair and again collected the signals. Following the above-mentioned protocols, leak and no-leak signals of 993 cases in total from 75 different sites in Hong Kong were collected. The collection of signals from such a large number of sites at different locations brought representativeness of the whole WDN as most of the boundary conditions and functional characteristics of the distribution system were taken into account.

[Insert Figure 3]

[Insert Figure 4]

4.4. Signal processing

Our study used a standard deviation-based monitoring algorithm developed by Martini et al. (2017) for signal processing and the development of leak thresholds. The main advantages of this algorithm include 1) its suitability for raw signals derived from accelerometers; 2) no additional requirement of any signal pre-processing; and 3) proven effectiveness both in lab-based and field-based experiments. El-Zahab et al. (2018) used this algorithm to differentiate between leak and no-leak states employing lab-scale experiments, whereas Martini et al. (2017) found its efficacy in real networks for above-ground service connection pipes. However, none of the studies tested the practicality of this algorithm for underground pipes. Our preliminary studies in the Hong Kong water supply network, before designing the actual experiments, confirmed the applicability of this algorithm for underground pipes. This utility was further confirmed by the local contractor, who has been involved in the leak detection process for over 20 years in Hong Kong. Throughout the course of our study, the standard deviation-based monitoring algorithm has been found to be extremely effective in detecting leaks, even more than all the other traditional feature-based algorithms (tested in our study) which required extensive signals pre-processing. The mathematical approach behind the algorithm can be summarized in the following steps:

1) Acceleration readings per second (g) were collected as mentioned in the previous sections;
2) Every 100 seconds, the standard deviations of the g readings were computed. For example, for an hour-long (3600 seconds) reading, 36 standard deviations were computed (Figure 5);
3) To establish the leak thresholds, only a subset of the 10 lowest standard deviations of no-leak signals were used for each mid-night in equation 2.

$$MI_j = mean\left(\sigma_j, 10\right) \ldots \ldots \ldots \ldots (2)$$

The selection of the subset helped to minimize perturbations caused by external sources. The lowest leak threshold was named as the monitoring index (MI_o). Monitoring indexes were established for different pipe types and diameters as given in Table 1.

4) Since the presence of a leak in a pipe causes the MI_j values to increase and the comparison of such values with MI_o can be used for providing a warning about the leak. Therefore, the monitoring index efficiencies MI_E of each pipe type were established using equation (3) which provided a clear indication of a leak in case the value was exceeded.

$$MI_{Ex} = \frac{MI_j}{MI_o} \dots \dots \dots \dots (3)$$

5) In case of a leak, the MI_E values were computed and compared with the MI_E values for specific pipe types. Typically, the MI_E values of the leaked pipes exceeded the MI_E values of non-leaked pipes (Figure 6 for example).

[Insert Figure 5]

[Insert Table 1]

[Insert Figure 6]

5. Development of Leak Detection Model

Leak detection in WDNs is a binary classification problem that can be solved by applying machine learning algorithms to accurately recognize leak and no-leak states using acceleration signals/data (Ravichandran et al. 2021). The main component typically include 1) collection of acceleration data from lab or field-based experiments (field-based in our case); 2) feature extraction using a signal processing algorithm; 3) the development of machine learning algorithms-based leak detection models using extracted features; and 4) application of machine learning models to reach a binary decision on leak or no-leak classification on the a) testing dataset; and b) validation dataset. It is now a widely accepted fact that machine learning is an effective approach in dealing with binary classification problems (Ravichandran et al. 2021). A simple schematic diagram representing the general framework of such a binary classification process is given in Figure 7, explained as follows.

[Insert Figure 7]

- As per this framework, firstly, a large set of diverse data is collected. This data are then divided into modeling and validation datasets. Likewise, the acceleration data in our case was divided into modeling and validation datasets. The validation dataset was separated randomly before the application of machine learning whereas the modeling dataset was divided into training and testing datasets using the ratio of 80/20 (i.e. 80% of the data was used for training the models and 20% was used for testing the models).
- In the second step, the data was used to extract quality features that are very crucial for the performance of machine learning models. These features vary with the type of application and without extracting good features, the development of accurate machine learning models is not possible. In our case, the signal processing algorithm was used to extract the useful MI_E values for developing the high-performance machine learning models.
- In the third step, machine learning models were developed using different machine learning algorithms. The choice of machine learning algorithms is again application dependent and depends on factors including training data size, output accuracy, speed of training time, data structure, number of features, trial and error, etc. Machine learning can be supervised, unsupervised, and reinforcement type. In our case, supervised machine learning was used. Machine learning models were developed using KNN, Decision Tree, Random Forest, and Adaboost algorithms. The process involved training the models using the training dataset. This process was repeated for each machine learning algorithm.
- In the fourth step, the performance of the trained machine learning models was evaluated on the testing data and the average AUROC (area under receiver operating curve) was derived for each model. However, AUROC might be misleading for an imbalanced dataset (Wang et al. 2021), so

other performance measures including class precision, class recall, and F1 scores were computed for the testing dataset. Finally, the performance of the trained machine learning models were compared on the testing dataset and the best models were selected. The best-performing models were then applied to the validation dataset.

The overall procedure of developing machine learning models for our study is illustrated in Figure 8.

[Insert Figure 8]

5.1. Input data for machine learning models

The input of the model was 1) MI_E values acquired through the signal processing algorithm for each leak and no-leak cases in the modeling dataset, and 2) standard deviations of g (acceleration) values of the same cases. Other combinations such as MI_E values and the averages of the g values; MI_E values alone; standard deviations of the g values alone, etc. were also attempted as inputs but didn't provide desired results. An open data science software RapidMiner, specially designed for developing predictive models using machine learning, was used in our study. Out of the total of 993 cases, 816 cases (166 leak and 650 no-leak cases) were used in the modeling dataset while the remaining 177 cases were separated randomly before applying the machine learning algorithms to be used as the validation set. The dataset comprised cases from 1) metal pipes including stainless steel (SS), galvanized iron (GI), cast iron (CI), and ductile iron (DI); and 2) non-metal pipes included polyethylene (PE), Un-plasticized poly-vinyl chloride (UPVC). Pipe diameters ranged from 50mm to 300mm.

5.2. Input data pre-processing

Input data pre-processing was carried out to understand the data before applying machine learning algorithms. Since most machine learning models work on the balanced class principle, an imbalanced dataset should be preprocessed for reliable classification (Choi et al. 2020). Imbalanced datasets are typically challenging for almost all machine learning algorithms as they tend to neglect minority cases. Therefore, the imbalanced data should be resampled to produce a class-balanced dataset.

In binary classification problems, a balanced dataset can be achieved through under-sampling and oversampling (Wang et al. 2021). Under-sampling balances the dataset by deleting the majority of observations arbitrarily. However, under-sampling suffers from the risk of discarding useful cases, thus, deteriorating the majority class distribution. Oversampling doesn't suffer from such risks as oversampling increases the minority observations. The synthetic minority oversampling technique (SMOTE) is the most representative example of oversampling which generates artificial minority cases by interpolating existing minority cases (Wang et al. 2021; Choi et al. 2020). Since SMOTE can effectively overcome overfitting to large extent, this technique was applied to the modeling dataset to address the class imbalance problem. The leak cases were segregated from the modeling dataset and duplicated to attain the same number of leak and no-leak cases. Data balancing was only conducted on the modeling dataset; the validation dataset was still imbalanced.

5.3. Performance evaluation of machine learning models

For binary classification, the performance of machine learning models can be evaluated using statistics derived from the confusion matrix. This matrix basically provides information about the predicted and target classes by a machine learning algorithm and contains four entries, as depicted in Figure 9, where true positive (TP) represents the number of positive cases identified correctly; true negative (TN) represents the number of negative cases identified correctly; false positive (FP) represents the number of negative cases incorrectly identified as positives; and false-negative (FN) represents the number of positive cases incorrectly identified as negatives. From these statistics, the ROC curve can be plotted considering two metrics simultaneously, namely true positive rate (TPR) and false-positive rate (FPR). TPR, as given in equation 4, is the ratio of the number of positive cases identified correctly to the number of true positive cases, whereas FPR, as given in equation 5, is the ratio of the number of negative cases identified as positive cases to the number of true negative cases. AUROC indicates the models' capability to successfully differentiate between positive and negative cases. AUROC ranges from 0 to 1; the higher the AUROC the better the performance of the model (Wang et al. 2021) and a value lower than 0.7 is considered poor (Choi et al. 2020). The overall accuracy of the model is defined in equation 6.

$$TPR = \frac{TP}{TP + FN} \dots\dots\dots\dots (4)$$

$$FPR = \frac{TP}{FP + TN} \dots\dots\dots\dots (5)$$

$$accuracy = \frac{TP + TN}{TP + FN + FP + TN} \dots\dots\dots\dots (6)$$

[Insert Figure 9]

However, AUROC is not sufficiently sensitive for an imbalanced set, therefore, class precision, class recall, and F1 scores were computed to evaluate the classification performance. Class recall, as given in equation 7, is basically the same as TPR. Class precision, as given in equation 8, is the ratio of the number of correctly identified positive cases to the total number of positive cases.

$$class\ recall\ = TPR = \frac{TP}{TP + FN} \dots\dots\dots\dots (7)$$

$$class\ precision = \frac{TP}{TP + FP} \dots\dots\dots\dots (8)$$

The F1 score, ranging from 0 to 1, provides a harmonic mean between class precision and recall. The higher the F1 score, the better the performance of the model. This score provides information about the model's preciseness (i.e. how many cases are classified accurately) and robustness (i.e. how many hard-to-classify cases are missed) (Kampelopoulos et al. 2020b; Mishra 2018). The formula for calculating the F1 score is given in equation (9).

$$class\ recall\ = 2\ X\ \frac{1}{\frac{1}{class\ precision} + \frac{1}{class\ recall}} \dots\dots\dots\dots (9)$$

5.4. Model implementation and performance

The application of machine learning models for leak and no leak classification in water distribution networks (WDN) is relatively new. Therefore, a thorough literature review of studies involving similar applications was conducted for the selection of appropriate algorithms. For example, Fan et al. (2021) used a hypothetical EPANET hydraulic model to mimic real WDN conditions and used an artificial neural network (ANN) to train and test the model. At the DMA level, Cantos et al. (2020) developed similar leak detection ANN and support vector machine (SVM) models based on hydraulic data. Guo et al. (2021) presented another classification model focusing on efficient leakage detection under noisy conditions. The study found the time-frequency convolutional neural network (CNN) performed more accurately under different signal-to-noise (SNR) conditions as compared to other machine learning models. El-Zahab et al. (2018) used Decision Tree, SVM, and Naïve Bayes to develop leak detection models using acceleration data from the lab-based experiments. Ravichandran et al. (2021) used ensemble-based algorithms and established machine learning models for leak detection using acoustic data from WDNs in North American Cities. They further compared the performance of machine learning models with K-nearest Neighbor (KNN) and ANN. Due to the scarcity of the relevant work published in the literature on leak detection using machine learning modeling in real WDNs, it is useful to learn from other similar applications of supervised learning. Machine learning models for leak detection in gas pipelines offer an interesting reference point. Kim et al. (2021) recently presented a deep learning-based model for leak localization and leak size estimation in subsea gas pipelines. da Cruz et al. (2020) used acoustic data for developing an ensemble-based leak detection model for low-pressure gas pipelines. Likewise, Ramotsoela et al. (2019) used two machine learning-based models based on SVM and decision trees for leak detection in noisy industrial conditions. They found SVM to perform better in noisy conditions. Apart from gas pipelines, supervised learning has also been employed in condition assessment and failure prediction of sewer pipelines. Moradi et al. (2019) elaborated on the use of several machine learning and deep learning-based models for the purpose, especially CNN, for sewer defect classification, defect detection, and classification. However, training of CNN-based models is expensive.

The literature review provided ample insights into various requirements of implementing machine learning algorithms for real WDNs. For example, ANN and ensembles work well on large data and are not-so-sensitive to outliers (Kayaalp et al. 2017; Kutylowska 2015). KNN and SVM are less prone to overfitting, however, KNN is computationally more efficient (Ayadi et al. 2020; Kang et al. 2017). Decision Tree provides an easy interpretation of results through visualization (Winkler et al. 2018). Besides, all these above-mentioned algorithms are stable and robust (Alobaidi et al. 2019). Our study applied all these algorithms to check their viability in real WDNs and made a comparison of the results to establish the best models. Since signal attenuation is higher in non-metal pipes than metal pipes, therefore, it was appropriate to develop separate metal and non-metal-based models. For the sake of the development of accurate models and provide a suitable comparison, the algorithms that didn't achieve good accuracy for both metal and non-metal pipes were dropped for further analysis. For example, ANN and SVM didn't achieve acceptable accuracy for both metal and non-metal pipes whereas Naïve Bayes failed to achieve acceptable accuracy for non-metal pipes. Therefore, Decision Tree, KNN, and two Ensemble algorithms (Random Forest and Adaboost) based machine learning were implemented and compared. The best performing models can be used standalone for leak detection or can be used in a voting system i.e. a case will be declared as a leak if the majority of algorithms predict the case as a leak. 8:2 training and the testing datasets were used for each model and parameters for machine learning algorithms were kept the same for both metal and non-metal-based models.

5.4.1. Decision Tree

Decision Tree is one of the most widely used supervised learning-based algorithms. It gives a flow-chart-like structure with internal nodes, branches, and leaves (terminal nodes), where each internal node represents a test on the attribute, each branch donates the test outcome, and each leaf provides the class label. The main advantage of a decision tree lies in its easy interpretability through a visual flowchart and requires little effort for data preparation. However, decision trees tend to overfit the data (El-Zahab et al. 2018; Hamoud et al. 2018; Sharma and Kumar 2016).

The parameters for Decision Tree were set as follows: Criterion = gain_ratio; maximal depth = 10; apply pruning = yes; confidence = 0.1; minimal gain = 0.01; minimal leaf size = 2; minimal size for split = 4; number of pre-pruning alternatives = 3. The results for Decision Tree models for both non-metal and metal-based pipes are given in Tables 2 and 3, respectively. The Decision Tree approach showed a very high accuracy of 99.18% for metal pipes and didn't miss a single leak, and its class recall for leaks reached 100%. However, the Decision Tree model for non-metals pipe showed a relatively low accuracy of 84.78% and missed 14 leaks, therefore, the class recall is on the lower side at 79.71%.

To check the quality of retrieved data per class, class precisions were computed. The class precision for no-leak data is higher in the case of the non-metal-based model but was the same for the metal-based model. F1 scores almost reaching 1.00 show the robustness of the metal-based model. Figure 10 shows the Decision Tree model for metal and non-metal pipes.

[Insert Table 2]

[Insert Table 3]

[Insert Figure 10]

5.4.2. K-Nearest Neighbour

KNN predicts the category of new data, on the basis of similarities with K training data which are nearest to new data (Quy et al. 2019; Soldevila et al. 2017). Afterward, KNN places the new data in the category with the highest probability. KNN is typically very effective for large datasets and provides good results even for small datasets if the data is noise-free and labeled, however, finding the K number that provides the highest accuracy can be complex and time-consuming (Fereidooni et al. 2021; Ravichandran et al. 2021).

The parameters for KNN models were set as follows: K = 5; weighted vote = yes; measure types = mixed measures; mixed measures = mixed Euclidean distance. The results for KNN models for metal and non-metal pipes are given in Tables 2 and 3, respectively. The accuracy of the metal-based KNN model was exactly 96.72%% and the model classified both leak (except 1) and no-leak cases (except 3) precisely. For the non-metal-based model, the accuracy of the model was 89.86%. KNN, however, missed 8 leaks, and therefore its class recall 88.41%. The quality of retrieved data per class was similar in the case of no-leak and leak cases as depicted from class precisions of the non-metal-based model. In terms of F1 scores, the model showed similar performance for both leak and no-leak cases.

5.4.3. Ensembles

Ensemble modeling is the process of generating diverse models and using their combinations to predict the outcomes (Rayaroth and Sivaradje 2019; Kayaalp et al. 2019). Ensembles techniques are typically deployed to boost performance or reduce the likelihood of errors from one single model (Ravichandran et al. 2021; Shirzad and Safari 2019). Our study adopted two popular ensemble learning algorithms: Random Forest and Adaboost. Random Forest relies on multiple decision trees, built from the sub-set of the dataset. Random forest takes votes on predictions from each tree and based on the majority votes predicts the final outcome (Shirzad and Safari 2019; Butterfield et al. 2018). The parameters for Random Forest were set as follows: Number of trees = 100; criterion = gain ratio; maximal depth = 10; apply pruning = no; random splits = no; guess subset ratio = yes; voting strategy = confidence vote. Adaboost, on the other hand, assists in combining weak classifiers into one single strong classifier, by putting more weight on the cases that are difficult to handle than on the cases which are already handled well (Desarda 2019). Our study attempted Adaboost using Decision Tree as a weak classifier following the suggestion of Thongkam et al. (2008) and Kim and Upnega (2014). The parameters for Adaboost (Decision Tree) were set as follows: Iterations = 10; Criterion = gain_ratio; maximal depth = 10; apply pruning = yes; confidence = 0.1; minimal gain = 0.01; minimal leaf size = 2; minimal size for split = 4; number of pre-pruning alternatives = 3.

Both ensemble models, Random Forest and Adaboost (Decision Tree) increased the accuracy of individual Decision Tree models for non-metal pipes. The accuracy of Random Forest and Adaboost (Decision Tree) came out to be 94.93% and 94.20%, respectively. However, Random Forest increased the class recall to

91.30% and the model only missed 6 leaks. Adaboost (Decision Tree), on the other hand, missed 10 leaks. Therefore, its class recall for leaks was 85.51%. As a matter of fact, KNN missed a fewer number of leaks than Adaboost and its class recall is higher at 88.41%. In order to check the overfitting, F1 scores were computed which also showed extremely good results. Again, Random Forest showed better results in F1 matric among all the models.

For metal-based pipes, both machine learning ensemble models performed remarkably well. The accuracy of Random Forest reached 100%. The accuracy of Adaboost (Decision) came out to be 99.18%, exactly similar to the individual Decision Tree model and Adaboosting didn't help in increasing the accuracy. None of the ensemble models and Decision Tree missed any leak. Decision Tree and Adaboost (Decision Tree) missed one no-leak case.

5.5. Comparative ROC

Since SMOTE was used for data balancing, comparative ROC curves were plotted to further confirm the performance of the models (Figure 11). Figure 11(a) shows that the AUROC values for all four metal-based models were higher than 0.9. Among the non-metal-based models, the AUROC value for the Decision Tree model was less than 0.9, relatively lower than the other three models. Therefore, this model was dropped for the validation analysis. This omission was also justified from the accuracy histograms in Figure 12 which clearly shows the inferior performance of the Decision Tree non-metal-based model. In addition, this model missed the highest number of leaks in comparison to other models (Figure 13) and caused the highest number of false alarms (Figure 14).

In terms of comparative performance, Random Forest achieved the highest accuracy among both metal and non-metal-based models. All four metal-based models performed equally well in detecting leaks, as three of the models, except KNN, didn't even miss a single leak. However, Random Forest missed the lowest number of leaks among non-metal models. Metal-based models in general performed better than non-metal-based models.

[Insert Figure 11]

[Insert Figure 12]

[Insert Figure 13]

[Insert Figure 14]

5.6. Model correlations and sensitivity

The relation between independent variables and dependent variables can be checked through correlation coefficients. A positive correlation means that by increasing the value of the independent variable, the value of the dependent variable increases. In the case of binary classification (leak and no-leak), it means that by increasing the value of the independent variable the no-leak state changes to the leak state. Both independent variables (standard deviation and MIE values) are positively correlated to the dependent variable which shows the importance of both variables in detecting leaks. However, the correlation coefficient (0.211) of the MIE values is a bit higher than the correlation coefficient (0.140) of the standard deviation values.

To check the importance of independent variables and model sensitivity, the models' accuracy and the number of leaks missed were checked by training models using one independent variable at one time (i.e. models were developed using MIE values alone and standard deviation values alone) and the results were compared with models trained using both independent variables. From Table 4, it is clear that the models built on both MIE and standard deviation values together outperformed the model trained using single variables. Standard deviation-based models, nevertheless, performed better than MIE-based models. In fact, the standard deviation Adaboost model, in the case of non-metal pipes, missed fewer leaks in comparison to the original MIE plus standard deviation model, however, the overall accuracy was lower.

[Insert Table 4]

6. Validation of Models and Utilization

The models were validated through the validation of 177 cases which were separated from the model after the data collection. Data balancing was not conducted on the validation set, which contained 102 metal cases and 77 non-metal cases. Validation of the metal cases was carried out using all four models, however, for the validation of non-metal cases, the Decision Tree model was omitted due to its low accuracy on different performance metrics. Machine learning models were not supplied with leak and no-leak labels for the validation set and the accuracy of the validation set was later checked by comparing with the real leak states and the results were computed manually.

The results of both metals and non-metals pipes are given in Table 5. It is interesting to note that all models correctly identified no-leak cases for metal pipes. For metal pipes, although the accuracy of all four models was similar on the testing set, KNN detected the highest number of leaks accurately and only missed one leak. All the other three models missed three leaks and surprisingly all of them failed to correctly identify the state of the same three cases (IDs = 957, 958, and 959). KNN, however, was successful in identifying the leak state for cases 957 and 958, however, failed to identify case 959 with 59% confidence. For non-metal pipes, Random Forest and Adaboost (Decision Tree) correctly identified all leak and no leak cases. KNN also identified all leak cases correctly, however, failed to identify five no-leak cases correctly.

In Hong Kong, noise loggers lead to a high percentage of false alarms, achieving such a high accuracy with accelerometers in detecting leaks is quite promising. Both metal and non-metals were unable to classify only a very few cases (considering both training and validation set) which is acceptable as this classification accuracy can be enhanced by looking at the results of different models simultaneously and continuous training of models with new data. The accuracy of non-metal pipes can be further increased by developing models for individual pipe materials such as polyethylene-based models.

[Insert Table 5]

7. Conclusions, Limitations, and Future Works

This research investigated the application of MEMS-based accelerometers for leak detection in real pipe networks. The leak detection problem was formulated as a binomial problem to identify the leak and no-leak cases using acceleration data. Firstly, a Monitoring index-based algorithm was used for signal analysis and to distinguish between 993 leak and no-leak cases collected from 75 different sites in Hong Kong. The cases were collected from sites of different pipe materials and types. Secondly, monitoring index efficiencies obtained from the data analysis and standard deviations were used to demonstrate the effectiveness of machine learning models in detecting leaks. 816 cases were utilized to develop individual [KNN and Decision Tree] and ensemble (Random Forest and Adaboost (Decision Tree)] machine learning models. The remaining 177 cases were used as a validation set to further verify the performance of the models.

All types of individual and ensemble metal-based models acquired very high overall accuracy in leak detection; Random Forest, in fact, reached 100% accuracy. Apart from 1 missed leak by KNN, all the other models accurately identified all the leaks. However, on the validation set, KNN performed the best and missed only one leak. The other three models missed 3 leaks each.

Among the non-metal pipes, ensembles models performed well and the accuracy reached over 94% for both the models. Among individual models, KNN reached almost 90% accuracy. Decision Tree, however, didn't perform so well and reached around 85% accuracy and hence, drop for subsequent validation analysis. KNN, Random Forest, and Adaboost (Decision Tree) were also unable to classify some leak cases, with Random Forest missed the lowest number of leak cases at 6. The ensemble models performed remarkably well in detecting no-leak cases and only provided one false alarm each. The performance of KNN and ensemble models was further confirmed with a validation set. Ensemble models worked perfectly with the validation set, and KNN led to five false alarms. From the results, it can be safely deduced that all the selected metal and non-metal-based models can be extremely effective in detecting leaks using MEMS accelerometers in real networks.

MEMS-based accelerometers provide a cheap technology in comparison to noise loggers, and our study has proven the effectiveness of this technology in leak detection, however, there are certain limitations that might hinder the application of MEMS accelerometers-based leak detection models in water networks. Firstly, the accelerometers used in our study were battery-powered devices and a single accelerometer can only provide data for 30 minutes after charging. Therefore, these devices can only be used for temporary monitoring and require personnel to replace accelerometers every night. Secondly, the accelerometers can't be placed very far from the gateway unless a long-distance transmission antenna is used which might be costly. Basically, the technology used in our study needs further maturity in terms of wireless connectivity and data transfer. The third limitation is about the leak detection models themselves as the models were developed using monitoring indexes and standard deviations. The models didn't use traditional features such as spread, level, frequency centroid, etc. The impact of the inclusion of such features on the models' accuracy needs to be checked. Fourthly, ideally, separate models should have been developed using the dataset for each pipe type and diameter. However, due to the limitations on the data collection, two types of models, metal and non-metal-based models, were developed and tested on the different types of pipes and materials. Although the model succeeded in detecting leaks for a vast majority of cases, however, its accuracy can be further enhanced by developing individual models and administrating a large-scale validation.

Considering the limitations, future work will be conducted on developing feature-based machine learning models for accelerometers. These models will use the traditional features such as spread, level, maximum amplitude, peak frequency, kurtosis, etc., extracted from the acceleration signals collected in this study. The accuracy of the developed models will be compared with the models presented in this study. Later, feature-based machine learning models and the models presented in this study will be combined to further improve the leak detection accuracy and performance.

Acknowledgment

The authors gratefully acknowledge the support from the Innovation and Technology Fund (Innovation and Technology Support Programme (ITSP)), Hong Kong, and the Water Supplies Department, Hong Kong, under grant number ITS/067/19FP. The authors would also like to thank the two anonymous reviewers for their constructive and encouraging comments.

References

Alobaidi, M. H., Meguid, M. A., & Chebana, F. (2019). "Predicting seismic-induced liquefaction through ensemble learning frameworks". *Scientific Reports, 9*(1), 1-12.

Ayadi, A., Ghorbel, O., BenSalah, M. S., & Abid, M. (2020). "Spatio-temporal correlations for damages identification and localization in water pipeline systems based on WSNs". *Computer Networks, 171*, 107134.

Bakhtawar, B., & Zayed, T. (2021). "Review of Water Leak Detection and Localization Methods through Hydrophone Technology". *Journal of Pipeline Systems Engineering and Practice*, *12*(4), 03121002.

Boretti, A., and Rosa, L. (2019). "Reassessing the projections of the world water development report". *NPJ Clean Water*, *2*(1), 1-6.

Butterfield, J. D., Meyers, G., Meruane, V., Collins, R. P., and Beck, S. B. (2018). "Experimental investigation into techniques to predict leak shapes in water distribution systems using vibration measurements". *Journal of Hydroinformatics*, *20*(4), 815-828.

Cantos, W. P., Juran, I., and Tinelli, S. (2020). "Machine-Learning–Based Risk Assessment Method for Leak Detection and Geolocation in a Water Distribution System". *Journal of Infrastructure Systems*, *26*(1), 04019039.

Choi, J., Gu, B., Chin, S., & Lee, J. S. (2020). "Machine learning predictive model based on national data for fatal accidents of construction workers". *Automation in Construction*, *110*, 102974.

da Cruz, R. P., da Silva, F. V., and Fileti, A. M. F. (2020). "Machine learning and acoustic method applied to leak detection and location in low-pressure gas pipelines". *Clean Technologies and Environmental Policy*, *22*(3), 627-638.

Desarda, A. (2019). Understanding Adaboost. Towards Data Science Inc., Ontario. https://towardsdatascience.com/understanding-adaboost-2f94f22d5bfe

El-Zahab, S., Abdelkader, E. M., and Zayed, T. (2018). "An accelerometer-based leak detection system". *Mechanical Systems and Signal Processing*, *108*, 276-291.

El-Zahab, S., and Zayed, T. (2019). "Leak detection in water distribution networks: an introductory overview". *Smart Water*, *4*(1), 1-23.

El-Zahab, S., Mosleh, F., and Zayed, T. (2016). An accelerometer-based real-time monitoring and leak detection system for pressurized water pipelines. In *Pipelines 2016*, 17-20 July, Kansas City, 257-268.

Fan, X., Zhang, X., and Yu, X. B. (2021). "Machine learning model and strategy for fast and accurate detection of leaks in water supply network". *Journal of Infrastructure Preservation and Resilience*, *2*(1), 1-21.

Fereidooni, Z., Tahayori, H., and Bahadori-Jahromi, A. (2021). "A hybrid model-based method for leak detection in large scale water distribution networks". *Journal of Ambient Intelligence and Humanized Computing*, *12*(2), 1613-1629.

Gao, Y., Brennan, M., Joseph, P. F., Muggleton, J. M., and Hunaidi, O. (2005). "On the selection of acoustic/vibration sensors for leak detection in plastic water pipes". *Journal of Sound and Vibration*, *283*(3-5), 927-941.

Gupta, A. (2011). Hong Kong is wasting a third of its water. China Dialogue, London. https://chinadialogue.net/en/cities/9803-hong-kong-is-wasting-a-third-of-its-water/

Guo, G., Yu, X., Liu, S., Ma, Z., Wu, Y., Xu, X., Wang, X., Smith, K., Wu, X. (2021). "Leakage Detection in Water Distribution Systems Based on Time–Frequency Convolutional Neural Network". *Journal of Water Resources Planning and Management*, *147*(2), 04020101.

Hamilton, S., Charalambous, B. (2013). Leak detection. IWA Publishing, London.

Hamoud, A., Hashim, A. S., and Awadh, W. A. (2018). "Predicting student performance in higher education institutions using decision tree analysis". *International Journal of Interactive Multimedia and Artificial Intelligence*, *5*, 26-31.

Hu, Z., Tariq, S., and Zayed, T. (2021). A comprehensive review of acoustic based leak localization method in pressurized pipelines. *Mechanical Systems and Signal Processing*, *161*, 107994.

Hunaidi, O., and Chu, W. T. (1999). "Acoustical characteristics of leak signals in plastic water distribution pipes". *Applied Acoustics*, *58*(3), 235-254.

Ismail, M. I. M., Dziyauddin, R. A., Salleh, N. A. A., Muhammad-Sukki, F., Bani, N. A., Izhar, M. A. M., and Latiff, L. A. (2019). "A review of vibration detection methods using accelerometer sensors for water pipeline leakage". *IEEE Access*, *7*, 51965-51981.

Kampelopoulos, D., Papastavrou, G. N., Kousiopoulos, G. P., Karagiorgos, N., Porlidas, D., and Nikolaidis, S. (2020a). Defining thresholds for leak detection parameters through statistical analysis of the noise in water/gas pipelines. In *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 25-28 May, Dubrovnik, 1-6.

Kampelopoulos, D., Papastavrou, G. N., Kousiopoulos, G. P., Karagiorgos, N., Goudos, S. K., and Nikolaidis, S. (2020b). Machine Learning Model Comparison for Leak Detection in Noisy Industrial Pipelines. In *2020 9th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, 7-9 September, Bremen, 1-4.

Kanakoudis, V., and Muhammetoglu, H. (2014). "Urban water pipe networks management towards non-revenue water reduction: Two case studies from Greece and Turkey". *CLEAN–Soil, Air, Water*, *42*(7), 880-892.

Kang, J., Park, Y.-J., Lee, J., Wang, S.-H., & Eom, D.-S. (2017). "Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution systems". *IEEE Transactions on Industrial Electronics*, *65*(5), 4279-4289.

Kayaalp, F., Zengin, A., Kara, R., & Zavrak, S. (2017). "Leakage detection and localization on water transportation pipelines: a multi-label classification approach". *Neural Computing and Applications*, *28*(10), 2905-2914.

Kutyłowska, M. (2015). "Neural network approach for failure rate prediction". *Engineering Failure Analysis*, *47*, 41-48.

Kim, S. Y., & Upneja, A. (2014). "Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models". *Economic Modelling*, *36*, 354-362.

Kim, J., Chae, M., Han, J., Park, S., and Lee, Y. (2021). "The development of leak detection model in subsea gas pipeline using machine learning". *Journal of Natural Gas Science and Engineering*, *94*, 104134.

Marmarokopos, K., Doukakis, D., Frantziskonis, G., and Avlonitis, M. (2018). “Leak detection in plastic water supply pipes with a high signal-to-noise ratio accelerometer”. *Measurement and Control*, *51*(1-2), 27-37.

Martini, A., Troncossi, M., and Rivola, A. (2015). “Automatic leak detection in buried plastic pipes of water supply networks by means of vibration measurements”. *Shock and Vibration*, *2015*, 1-13.

Martini, A., Troncossi, M., and Rivola, A. (2017). “Leak detection in water-filled small-diameter polyethylene pipes by means of acoustic emission measurements”. *Applied Sciences*, *7*(1), 2.

Mishra, A. (2018). Metrics to evaluate your machine learning algorithm. Towards Data Science Inc., Ontario. https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234

Moradi, S., Zayed, T., and Golkhoo, F. (2019). “Review on computer aided sewer pipeline defect detection and condition assessment”. *Infrastructures*, *4*(1), 10.

Mostafapour, A., and Davoudi, S. (2013). “Analysis of leakage in high pressure pipe using acoustic emission method”. *Applied Acoustics*, *74*(3), 335-342.

Okosun, F., Cahill, P., Hazra, B., and Pakrashi, V. (2019). “Vibration-based leak detection and monitoring of water pipes using output-only piezoelectric sensors”. *The European Physical Journal Special Topics*, *228*(7), 1659-1675.

Quy, T. B., Muhammad, S., & Kim, J. M. (2019). “A reliable acoustic EMISSION based technique for the detection of a small leak in a pipeline system”. *Energies*, *12*(8), 1472.

Ramotsoela, D. T., Hancke, G. P., and Abu-Mahfouz, A. M. (2019). “Attack detection in water distribution systems using machine learning”. *Human-centric Computing and Information Sciences*, *9*(1), 1-22.

Ravichandran, T., Gavahi, K., Ponnambalam, K., Burtea, V., and Mousavi, S. J. (2021). “Ensemble-based machine learning approach for improved leak detection in water mains”. *Journal of Hydroinformatics*, *23*(2), 307-323.

Rayaroth, R., & Sivaradje, G. (2019). “Random bagging classifier and shuffled frog leaping based optimal sensor placement for leakage detection in WDS”. *Water Resources Management*, *33*(9), 3111-3125.

Sharma, H., and Kumar, S. (2016). “A survey on decision tree algorithms of classification in data mining”. *International Journal of Science and Research (IJSR)*, *5*(4), 2094-2097.

Shirzad, A., and Safari, M. J. S. (2019). “Pipe failure rate prediction in water distribution networks using multivariate adaptive regression splines and random forest techniques”. *Urban Water Journal*, *16*(9), 653-661.

Soldevila, A., Fernandez-Canti, R. M., Blesa, J., Tornil-Sin, S., & Puig, V. (2017). “Leak localization in water distribution networks using Bayesian classifiers”. *Journal of Process Control*, *55*, 1-9.

Tariq, S., Hu, Z., and Zayed, T. (2020). “Micro-electromechanical systems-based technologies for leak detection and localization in water supply networks: A bibliometric and systematic review”. *Journal of Cleaner Production*, *289*, 125751.

Thongkam, J., Xu, G., and Zhang, Y. (2008). AdaBoost algorithm with random forests for predicting breast cancer survivability. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 1-8 June, Hong Kong, 3062-3069.

UN-WWAP (United Nations World Water Assessment Programme). (2018). The United Nations World Water Development Report 2018: Nature-Based Solutions for Water. UNESCO, Paris. https://www.unwater.org/publications/world-water-development-report-2018/

USEPA (United States Environmental Protection Authority). (2010). Control and Mitigation of Drinking Water Losses in Distribution Systems. USEPA, Washington DC.

Van Zyl, J., and Clayton, C. R. I. (2007). "The effect of pressure on leakage in water distribution systems". *Proceedings of the Institution of Civil Engineers-Water Management*, *160*(2), 109-114.

Wang, Y., Shao, Z., and Tiong, R. L. (2021). Data-Driven Prediction of Contract Failure of Public-Private Partnership Projects. *Journal of Construction Engineering and Management*, *147*(8), 04021089.

Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W., & Tscheikner-Gratl, F. (2018). "Pipe failure modelling for water distribution networks using boosted decision trees". *Structure and Infrastructure Engineering*, *14*(10), 1402-1411.

WSD (Water Supplies Department). (2020). Replacement and Rehabilitation Programme of Water Mains. WSD, Hong Kong. https://www.wsd.gov.hk/en/core-businesses/major-infrastructure-projects/r-r-projects/index.html

Yue, D. P., and Tang, S. L. (2011). "Sustainable strategies on water supply management in Hong Kong". *Water and Environment Journal*, *25*(2), 192-199.

List of Tables

Table 1: Monitoring indexes for the no-leak state of various pipe type

Pipe type	Pipe diameter (mm)	MI_o
Polyethylene	50	0.005133
Polyethylene	80	0.001192
Polyethylene	100	0.001162
Polyethylene	150	0.001434
Polyethylene	200	0.001549
Galvanized Iron	50	0.001352
Galvanized Iron	80	0.001614
Galvanized Iron	100	0.001385
Ductile Iron	100	0.001546
Ductile Iron	150	0.001225
Ductile Iron	300	0.001221
Stainless Steel	100	0.001254
Un-plasticized Polyvinyl Chloride	40	0.006291
Un-plasticized Polyvinyl Chloride	100	0.001138
Asbestos Cement	150	0.001126

Table 2: Leak detection results for metal-based models

Models	Results				
KNN	Accuracy=96.72%	True No Leak	True Leak	Class Precision	F1 score
	Predicted as No Leak	58	1	98.31%	0.97
	Predicted as Leak	3	60	95.24%	0.97
	Class Recall (%)	95.08%	98.36%	-	-
Decision Tree	Accuracy=99.18%	True No Leak	True Leak	Class Precision	F1 score
	Predicted as No Leak	60	0	100%	0.99
	Predicted as Leak	1	61	98.39%	0.99
	Class Recall (%)	98.36%	100%	-	-
Random Forest	Accuracy =100%	True No Leak	True Leak	Class Precision	F1 score
	Predicted as No Leak	61	0	100%	1.00
	Predicted as Leak	0	61	100%	1.00
	Class Recall (%)	100%	100%	-	-
Adaboost (Decision Tree)	Accuracy=99.18%	True No Leak	True Leak	Class Precision	F1 score
	Predicted as No Leak	61	0	100%	0.99
	Predicted as Leak	1	61	100%	0.99
	Class Recall (%)	98.36%	100%	-	-

Table 3: Leak detection results for non-metal-based models

Models	**Results**				
KNN	Accuracy=89.86%	True No Leak	True Leak	Class Precision	F1 score
	Predicted as No Leak	63	8	88.73%	0.90
	Predicted as Leak	6	61	91.04%	0.90
	Class Recall (%)	91.30%	88.41%	-	-
Decision Tree	Accuracy=84.78%	True No Leak	True Leak	Class Precision	F1 score
	Predicted as No Leak	62	14	81.58%	0.85
	Predicted as Leak	7	55	88.71%	0.84
	Class Recall (%)	89.86%	79.71%	-	-
Random Forest	Accuracy =94.93%	True No Leak	True Leak	Class Precision	F1 score
	Predicted as No Leak	68	6	91.89%	0.95
	Predicted as Leak	1	63	98.44%	0.95
	Class Recall (%)	98.55%	91.30%	-	-
Adaboost (Decision Tree)	Accuracy=94.20%	True No Leak	True Leak	Class Precision	F1 score
	Predicted as No Leak	68	10	87.18%	0.93
	Predicted as Leak	1	59	98.33%	0.92
	Class Recall (%)	98.55%	85.51%	-	-

Table 4: Comparative analysis of models developed by changing the number of attributes

Model	Missed				Accuracy		
	Type	MIE	Std	MIE+Std	MIE	Std	MIE+Std
KNN	Metal	4	3	1	95.08%	95.90%	96.72%
Decision Tree	Metal	5	2	0	95.90%	96.72%	99.18%
Random Forest	Metal	3	2	0	95.90%	97.54%	100%
Adaboost (Decision Tree)	Metal	5	2	0	95.90%	96.72%	99.18%
KNN	Non-metal	13	12	8	86.96%	89.13%	89.86%
Random Forest	Non-metal	16	9	6	84.78%	91.30%	94.93%
Adaboost (Decision Tree)	Non-metal	15	8	10	86.96%	92.03%	94.20%

Table 5: Validation set results

Models	Results				
Metal-based models					
KNN	Accuracy=99.02%	True No Leak	True Leak	Class Precision	F1 score
	Predicted as No Leak	39	1	97.50%	0.99
	Predicted as Leak	0	62	100%	1.00
	Class Recall (%)	100%	98.41%	-	-
Decision Tree	Accuracy=97.06%	True No Leak	True Leak	Class Precision	F1 score
	Predicted as No Leak	39	3	92.86%	0.96
	Predicted as Leak	0	60	100%	0.98
	Class Recall (%)	100%	95.24%	-	-
Random Forest	Accuracy=97.06%	True No Leak	True Leak	Class Precision	F1 score
	Predicted as No Leak	39	3	92.86%	0.96
	Predicted as Leak	0	60	100%	0.98
	Class Recall (%)	100%	95.24%	-	-
Adaboost (Decision Tree)	Accuracy=97.06%	True No Leak	True Leak	Class Precision	F1 score
	Predicted as No Leak	39	3	92.86%	0.96
	Predicted as Leak	0	60	100%	0.98
	Class Recall (%)	100%	95.24%	-	-
Non-metal-based models					
KNN	Accuracy=93.33%	True No Leak	True Leak	Class Precision	F1 score
	Predicted as No Leak	19	0	100%	0.88
	Predicted as Leak	5	51	91.07%	0.95
	Class Recall (%)	79.17%	100%	-	-
Random Forest	Accuracy=100%	True No Leak	True Leak	Class Precision	F1 score
	Predicted as No Leak	24	0	100%	1
	Predicted as Leak	0	51	100%	1
	Class Recall (%)	100%	100%	-	-
Adaboost (Decision Tree)	Accuracy=100%	True No Leak	True Leak	Class Precision	F1 score
	Predicted as No Leak	24	0	100%	1
	Predicted as Leak	0	51	100%	1
	Class Recall (%)	100%	100%	-	-

List of Figures

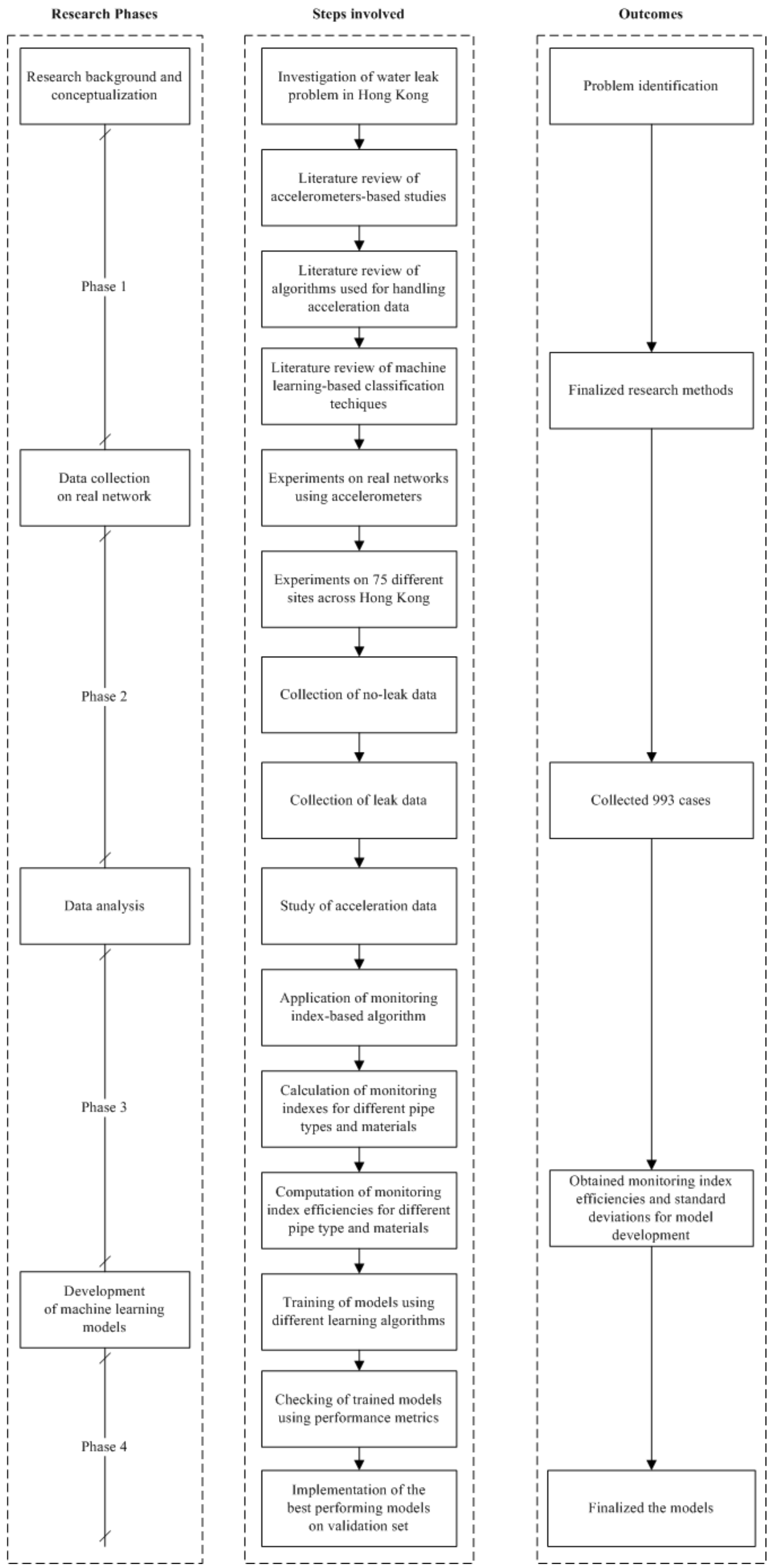

Figure 1. Research methodology

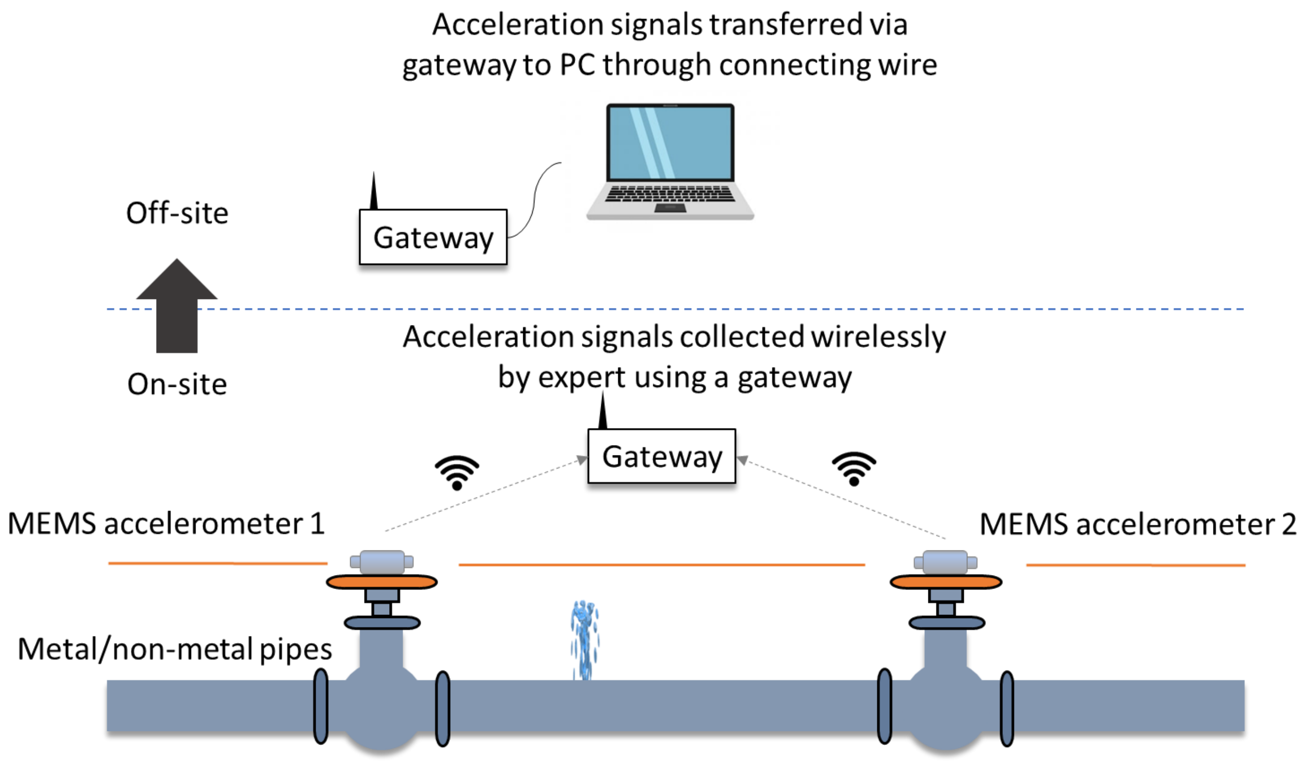

Figure 2. MEMS accelerometers-based data acquisition system

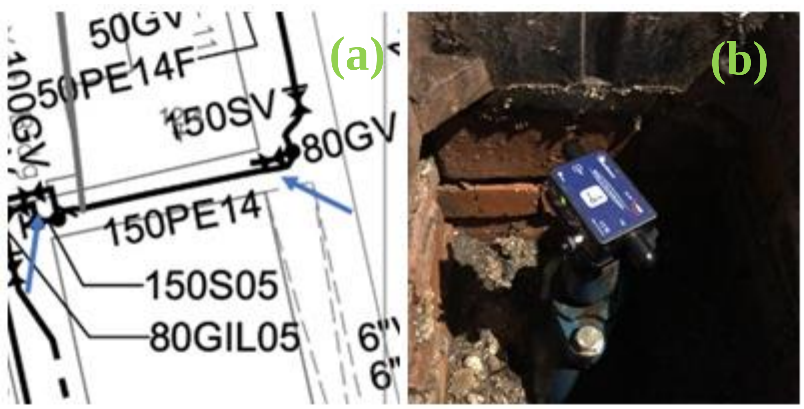

Figure 3. (a) Map showing leak on 150PE pipe - (b) Placement of accelerometer on gate valve (80GV)

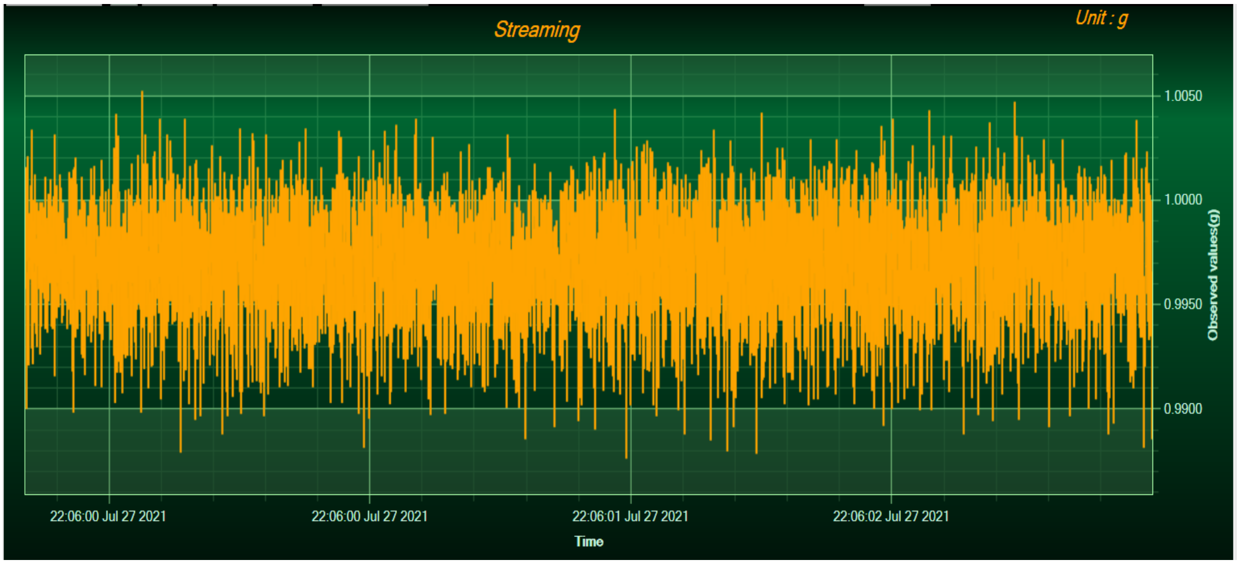

Figure 4. An example of acceleration signal acquired through the data acquisition system

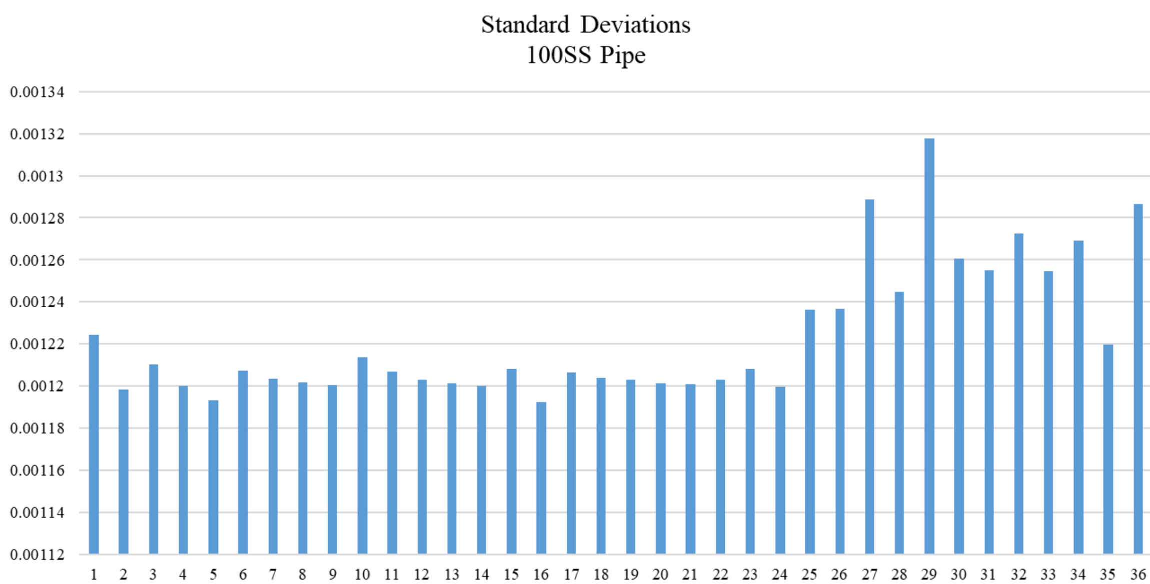

Figure 5. Standard deviations for an hour-long no-leak data (100SS pipe)

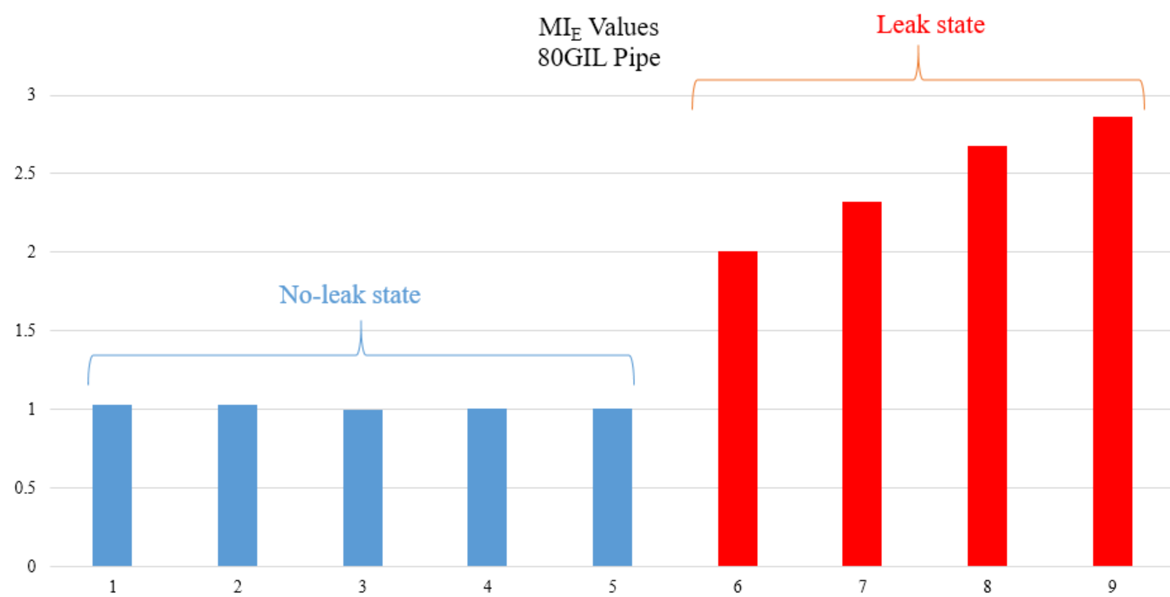

Figure 6. An example of monitoring index efficiencies (80GIL pipe)

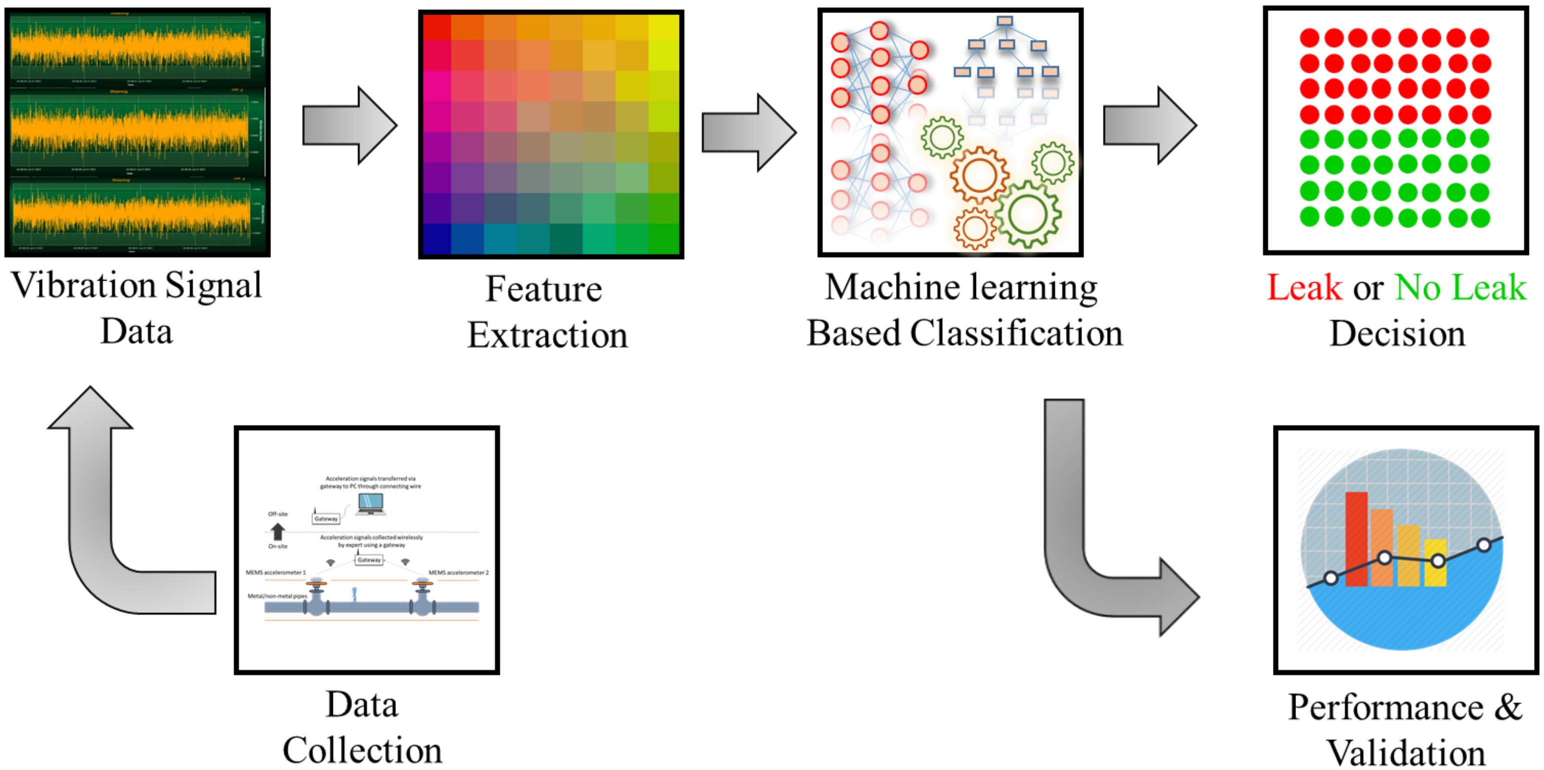

Figure 7. Framework for machine learning-based binary classification process

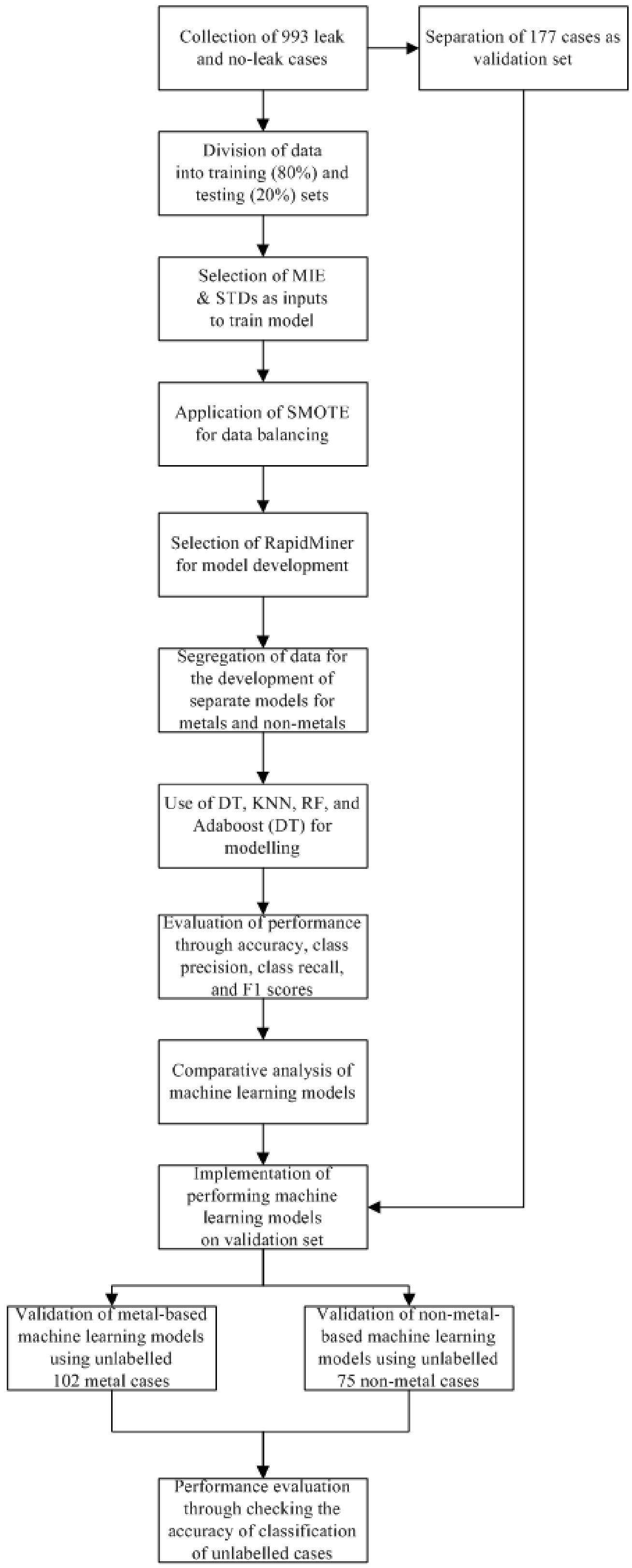
Collection of 993 leak and no-leak cases
Separation of 177 cases as validation set
Division of data into training (80%) and testing (20%) sets
Selection of MIE & STDs as inputs to train model
Application of SMOTE for data balancing
Selection of RapidMiner for model development
Segregation of data for the development of separate models for metals and non-metals
Use of DT, KNN, RF, and Adaboost (DT) for modelling
Evaluation of performance through accuracy, class precision, class recall, and F1 scores
Comparative analysis of machine learning models
Implementation of performing machine learning models on validation set
Validation of metal-based machine learning models using unlabelled 102 metal cases
Validation of non-metal-based machine learning models using unlabelled 75 non-metal cases
Performance evaluation through checking the accuracy of classification of unlabelled cases

Figure 8. Development of machine learning models for leak detection

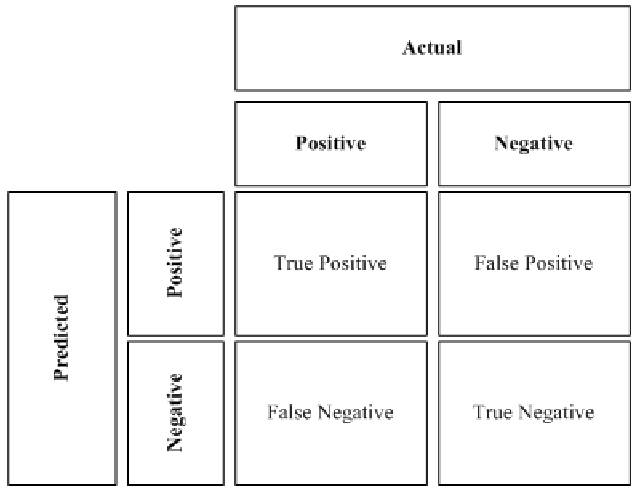

Figure 9. Confusion Matrix

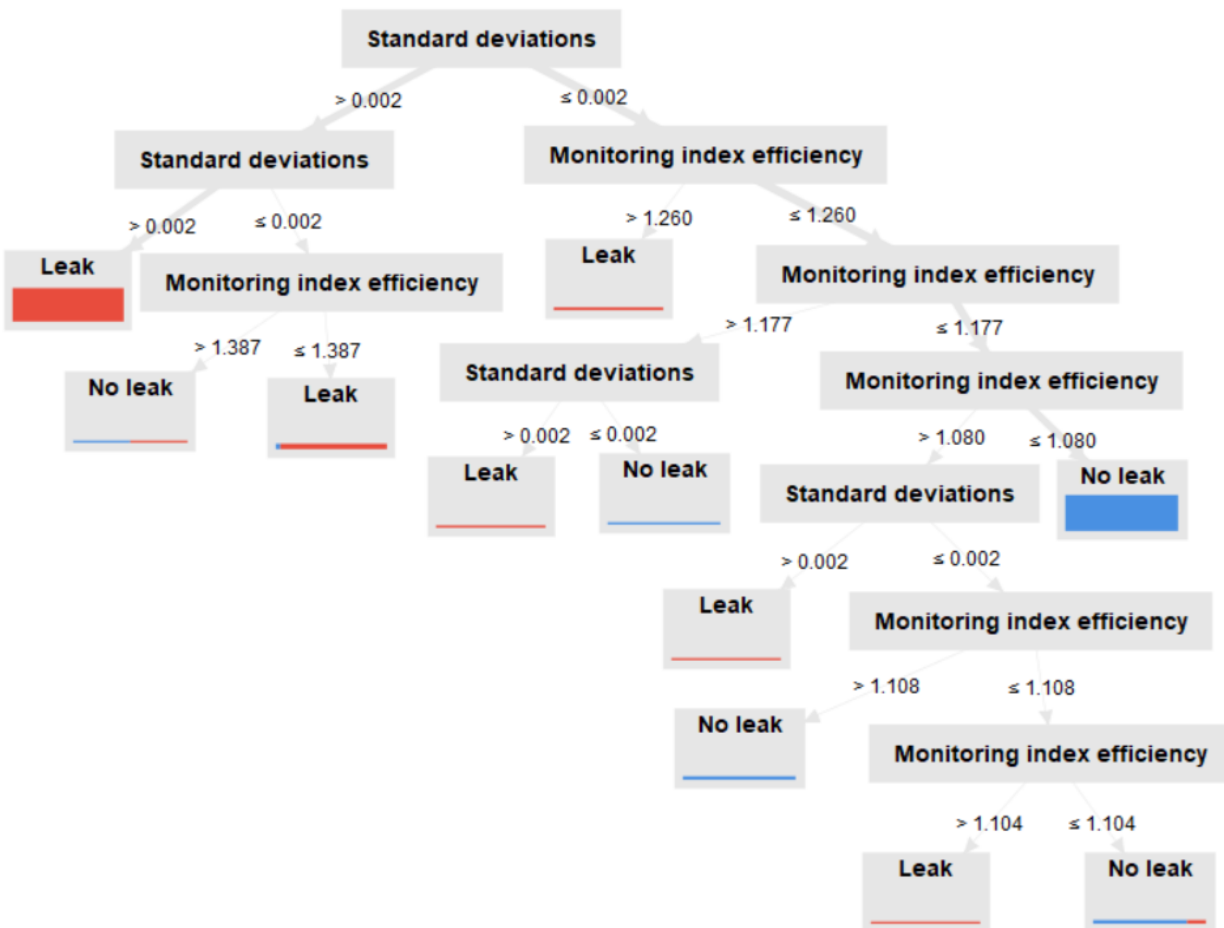

Figure 10 (a). Decision Tree model for metal pipes

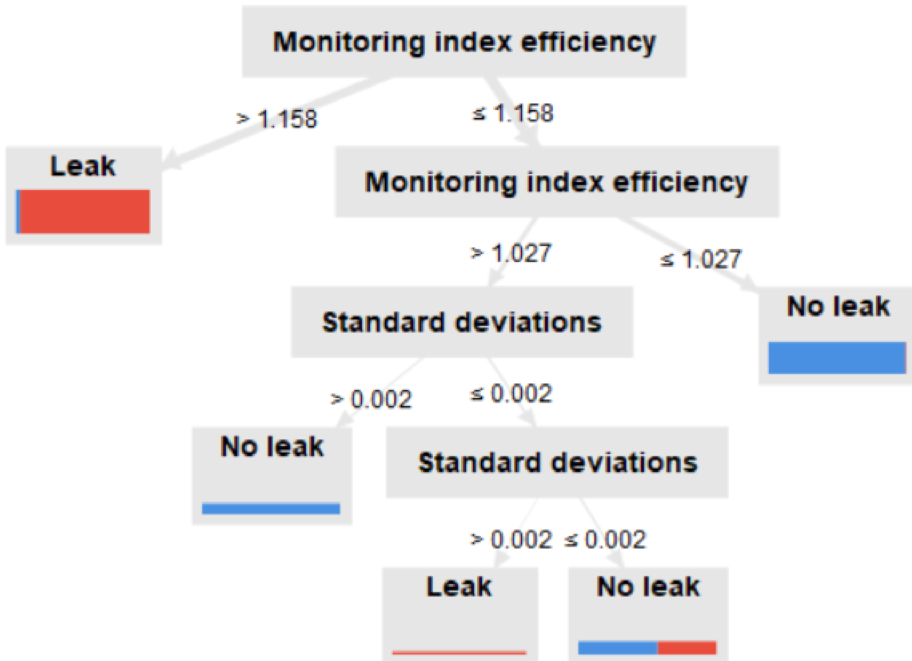

Figure 10 (b). Decision Tree model for non-metal pipes

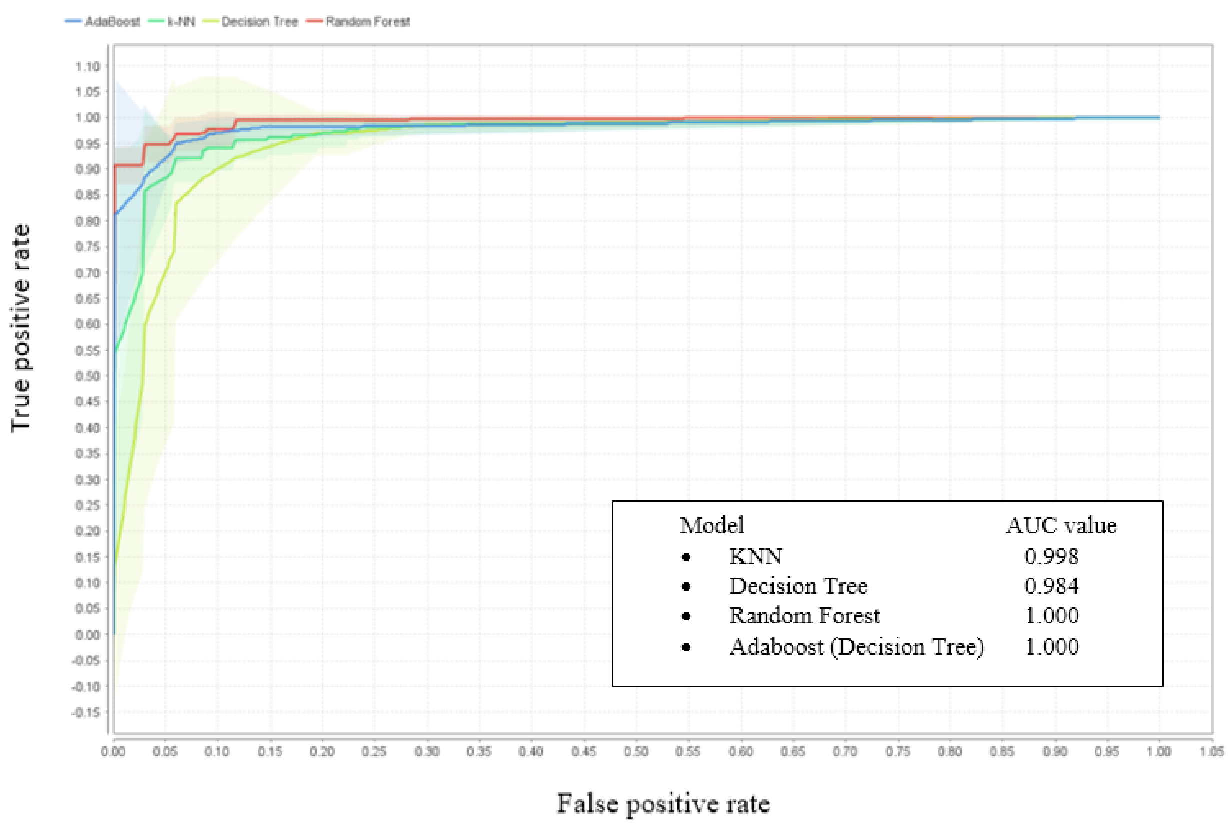

Figure 11 (a). AUROC for metal-based models

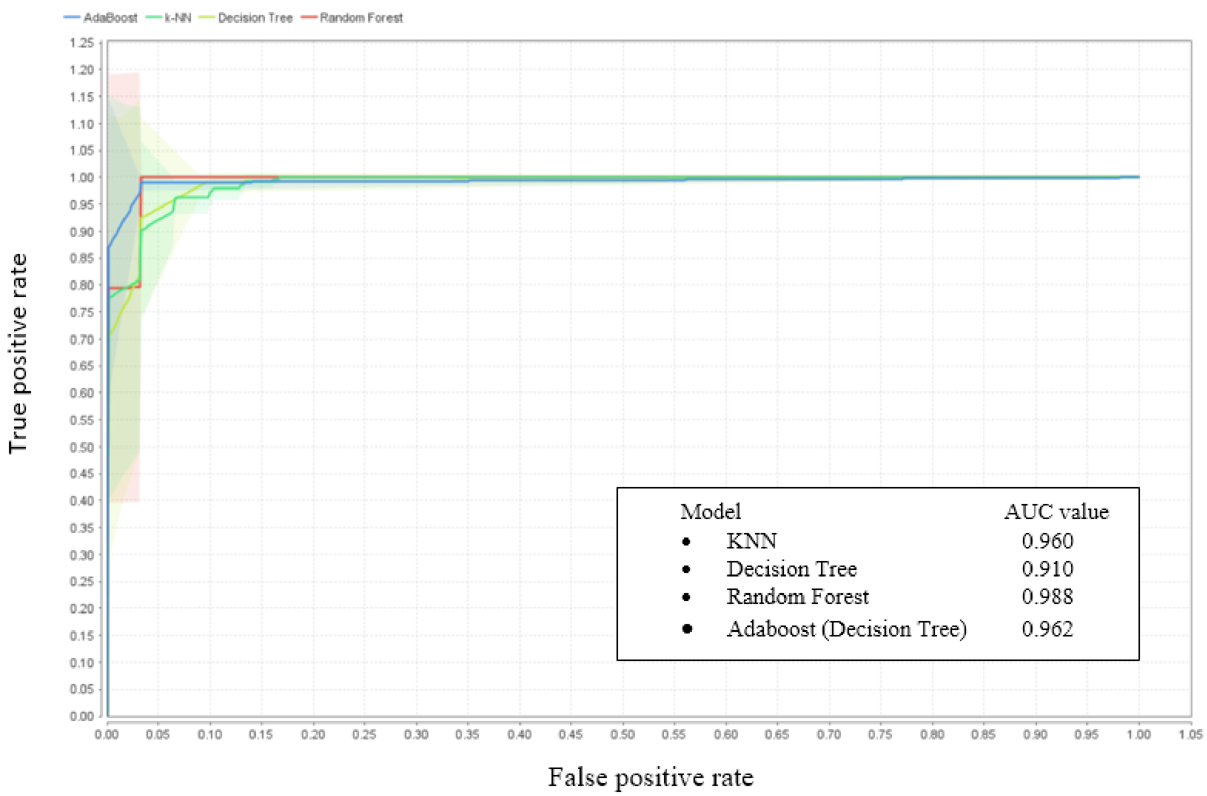

Figure 11(b). AUROC for non-metal-based models

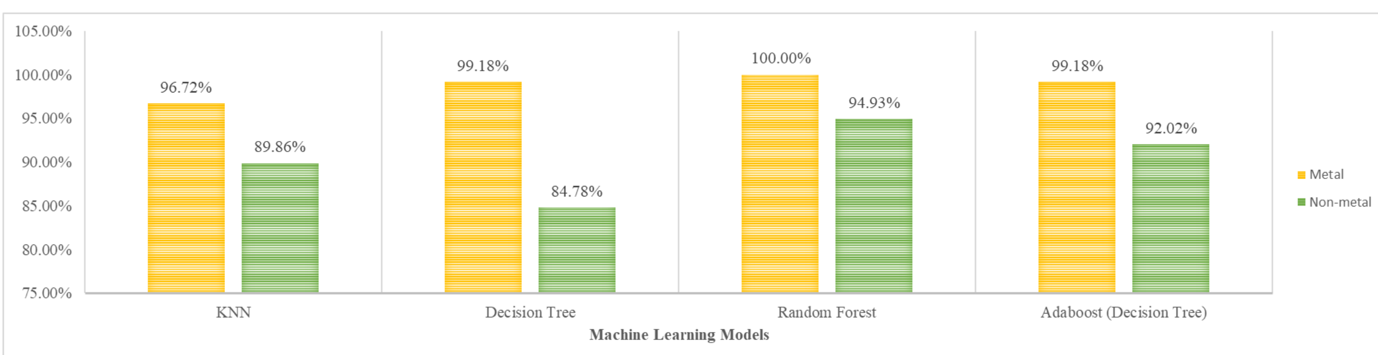

Figure 12. Comparison of accuracy of machine learning models

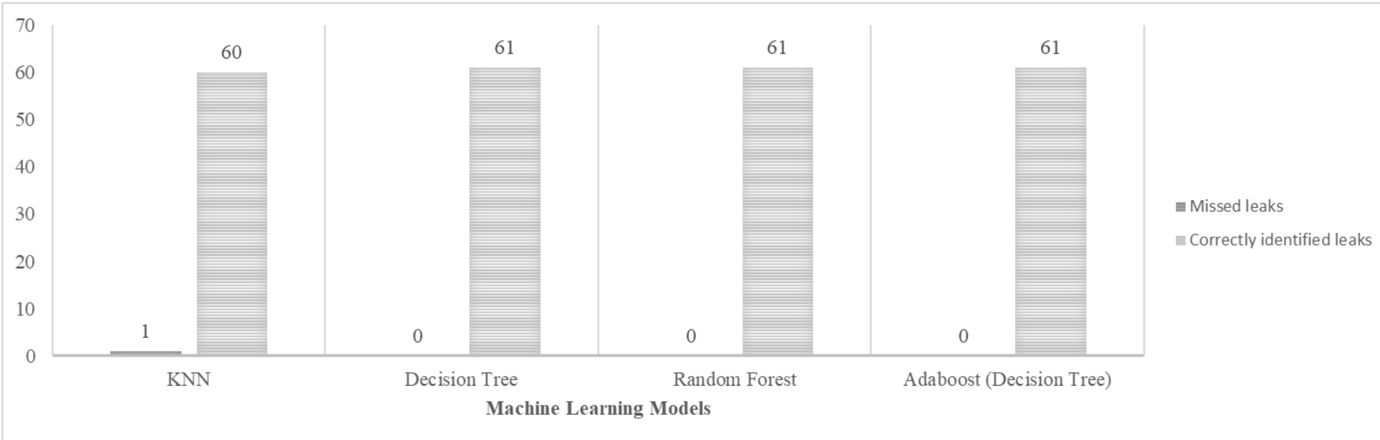

Figure 13(a). Number of leaks missed by metal-based machine learning models

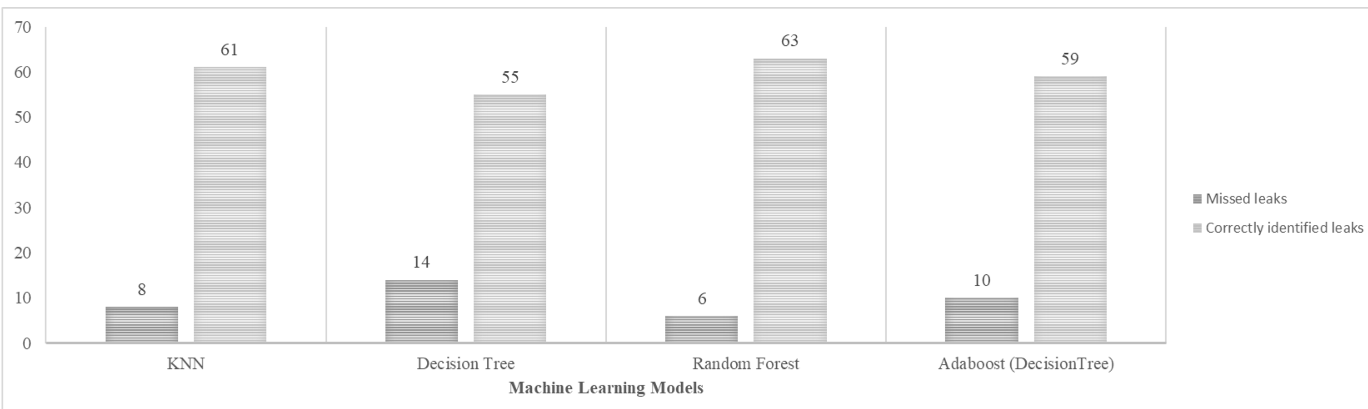

Figure 13(b). Number of leaks missed by non-metal-based machine learning models

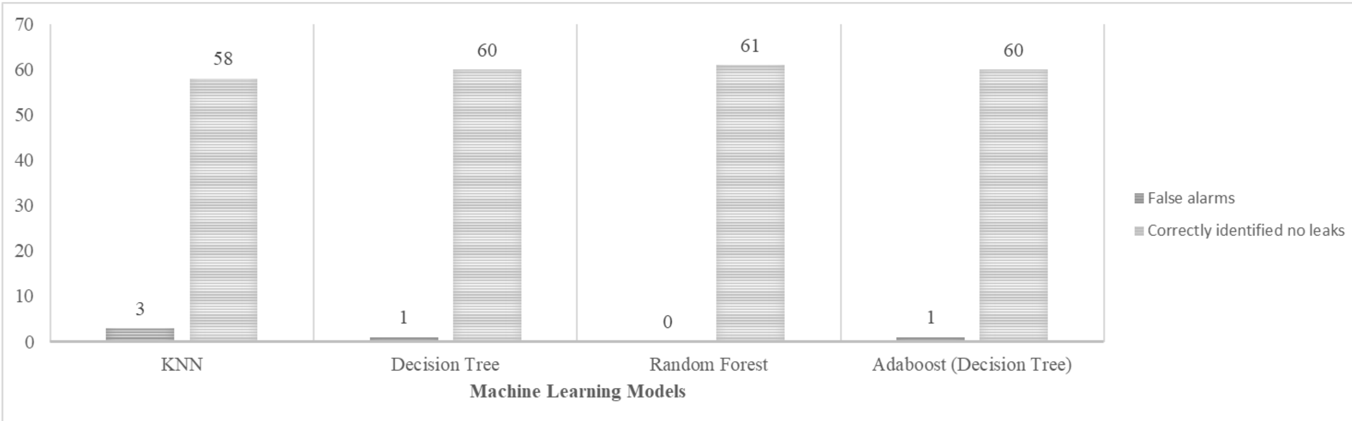

Figure 14(a). False alarms metal-based machine learning models

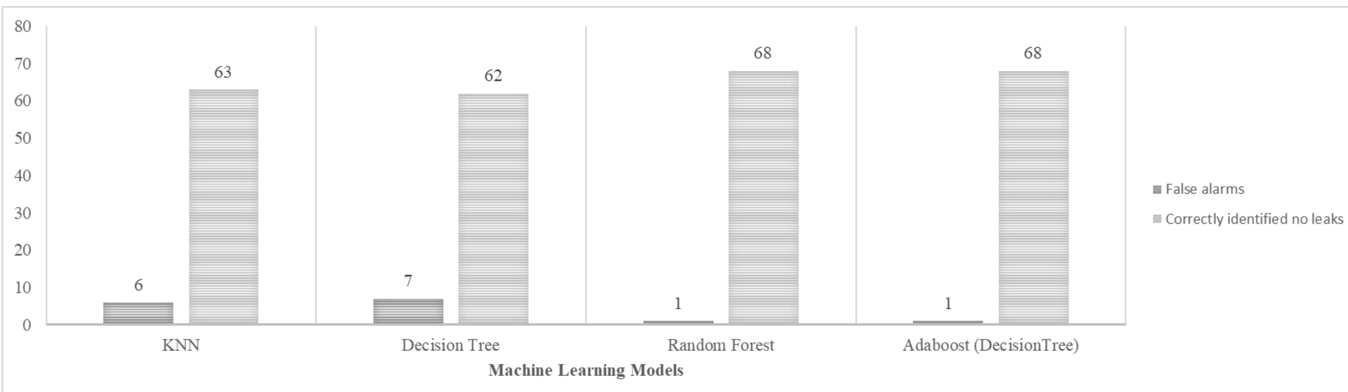

Figure 14(b). False alarms by non-metal-based machine learning models