

# Ensemble-based machine learning approach for improved leak detection in water mains

Thambirajah Ravichandran, Keyhan Gavahi,  
Kumaraswamy Ponnambalam, Valentin Burtea and S. Jamshid Mousavi

## ABSTRACT

This paper presents an acoustic leak detection system for distribution water mains using machine learning methods. The problem is formulated as a binary classifier to identify leak and no-leak cases using acoustic signals. A supervised learning methodology has been employed using several detection features extracted from acoustic signals, such as power spectral density and time-series data. The training and validation data sets have been collected over several months from multiple cities across North America. The proposed solution includes a multi-strategy ensemble learning (MEL) using a gradient boosting tree (GBT) classification model, which has performed better in maximizing detection rate and minimizing false positives as compared with other classification models such as KNN, ANN, and rule-based techniques. Further improvements have been achieved using a multitude of GBT classifiers combined in a parallel ensemble method called bagging algorithm. The proposed MEL approach demonstrates a significant improvement in performance, resulting in a reduction of false positives reports by an order of magnitude.

**Key words** | acoustic features, leak detection, machine learning, multi-strategy ensemble learning

**Thambirajah Ravichandran  
Kumaraswamy Ponnambalam**  
Department of Systems Design Engineering,  
University of Waterloo,  
Waterloo, Canada

**Keyhan Gavahi**  
Department of Civil, Construction and  
Environmental Engineering,  
University of Alabama,  
Tuscaloosa, AL, USA

**Valentin Burtea**  
Mueller Water Products – Echologics,  
Toronto, Ontario, Canada

**S. Jamshid Mousavi** (corresponding author)  
School of Civil and Environmental Engineering,  
Amirkabir University of Technology,  
Tehran, Iran  
and  
Department of Electrical and Computer  
Engineering,  
University of Waterloo,  
Waterloo, Canada  
E-mail: [jmosavi@aut.ac.ir](mailto:jmosavi@aut.ac.ir)

## HIGHLIGHTS

- State-of-the-art machine learning (ML) algorithms are used for solving the leak detection problem in water mains.
- A large number of acoustic signals and data are collected and used along with dimensionality reduction techniques as input features to ML algorithms.
- A novel multi-strategy ensemble-based algorithm is applied to improve further the performance of the investigated leak detection classification problem.

## INTRODUCTION

Safe drinking water is supplied to homes and businesses by water mains laid down decades ago that are nearing end of life. The break rate of water mains has increased by 27% from 11.0 to 14.0 breaks per 100 miles per year in the USA

and Canada during the 2012–2018 period (Folkman 2018). Even more concerning is that break rates of cast iron and asbestos cement pipe, which make up 41% of the installed water mains in these two countries, have increased by more than 40% over the same period. It is estimated that water utilities lose 10% of clean water due to leakage (Folkman 2018). Out of sight, water leaks may go undetected for a long time until they turn into catastrophic bursts and cost cities millions of dollars every year. The increase in operational costs to

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

doi: 10.2166/hydro.2021.093

supply water and the need to preserve an important natural resource are pushing water utilities to invest in new methods to detect, locate, and fix leaks efficiently (Yang *et al.* 2013). To improve operational efficiency utilities are now considering automatic remote monitoring systems to obtain real-time information for large-scale water distribution networks, which provide early detection and localization of leakage.

Several technologies have been developed during the past decade to address this issue. Some rely on hydraulic measurements such as minimum night flow (MNF) analysis and inverse transient analysis (ITA) (Liggett & Chen 1994), others use acoustic devices (Hamilton & Charalambous 2013), infrared thermography (Fahmy & Moselhi 2010), tracer gas (Hamilton & Charalambous 2013), ground-penetrating radar (Eyuboglu *et al.* 2003), leak detecting robots, and satellite imagery.

MNF analysis in district metered areas (DMAs) is the traditional approach to leak detection (AL-Washali *et al.* 2018). This method uses the input–output flow balance during night time to estimate water loss and the variance of pressure and flow readings as a measure of water consumption. While it can detect flow rates as small as 5 gal/min (Hamilton 2014), it is often difficult to separate real losses (leaks) from apparent losses caused by meter inaccuracy. Another limitation comes from the inability to locate the leaks. For these reasons, utilities apply DMA analysis as a survey for water loss, while they employ other technologies to locate and repair water mains defects. This process involves periodic inspections and costly manual work.

ITA was proposed by Liggett & Chen (1994) for the task of leak detection and localization. The method relies on minimizing the error between pressure and flow measurements and their corresponding signals obtained through numerical simulations. The parameters of a given network are adjusted until the best match is achieved. The difference between the reference model and the resulting configuration identifies possible defects in the pipe network. The numerical simulations require significant computational resources and the results depend on the accurate knowledge of the water network. This requirement is a significant obstacle in applying this method on a large scale, because detailed and reliable information about buried infrastructure is often difficult to obtain. In short, ITA provides reliable results in a known network, but the process is computationally intensive and difficult to scale.

Another popular method to leak detection is the tracer gas method. Pressurized non-toxic and insoluble gasses are infused into the water mains. If a pipe is leaking, the gas will escape and it could be detected from above ground using specific detectors, identifying the location of the leak (Hamilton & Charalambous 2013). This technique requires a human presence in the field and the process is laborious.

Among the aforementioned methods, acoustic monitoring devices are the most applied approach due to their accuracy, low cost, and easy to set up capabilities (El-Zahab & Zayed 2019). The technology relies on detecting acoustic signals generated by the release of pressurized fluids. Acoustic waves propagate over long distances in water mains and could be detected using vibration sensors, such as piezo-electric accelerometers, or hydrophones. The typical sensor spacing ranges from 100 m to 1 km, depending on the specific attenuation of acoustic waves for the given pipe specifications (Gao *et al.* 2005; Perrier *et al.* 2019). Consequently, the sensors could be installed on accessible appurtenances, such as existing hydrants or valves, providing a non-invasive and non-disruptive alternative to detect and locate leaks in buried water mains. With this arrangement, acoustic leak detection systems can detect leaks of 5 gal/min or more, assuming a minimum pressure of 30 psi in the pipelines. It is not uncommon that leaks as small 1 gal/min could be discovered if they occur in the immediate proximity of the sensors.

Many researchers have used acoustic signals as part of their leak detection systems (Fuchs & Riehle 1991; Hunaidi & Chu 1999; Khulief *et al.* 2012; Mostafapour & Davoudi 2013). The methods range from utilizing signal processing techniques to application of machine learning (ML) algorithms. Gao *et al.* (2017) focused on plastic pipeline systems and applied several signal processing techniques to improve the shape of the cross-correlation function for leak detection using acoustic signals. Wu *et al.* (2018) developed a novel denoising method by using acoustic signals sensed by a dynamic pressure transducer (DPT), improving the robustness and the stability of the detection system. Ting *et al.* (2019) used dual-tree complex wavelet transform (DTCWT) to reduce the noise in the acoustic signals and improve the localization error. In this study, the acoustic signals will be processed using ML, which is capable of recognizing the inherent patterns from input–output data relationships rather than analytical analysis of true signals.

As one of the effective means for tackling hydrosystems management problems, ML techniques have recently received notable attention (Maier & Dandy 2000; Han *et al.* 2007; Jung *et al.* 2010; Ebtehaj *et al.* 2017; Gavahi *et al.* 2018, 2019; Tyralis *et al.* 2019; Alipour *et al.* 2020; Jafarzadegan *et al.* 2020). This is due to their capability of mapping complex input/output relations that are often hard to fathom or inconceivable by humans (Solomatine & Ostfeld 2008). Especially in pipe leak detection, ML techniques can utilize network hydraulic and water-quality parameters for real-time pipelines monitoring (Cantos *et al.* 2020). To provide some examples, artificial neural networks (ANNs) have been extensively used for identifying failure events or predicting failure patterns in pipe systems (Pudar & Liggett 1992; Mounce *et al.* 2002, 2003, 2010; Mounce & Machell 2006; Romano *et al.* 2012, 2014). Mounce & Machell (2006) trained two types of static and time delay ANN using data recorded in a real water distribution system. The result showed that the networks were able to learn the leaks' patterns and the time delay ANN outperformed the simple multilayer perceptron network. The support vector machine (SVM) method was used by Mashford *et al.* (2012) to interpret batches of data collected from pressure sensors and flow measuring devices to be able to predict location and size of the leaks in a water pipe network. Mamo *et al.* (2014) developed a decision-making framework based on multiclass SVM to capture the presence of leaks in various sizes and locations. They showed that their framework is successful in identifying pipelines that required urgent action and can provide an early leak detection and monitoring model to the system operators and water utility companies. Romano *et al.* (2014) utilized a combination of artificial intelligence (AI) and statistical analysis techniques and proposed an automated near-real-time methodology to detect events such as pipe bursts and abnormal pressure variations at the district metered area level. The techniques included wavelets transform for denoising flow signals, ANN for short-term forecasting of flow values, and Bayesian inference systems for representing the probability of detected events and alarm raising sensitivity. Snider & McBean (2018) investigated a gradient boosting ML technique to improve the accuracy of time to failure prediction for individual pipe segments and demonstrated its better performance compared

with ANN and Random Forest methods. They also studied the benefits of using hydraulic and spatial predictor variables (average pressure and pipe break density variables) to further increase the accuracy of time to failure prediction. Cody *et al.* (2018) used singular spectrum analysis (SSA) to extract leak components from noisy measurements and showed that SSA is efficiently capable of extracting hydro-acoustic signals. They combined the SSA decomposition of leak-free data with an ensemble one-class SVM and demonstrated the effectiveness of SSA in combination with an ML technique for leak detection in water pipeline systems. As an emerging type on neural networks, convolutional neural networks (CNNs) are recently being used due to their ability of automatic feature extraction (Shen 2018). Chuang *et al.* (2019) and Zhou *et al.* (2019) utilized CNN for pipe leak detection using the images from the conversions of the reconstructed signals. Their result showed that CNNs are capable of leak detection and their corresponding location. These research studies show the growing interest in using ML techniques for leak/burst detection of pipelines at the district metered area.

The focus of this work is to contribute to the development of an effective water pipe leak detection system using acoustic techniques and various ML methods. In particular, we propose a multi-strategy ensemble learning (MEL) approach (Webb & Zheng 2004) for the leak detection problem in water mains. The leak detection is formulated as a binary classification problem to identify leak and no-leak cases using acoustic signals. Detection features used in this binary classifier are extracted from power spectral density and time-series reports collected over 3 months from several locations in the USA and Canada. Initially, three binary classifiers have been considered, namely K-nearest neighbors (KNN), ANN, and gradient boosting tree (GBT). Since the best performance has been obtained using GBT, several GBT classifiers have been combined then using a parallel ensemble method called bagging algorithm (BA). The performance of the proposed MEL algorithm for leak detection is evaluated in terms of sensitivity, specificity, and accuracy.

---

## ACOUSTIC WATER PIPE LEAK DETECTION SYSTEM

The industrial partner of this research is Echologics, a division of Mueller Water Products, a world leader in water

technology. Echologics provides leak detection services using acoustic technology for water utilities in North America, Europe, and Asia. Echologics uses proprietary wireless sensors (*EchoShore<sup>®</sup>-DX*) that are located inside fire hydrants (see Figure 1) and are listening for suspicious noise originating from water mains.

Leakage events generate acoustic waves with specific characteristics that propagate through the water mains and eventually reach the acoustic sensors.

Figure 2 shows an example of *EchoShore<sup>®</sup>-DX* deployment; the devices colored in red are excited by a nearby leak, the devices colored in yellow detect elevated acoustic

power but the source is not a leak, while the devices marked green are not detecting any change in acoustic power.

When one sensor detects specific vibration patterns it alerts a central hub, which then uses data from neighboring sensors to classify the event and locate its source. This paper discusses supervised learning algorithms to solve this specific classification problem. Detection features are extracted from acoustic reports collected from multiple nodes over time. For adequate performance, the training data set must exceed 10,000 acoustic samples collected from at least 100 distinct locations with different hydraulic and acoustic characteristics to ensure sufficient diversity in the data set.

Each node collects acoustic samples each day using the following process:

- Acoustic signals are recorded during night time to minimize spurious ambient noise.
- The sensors record data using a statistical sampling approach. Over 2 h, the sensors collect a total of 3 min of acoustic signals sampled at different moments in time, spread uniformly across the recording interval. This approach prevents collecting all data during a loud external event such as rain, wind, or traffic.
- The sound is recorded with a sampling rate of 4 kHz.
- Power spectral density and time series of sound loudness are computed.
- Relevant detection features are extracted.

The detection features could be segmented into three distinct classes, including (1) features obtained from the variation of sound loudness in time, (2) features obtained from the power spectral density of each sample, and (3) features obtained from the change in acoustic power.

The variation of loudness in time characterizes the ambient noise recorded at a specific location. The sound power is computed over each frame of 0.25 s resulting in a time series of 720 data points (see Figure 3) for each sample. A similar time series is produced using the instantaneous peak value for each frame. Statistical features are extracted from these two time series to characterize the acoustic context during a recording. The average power and the maximum instantaneous value for each frame  $k$

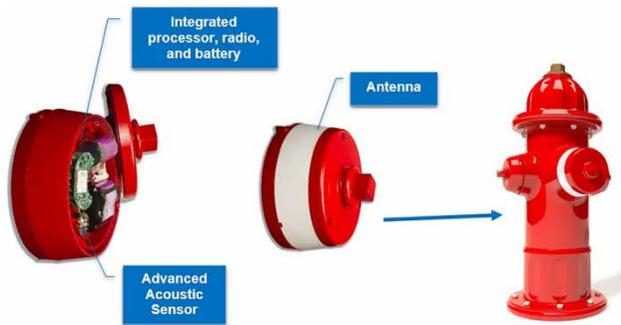


Figure 1 | Acoustic leak detection system (*EchoShore<sup>®</sup>-DX*).

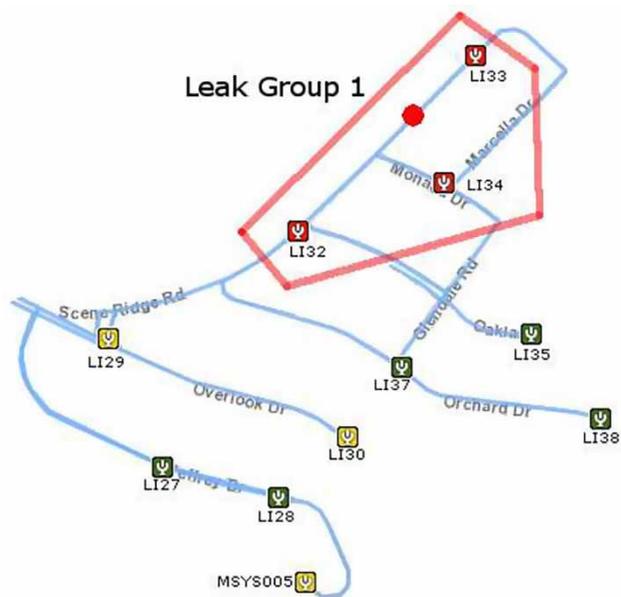


Figure 2 | Example of leak detection sensors deployed on a water pipe network. Please refer to the online version of this paper to see this figure in color: <https://doi.org/10.2166/hydro.2021.093>.

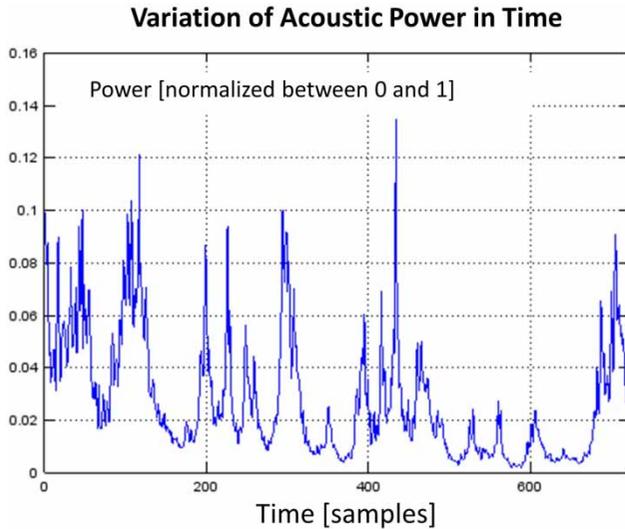


Figure 3 | Variation of acoustic power in time.

of length  $N$  are, respectively, calculated as follows:

$$P_k = \frac{1}{N} \sum_{i=(k-1)N}^{kN-1} x_i^2 \quad (1)$$

and

$$M_k = \max\{|x_i| \mid i = (k-1)N \dots kN-1\} \quad (2)$$

where  $x_i$  are the acoustic samples.

The power spectral density (spectrum) is estimated using the Bartlett's method of averaging periodograms. The periodogram is computed for each frame  $k$  of length  $N$  as the absolute value of the Discrete Fourier Transform. The power spectral density  $S(f)$  for the frequency component  $f$  is expressed as follows:

$$S(f) = \frac{1}{NF} \sum_{k=1}^{NF} \left| \sum_{n=0}^{N-1} x_n \cdot e^{-j2\pi f \left(\frac{n}{N}\right)} \right| \quad (3)$$

where 'NF' is the total number of frames. A Fast Fourier Transform algorithm is used to accelerate the computations. The reported spectrum is organized into 16 frequency bins harmonically ranged between 0 and 2000 Hz. Figure 4 shows an example of power spectral density (spectrum)

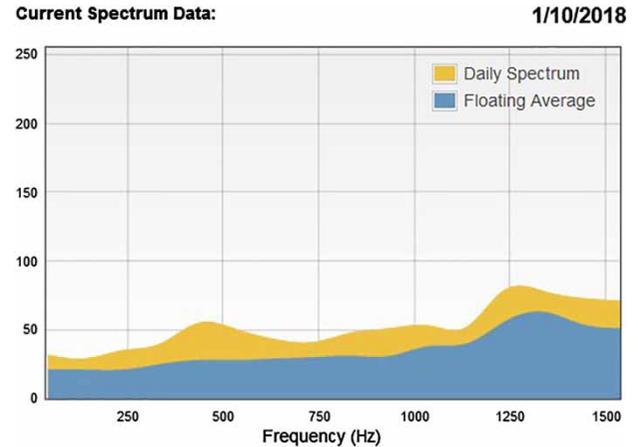


Figure 4 | Power spectral density and long-term average spectrum.

data and its long-term average of the spectrum. The long-term average spectrum is computed using a recursive exponential filter with a time constant  $T_c$ .

$$S_{AVG}(f, t) = S_{AVG}(f, t-1) + \frac{1}{T_c} (S(f, t) - S_{AVG}(f, t-1)) \quad (4)$$

The third class of features is extracted from the change between the recent spectrum and a long-term average. The change features capture several aspects of change, such as the maximum and average change in power spectral density, the variance of spectral change, the histogram of change across frequency bins, and the frequency range with the dominant change in power spectral density.

In total, 48 supervised features have been created from the three classes of available information.

Leakage events have been automatically located by the DX system using cross-correlation between acoustic signals recorded by adjacent nodes.

Figure 5 shows an example of cross-correlation signal used for leak location. The acoustic waves generated by the leak propagate through the pipe system and eventually reach the monitoring sensors. Depending on the distance between the source and each sensor, the acoustic waves will arrive at different times. Each sensor receives a similar signal component originating from the same source. This similarity produces a peak in the cross-correlation function, which indicates the time delay  $\Delta t$  between the received



system is given in Figure 6. The main components are the acoustic signal data collection, the feature extraction algorithm, and the binary classifier. Having explained in detail in the ‘Acoustic water pipe leak detection system’ section, the acoustic data collection from water mains and the feature extraction procedure using signal processing algorithms, this section provides details on the binary classifier design for detecting leaks in water mains.

It is well known that the classification problem is one of the major tasks very effectively solved using ML techniques. Typically, a classifier or classification model for solving binary or multiclass classification problems is designed and developed using the general framework, as shown in Figure 6. In this framework, the first step is to gather the larger and more diverse data sets that will be divided into training and testing data sets and used later in the pipeline by the Train Model and Test Model modules, respectively. Next, the Design Features module is responsible for extracting quality features from the raw data which can be a very application-dependent and challenging task as described in the previous section, where good features are extracted from raw acoustic data. It is also well understood that in order to develop a classifier with high performance, it is crucial that we need to provide it well-designed features.

Having extracted good features, the next step involves training a classification model by selecting an appropriate learning algorithm for the problem at hand. ML algorithms are broadly categorized into supervised, unsupervised, semi-supervised, and reinforcement learning types. In our specific case, the proposed learning model is supervised. The input and the output of the model are known. The output is the expected outcome of the model as defined by

a labeled data set (leak or not leak). The input consists of a set of features derived from physical principles. Popular ML algorithms may include KNN, ANNs, decision trees, support vector machines, etc. By training multiple classification models, it is possible to develop an ensemble of diverse classifiers to boost performance. Random forests and gradient boosting trees are prominent examples of this type of ensemble learning algorithms. Finally, the Test Model module is responsible for evaluating the performance of the trained classification model on the testing data set using a procedure known as cross-validation. For general classification and regression problems, commonly used cross-validation procedures hold out cross-validation, K-fold cross-validation, and bootstrapping-based cross-validation.

In addition to overall classification accuracy, the following performance measures are considered for effectively evaluating classification models: confusion matrix, precision, recall or sensitivity, specificity, and the receiver operating characteristic (ROC) curve and its area under the curve (AUC) measure. The confusion matrix provides a more effective means to evaluate the performance of a classifier. In a confusion matrix, a target class is represented by a row, while each column represents a predicted class. For a binary classifier, the four entries of this confusion matrix are true positive (TP: positive class correctly classified) and false positive (FP: negative class wrongly classified) given in the first row followed by false negative (FN: positive class wrongly classified) and true negative (TN: negative class correctly classified) given in the second row. Using these four entries of the confusion matrix, the other performance measures can be defined as follows. The overall classification accuracy is given by the

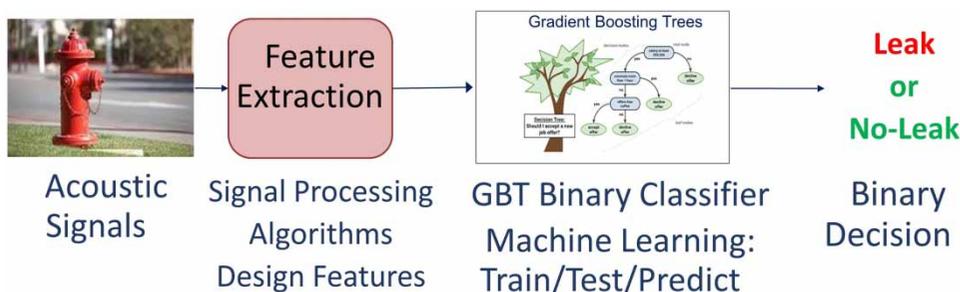


Figure 6 | Leak detection system components.

following ratio:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \quad (7)$$

The accuracy of the positive predictions is given by the performance called the precision of the classifier and is given as follows:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

The ratio of positive instances correctly identified by the classifier is known as the recall or sensitivity or true positive rate (TPR) of the classifier and is defined as follows:

$$\text{recall or sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

The ratio of negative instances correctly identified by the classifier is known as the specificity or true negative rate (TNR) of the classifier and is given as follows:

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (10)$$

The ROC curve is a plot of TPR (or sensitivity or recall) versus 1 – specificity or false positive rate (FPR). The area under the ROC curve is commonly used to compare classifiers. This definition implies that better performing classifiers have bigger areas under the ROC curve, and the classifiers having their ROC curves closer to the upper left corner (corresponding to a TPR of one and an FPR of zero) in the grid are preferred.

### GBT classifier

GBT classifier is a powerful ML algorithm that employs a sequential ensemble learning approach to convert multiple base classifiers also known as weak learners (typically decision trees are used as base classifiers) into one strong learner with improved classification performance. In general, the GBT algorithm can be applied to perform classification, regression, and feature ranking tasks (Hastie *et al.* 2009).

The GBT algorithm consists of three major components, namely a set of weak learners, a loss function, and an additive model which combines many weak learners into one strong learner to provide the desired GBT classifier. Decision trees are usually selected as base learners for developing the GBT classifier. Decision trees are generated in greedy manner, choosing the best split points to minimize the specified loss function. The base learners can be chosen as short decision trees having a single split or they can be large trees with maximum of up to 8 levels.

For classification, the GBT algorithm typically uses logarithmic loss function, but many other standard loss functions which are differentiable can be used as well. This flexibility of allowing any differentiable loss function to be used within the gradient boosting framework avoids the need for deriving new boosting algorithm for each loss function the user may want to specify, and this is a major benefit of the GBT algorithm.

In gradient boosting, weak learners are added in sequence one at a time using an additive model in which a gradient descent strategy is employed to minimize the loss function. In fact, gradient boosting algorithm casts the problem of combining weak learners into a strong learner (ensemble model) as a sequential gradient descent optimization problem: at each iteration, after calculating the loss, to perform the gradient descent procedure, a weak learner is added to the model that reduces the loss by parameterizing the decision tree and modifying its parameters to move in the gradient decent direction. Generally, this approach is called functional gradient descent or gradient descent with functions. The output for the new tree is then added to the output of the existing sequence of trees in an effort to correct or improve the final output of the model. A fixed number of trees are added, or training stops once loss reaches an acceptable level or no longer improves on an external validation data set.

### MEL APPROACH

Ensemble learning is an aggregation of multiple learners (classification or regression models) using some combination methods to form a strong learner (ensemble model). Unlike ordinary learning approaches, which try to

construct a single model from training data, ensemble learning methods try to construct multiple models to solve the same problem. Ensemble learning generally provides solutions with improved accuracy and/or robustness in most applications due to the availability of accurate and diverse multiple models for combining them into a single solution. Well-known ensemble learning algorithms include stacking (Wolpert 1992; Breiman 1996a), bagging (Breiman 1996b), and boosting (Freund & Schapire 1996) algorithms.

Generally, ensemble learning is implemented in three phases as shown in Figure 7, including (1) generation of base models, (2) selection of base models, and (3) aggregation of the selected base models using some combination methods. In the first phase, a pool of base models is generated, and the pool may consist of homogeneous base models (same model types) or heterogeneous base models (mixture of different model types). Base learners are usually generated from training data by a base learning algorithm such as decision trees, neural networks, or other methods. In the second phase, a subset of base models is selected. Finally, a model is formed by aggregating the selected models using a combination method. The generalization ability of an ensemble is often much stronger than that of base learners. To get a final model with improved generalization, it is essential that the base

models should be as accurate as possible, and as diverse as possible.

Depending on how the base learners are generated, ensemble methods are divided into two types: sequential ensemble methods and parallel ensemble methods. In sequential ensemble methods, base learners are generated sequentially and the above-described GBT algorithm is a good example of this type of sequential ensemble methods. In parallel ensemble methods, the base learners are generated in parallel and some details on this type of ensemble methods are given below.

### Parallel ensemble method

The main motivation of parallel ensemble methods is to exploit the independence between the base learners to improve the generalization performance of the ensemble model generated by combining the independent base learners. Well-known examples of this parallel ensemble paradigm include bagging, pasting, and random forest methods. Popular mechanisms to generate diverse and independent homogeneous base learners include data sample manipulation and input feature manipulation methods. Bagging and pasting methods adopt the data sample manipulation approach for generating diverse and independent base learners, and some more details on these two parallel ensemble methods are given below.

For data sample manipulation, bagging employs the bootstrap sampling mechanism to generate different random data subsets for training the base learners. In bootstrap sampling, training data instances are generated by sampling with replacement, which means some original data instances appear more than once while some other original data instances are not included at all in the chosen training data set. When this sampling is performed without replacement, the resulting ensemble method is called pasting. Once all the base learners are trained, for classification task, bagging and pasting methods typically adopt a voting strategy (also known as majority voting) for combining the outputs of the base learners. That is when provided with a new input data instance, bagging/pasting collects the output labels of base classifiers and then performs voting to decide the winner label as the final prediction. The pseudo-code of the bagging/pasting algorithm is given in Figure 8.

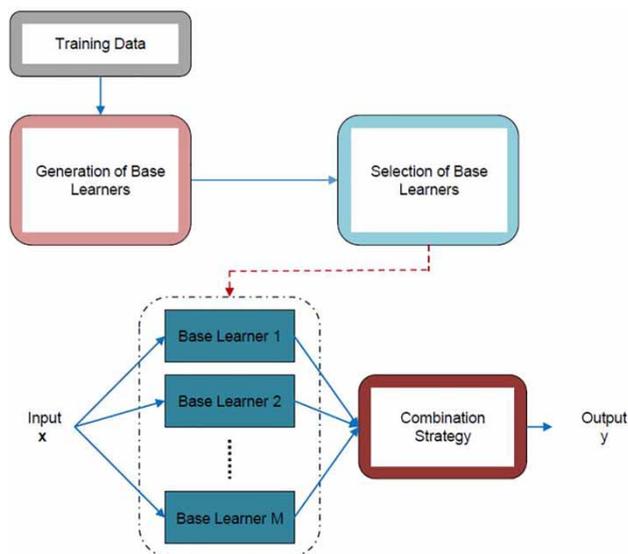


Figure 7 | Ensemble learning process.

---

### Bagging/Pasting Algorithm

---

**Inputs:** Data set  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ; Base Learning Algorithm  $L$ ; Number of Base Learners  $M$ .

**Outputs:** Ensemble classifier  $H(x)$

---

1. Train  $M$  base learners using the given base learning algorithm with the training data  $D_{\text{train}}$  generated using the bootstrap distribution (bagging) or randomly selected major portion of the data set  $D$ :

```

for  $m = 1, \dots, M$ 
   $h_m = L(D_{\text{train}})$ 
end

```

2. Form the ensemble classifier by combining the base learner using *majority voting* procedure:

$$H(x) = \underset{y \in Y}{\text{arg max}} \sum_{m=1}^M I(h_m(x) = y)$$


---

**Figure 8** | Pseudo-code of the bagging/pasting algorithm.

Typically, the computational cost of constructing an ensemble of base learners is not much larger than creating a single base learner. This is because typically we need to generate multiple versions of the model for model selection when we want to construct a single model, and this is comparable to generating base learners in ensemble learning, while the computational cost for combining base learners is often small.

As noted above, parallel and sequential ensemble methods are implemented by different mechanisms depending on how the base learners are generated. Therefore, it might be possible to enhance performance by combining these two ensemble-based methods. For example, bagging and AdaBoost algorithms as representatives of parallel and sequential methods exhibit the following different performances, respectively; AdaBoost is good at reducing both bias and variance, and bagging performs better in reducing variance more effectively than AdaBoost (Bauer & Kohavi 1999). Therefore, a model combining these methods may retain advantages of both (Webb & Zheng 2004), motivating researchers to investigate MEL approaches by combining simple ensemble learning methods. We adopt in this study a MEL approach by combining multiple and diverse GBT classifiers (sequential ensembles) trained as base learners using the bagging/pasting algorithm (parallel ensemble) as shown in Figure 7.

### CASE STUDY

The data set used in this study includes over 14,000 acoustic samples collected from more than 300 locations spread across the USA and Canada. The acoustic sensors have been installed in distribution water networks on a variety of metallic pipes. A combination of cast iron and ductile iron pipes was selected for this study. Different pipe diameters, ranging from 6 to 12 in. have been included. The average static pressure varies among locations from 50 to 120 psi. The operational data set includes 13,861 negative samples and 54 leak cases observed over 3 months. The observed flow rates range from 5 gpm to 50 gpm. Some leaks occurred directly on water mains, while others developed on service lines.

One difficulty is related to the asymmetrical composition of the data set as it includes a relatively low sample count of positive events. Indeed, there are very few leaks occurring naturally in urban networks, so most of the samples collected were non-leaks. To increase the number of leakage events, we resorted to simulations by creating artificial flows on the network. With these controlled experiments, we created a simulated data set of 240 samples, consisting of 194 non-leaks and 46 simulated leaks.

These artificial events have been created by flowing water, generally from hydrants. The noise generated by the

release of fluid under pressure is expected to have similar characteristics to a leak as the physical phenomena are similar. While similar there are some notable differences: when flowing a hydrant, the fluid is released in the atmosphere, while in the case of a leak, the fluid is released in the surrounding soil. The back pressure of the soil and the water saturation of the surrounding medium may affect the sound generated by the leak.

On the other hand, there are several similarities. The frequency range is similar for both real and simulated events. The lower frequency is dictated mainly by pipe specification and sensor technology. Accelerometers act as high-pass filters (Gao et al. 2005) for the low-frequency range. The wavelength of the acoustic waves is related to the pipe diameter and material. Larger pipes propagate longer wavelengths, thus lower frequencies. The acoustic waves in plastic pipes have lower frequencies compared with metallic pipes (Muggleton et al. 2003). In general, the lower bound of the frequency range is below 200 Hz (Gao et al. 2005). High frequencies attenuate faster when the acoustic waves are propagating through the pipe network. It means that there is an upper bound of the frequency range given the pipe specifications and the distance from the source (Almeida et al. 2014; Perrier et al. 2019). To ensure the diversity of the training set, the simulated events have been placed at various distances from the sensors. The distances from the leak to the sensors varies between 50 and 500 ft. The real leakage events occurred within 100–350 ft from adjacent sensors.

Another similarity is that both acoustic sources are broadband, exciting a wide range of frequencies.

It is also remarkable that the acoustic characteristics of leaks could vary significantly among cases including both real and simulated events. Difference in pressure, soil, pipe material, and pipe configuration can affect significantly the sound measured by the sensors. The spectrum of leakage events is not flat; the variation is caused by sound reflections from appurtenances and water network topology.

The simulated leaks preserve the characteristics listed above. For these reasons, this approach was considered suitable to increase the sample size for classifier training. At the same time, specific differences have been considered. The classifier was trained using a combination of simulated sources and real events. Another constraint was applied to the cost function of the classifier: the cost

of missing a leak is set much higher than creating a false detection. Because of this, the classifier was forced to detect all the leakage events while accepting a larger number of false positives. In fact, the performance of the classifiers is compared on how many false alerts are generated, while all the leakage events are classified correctly and no leak goes undetected.

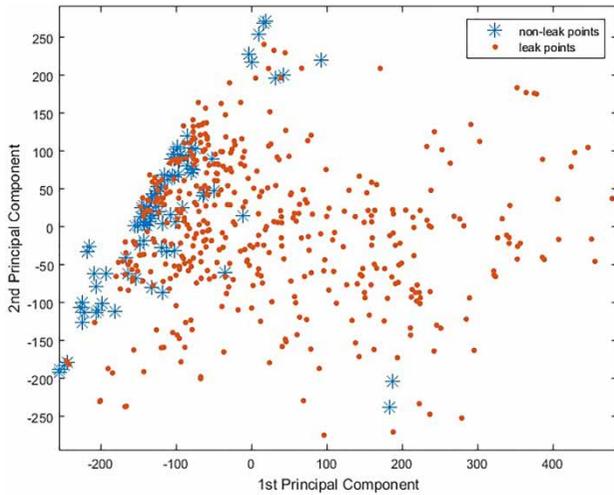
The two sets combined present 100 leaks and 14,055 negative events.

## RESULTS AND DISCUSSION

### General results

In order to demonstrate the performance of the proposed ML based approach for the acoustic water pipe leak detection system for discriminating between leak and no-leak cases effectively, in this subsection, classification results obtained using a simulated data set are presented. First, a simulated data set with 540 data samples (or instances) and 48 features as described in the ‘Acoustic water pipe leak detection system’ section is used for designing different classifiers, namely KNN, ANN, and GBT. In this data set, 446 data instances (83%) are in the Leak class and 94 data instances (17%) are in the No-Leak class.

Given this data set with 540 data samples, it is necessary to reduce the dimensionality of the input features at least in the case of ANN classifier design without overfitting. For this purpose, a well-known dimensionality reduction technique called principal component analysis (PCA) algorithm was employed. PCA algorithm projects the given data onto a lower-dimensional hyperplane while maintaining the variance within the data as much as possible. Using the PCA algorithm, the first eleven principal components are found to be explaining 90.21% of the variance within the above data set with 540 data samples. Therefore, the first 11 principal components are selected as input features that can be used to design KNN and ANN classifiers for detecting water pipe leaks. To illustrate the discriminative power of the reduced number of features obtained using the PCA algorithm, the scattered plot of leak and no-leak cases from the above data set is shown in Figure 9 using the first two principal components.



**Figure 9** | Scattered plot of leak and no-leak cases from the simulated data using the first two principal components.

As baseline classifiers based on the simulated data set, KNN and feedforward ANN classifiers were designed using the first 11 principal components as input features. For the KNN classifier,  $K$  was chosen as 5, and a single hidden layer with 15 neurons was selected as the architecture for the ANN classifier.

A GBT classifier was also designed using all 48 features extracted as explained in the ‘Acoustic water pipe leak detection system’ section. For this GBT classifier, the maximum number of trees and the maximum tree depth were chosen as 300 and 3, respectively. A greater number of trees allows better learning capability for the GBT classifier; however, it can also lead to slowing down of the training process as well as resulting in an overfitting situation. The tree depth represents how deep each tree can be built, and the deeper the tree, more information about the data can be captured by the GBT classifier. Higher depth values will also lead to overfitting of the data by the classifier. Therefore, it is important to choose these hyper-parameters of the GBT classifier appropriately either by some trial and error experiments or using a parameter sweep procedure. For developing the above classifiers, the simulated data set were randomly divided into 75% for the training set and the remaining 25% for the testing set. After that, the non-leaks points (the minority class) in the training set were randomly over-sampled. Then, the classifiers were trained and validated by performing multiple runs, and during each run, the seed

number was changed in order to have different randomly selected training and test sets. Table 1 compares the classification performance values (averaged over multiple runs) on the test data set for the above-mentioned classifiers.

Table 1 presents the average performance results along with their spread values (given within brackets) obtained from 20 repeated runs of each model using a bootstrap strategy. The ANN classifier shows an improvement of 5.1, 5.3 and 2.6% over the KNN classifier in terms of average classification accuracies for leak, non-leak, and the overall cases, respectively. However, considering the spread values given in Table 1, the difference in performance values between KNN and ANN classifiers is not very significant. The GBT classifier illustrates a further improvement of 7.8, 21, and 12.2% over the ANN classifier in terms of average classification accuracies for leak, non-leak, and overall cases, respectively. The GBT classifier shows a reduction of 45.2 and 40% in average False Positive Rate compared with the KNN and ANN classifiers, respectively. The comparison results shown in Table 1 clearly demonstrate the superiority of the GBT classifier in detecting leak and no-leak cases accurately.

A particularity of the leak detection classification problem is that high sensitivity is easier to achieve than high specificity and precision. The leak detection network is designed in such a way that the sound generated by potential leaks has a high probability of exciting the sensors. The sensor placement takes into account the statistics of acoustic power generated by leakage events and the attenuation of acoustic waves in water pipes. Thus, sensitivity depends greatly on the quality of the hardware and system configuration. A system able to detect changes in sound power originated from leaks for the majority of cases has a relatively high sensitivity (Equation (8)). The no-leak cases are more difficult to classify as the

**Table 1** | Classification performance (averaged over multiple runs) on test data for KNN, ANN, and GBT classifiers

Accuracy	KNN-PCA	ANN-PCA	GBT using all features
Overall	81.2% (2.35%)	83.3% (1.08%)	93.5% (0.97%)
Leaks	85.2% (2.69%)	89.5% (1.24%)	96.5% (0.98%)
No-Leaks	62.4% (3.36%)	65.7% (2.76%)	79.4% (1.16%)
False Positive Rate	37.6% (2.97%)	34.3% (2.54%)	20.6% (1.68%)

sensors could be excited by other noise sources than leaks, such as traffic, wind, rain, or other ambient noise sources. This is reflected in Table 1 by the higher classification rate for leaks (over 85%) and lower figures for no-leaks.

The proposed MEL approach is aimed at improving the no-leak classification accuracies further. As discussed before, the GBT classifier is well known for its capability to perform ranking of inputs or features based on their relative importance. Typically, in ML applications, rarely all the input or feature variables play equal to influence on the output response. It is often useful to know the relative importance of each feature variable in predicting the output response, and this knowledge is provided by the GBT classifier. In order to exploit this capability for feature ranking, the relative importance of the 48 acoustic features (as extracted in the ‘Acoustic water pipe leak detection system’ section) of the simulated data as shown in Figure 10 is utilized to select most relevant features for designing the GBT classifier with better generalization performance. Table 2 shows the performance comparison in terms of classification accuracy (averaged over multiple runs) between the GBT classifiers designed with all 48 features and the most relevant 30 features. For both of these GBT classifiers, the maximum number of trees and the maximum tree depth were chosen as 300 and 3, respectively.

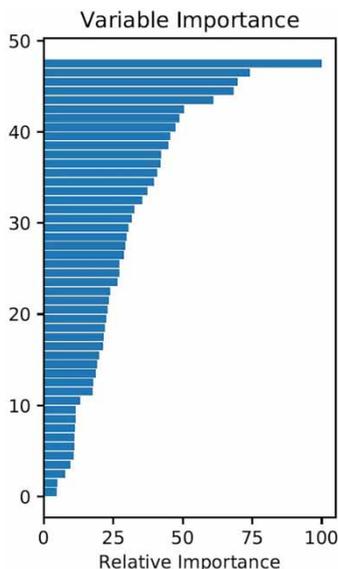


Figure 10 | Relative importance of the acoustic features for the simulated data.

Table 2 | Classification performance (averaged over multiple runs) on test data for GBT classifier

Accuracy	GBT using all features	GBT using top 30 features	Bagging on GBT using top 30 features
Overall	93.5%	93.8%	95.6%
Leaks	96.5%	96.5%	97.5%
No-Leaks	79.4%	81.0%	86.2%
False Positive Rate	21.6%	19.0%	13.8%

In order to demonstrate the performance improvement achievable with the MEL approach for discriminating between leak and no-leak cases effectively, a parallel ensemble method is employed to combine multiple GBT classifiers trained as base learners (using a sequential ensemble method) using the above-mentioned simulated data set. Here, the parallel ensemble method is implemented using the bagging/pasting scheme which employs an out-of-bag evaluation method along with a majority voting technique to implement the MEL approach. During the phase of generating base learners, the data set is randomly divided using a 75–25% split into a training set and a testing set which are also called out-of-bag instances (in the case of bootstrapping-based data split). It should be noted that these out-of-bag instances are not the same for all base learners. When multiple runs are used to generate base learners, where during each run a base learner is generated, a given data instance is not seen by on average about 37% of the base learners. Therefore, for each data instance, around 37% of the base learners can take part during the majority voting process. Using this scheme of out-bag-evaluation with majority voting, the proposed multi-strategy ensemble classifier was developed using the GBT classifiers with top 30 features as base learners. The resulting classification performance values are compared in Table 2 against those obtained with the base learners.

The comparison results shown in Table 2 clearly illustrate the advantage using the proposed MEL approach which employs the bagging method to combine multiple and diverse GBT classifiers to further improve the classification accuracy for no-leak cases. Reducing the number of features from 48 to 30 shows statistically similar performance, but a significant reduction in classifier complexity. Thus, the ensemble method used the same reduced feature

set. The ensemble classifier implemented using the bagging method shows an improvement of 8.6 and 6.4% in terms of no-leak classification accuracy over the GBT classifiers using 48 and 30 features, respectively. Compared with the base GBT classifier, the ensemble bagging method shows a reduction of 27.4% in False Positive Rate. These improvements were due to the unique aspect of the bagging method in utilizing the diversity of multiple GBT classifiers to improve the classification performance over individual base learners.

### Further results on the multi-strategy ensemble classifier

In order to further demonstrate the effectiveness of the MEL approach for the acoustic water pipe leak detection system for discriminating between leak and no-leak cases, classification results obtained using two different field data sets combined are presented. First, a simulated data set with 240 data instances (collected from 23 node segments) along with 48 features as described in the 'Acoustic water pipe leak detection system' section is considered for designing base learners (GBT classifiers). In this simulated data set, 46 data instances (from 10 node segments) are in the Leak class and 194 data instances (from 23 node segments) are in the No-Leak class. Second, a real-field operation data set with 13,915 data instances (collected from 302 node segments) along with 48 features is considered for designing base learners (GBT classifiers). In this field data set, 54 data instances (from 6 node segments) are in the Leak class and 13,861 data instances (from 302 node segments) are in the No-Leak class. These two data sets were combined to develop the proposed multi-strategy ensemble classifier using the following data split for training and testing sets.

For the Leak class from both data sets, data instances from 70% of the node segments were considered for the training data set, and the testing data instances were chosen from 30% of the node segments. For the No-Leak class from the simulated data set, the same 70–30% data split was considered. However, for the No-Leak class from the field data set, data instances from 10% of the node segments were considered for the training data set, and the

testing data instances were chosen from 90% of the node segments. During these training and testing data splits, all the data instances from any given node are kept in either training or testing data set but not in both. As before, the leak points (the minority class) in the training set were randomly oversampled.

Using the above-mentioned combined data set, multiple GBT classifiers were trained as base learners (based on a sequential ensemble method) using 48 features along with the maximum number of trees and the maximum tree depth, chosen as 300 and 3, respectively. Then, a parallel ensemble method was employed to combine these multiple base learners. As done before, the parallel ensemble method is implemented using the bagging/pasting scheme which employs an out-of-bag evaluation method along with a majority voting technique to implement the proposed MEL approach. Using this scheme of out-of-bag-evaluation with majority voting, the proposed multi-strategy ensemble classifier was developed using the GBT classifiers as base learners.

During the phase of generating base learners, multiple runs were performed by changing the seed number to have a different random data set during each run. Testing was performed using 1,000 runs with 30% of randomly selected data samples from each class. This design strategy ensures that the resulting classifier is general and does not over-fit to a particular data set.

The 'best' single GBT classifier is found to yield the following performance: Overall accuracy: 99.28%, Sensitivity: 100%, Specificity: 99.27%, and False Positive Rate of 0.73%.

To build the ensemble learner, we selected 25 'top' performing GBT classifiers. Diversity was also considered, by ensuring that the outputs of these classifiers were uncorrelated. Based on the combined all data instances, the ensemble classifier provides the following classification performance: Overall accuracy: 99.84%, Sensitivity: 100%, Specificity: 99.84% and False Positive Rate of 0.16%.

The overall accuracy is relatively high in all cases due to a large imbalance in the labeled data set: the vast majority of samples are no-leak cases (14,055) and only 100 samples are leaks. Most no-leak cases are quiet and do not generate a change in acoustic power; these no-leak cases can be easily classified with simple detection rules by monitoring changes in acoustic power. This classification was provided directly as the output of the current

EchoShore<sup>®</sup>-DX leak detection system. The performance of this ‘rule-based’ detection is presented as a baseline for comparison.

All classifiers have been tuned to achieve a Sensitivity of 100%, which ensures that all leak events are detected. Thus, the classifiers are compared in terms of Specificity and False Positive Rate.

The confusion matrices for the ‘rule-based’ classifier, the ‘best’ single GBT classifier, and the ensemble classifier are given below, and the classification performances for these three classifiers are compared in Table 3.

Confusion Matrix of “Rule-based” classifier:  $\begin{bmatrix} 13817 (TN) & 238 (FP) \\ 0 (FN) & 100 (TP) \end{bmatrix}$

Confusion Matrix of “Best” classifier:  $\begin{bmatrix} 13953 (TN) & 102 (FP) \\ 0 (FN) & 100 (TP) \end{bmatrix}$

Confusion Matrix of Ensemble classifier:  $\begin{bmatrix} 14032 (TN) & 23 (FP) \\ 0 (FN) & 100 (TP) \end{bmatrix}$

The comparison results shown in Table 3 are directly calculated from the above confusion matrices. Overall, the results demonstrate the advantage of the ensemble classifier over the single GBT classifier. Both ML classifiers perform better than the ‘rule-based’ solution. The ensemble classifier shows a reduction of 90.5 and 78.1% in False Positive Rate compared with the ‘rule-based’ and the ‘best’ single GBT classifiers, respectively. The above confusion matrices and the results in Table 3 show that the MEL method shows significant improvement by reducing the False Positives by an order of magnitude (from 238 to 23) while correctly identifying all leakage events (100% Sensitivity). These results also demonstrate that performance and robustness enhancements are possible for the design of a leak detection system for water mains by adopting this proposed MEL method.

**Table 3** | Classification performance on test data for multi-strategy ensemble classifier

Accuracy	Rule-based	‘Best’ classifier	Ensemble classifier
Overall	98.32%	99.28%	99.84%
Leaks	100%	100%	100%
No-Leaks	98.31%	99.27%	99.84%
False Positive Rate	1.69%	0.73%	0.16%

## CONCLUSIONS

In this paper, a MEL approach was presented as an effective solution for the development of an improved water pipe leak detection system using acoustic techniques. The leak detection was formulated as a binary classification problem to identify leak and no-leak cases using acoustic signals. Detection features used in this binary classifier were extracted from power spectral density and time-series reports collected over 3 months from multiple cities across North America. In total, 48 acoustic features were created from three classes of available information for the development of binary classifiers to detect leaks in water mains. To demonstrate the performance of the proposed ML based approach for discriminating between leak and no-leak cases effectively, a simulated data set with 540 data samples and 48 features were used for designing different classifiers such as KNN, ANN, and GBT. PCA was used for dimensionality reduction before training KNN and ANN classifiers. Results indicated GBTs performed better than KNN and ANN in maximizing detection and minimizing false positives.

To further demonstrate the effectiveness of the proposed MEL approach, classification results were obtained using two different field data sets combined. To implement the proposed MEL method, several GBT classifiers were combined using a parallel ensemble method also known as the BA. It was demonstrated that performance and robustness enhancements were possible by adopting the proposed MEL approach benefiting from the BA. The binary classifier designed using the MEL method showed significant improvement. It reduced the number of False Positives by an order of magnitude (from 238 to 23), and all leakage events were correctly identified. These performance improvements were due to the unique aspect of the MEL method to combine the individual strengths of sequential and parallel ensemble methods in achieving both the bias reduction and variance reduction simultaneously.

A reduction in false positive events has significant operational implications for water utilities by reducing the cost to investigate these events. For future work, it is possible to utilize a model-based leak detection and isolation system using hydraulic simulation models to generate

training data for building a more diverse ensemble of multiple classifiers and improve further the performance of the proposed approach.

## ACKNOWLEDGEMENTS

We thank Mueller Water Products – Echologics, Toronto, Ontario, Canada, and Alex Wong who was a co-applicant with K. Ponnambalam for the Fed Dev Grant funded with assistance from Southern Ontario Water Consortium.

## DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

## REFERENCES

- Alipour, A., Ahmadalipour, A., Abbaszadeh, P. & Moradkhani, H. 2020 Leveraging machine learning for predicting flash flood damages in the Southeast US. *Environmental Research Letters* **15** (2), 024011.
- Almeida, F., Brennan, M., Joseph, P., Whitfield, S., Dray, S. & Paschoalini, A. 2014 On the acoustic filtering of the pipe and sensor in a buried plastic water pipe and its effect on leak detection: an experimental investigation. *Sensors (Switzerland)* **14** (3), 5595–5610.
- AL-Washali, T., Sharma, S., AL-Nozaily, F., Haidera, M. & Kennedy, M. 2018 Modelling the leakage rate and reduction using minimum night flow analysis in an intermittent supply system. *Water (Switzerland)* **11** (1), 48.
- Bauer, E. & Kohavi, R. 1999 Empirical comparison of voting classification algorithms: bagging, boosting, and variants. *Machine Learning* **36** (1), 105–139.
- Breiman, L. 1996a Stacked regressions. *Machine Learning* **24** (1), 49–64.
- Breiman, L. 1996b Bagging predictors. *Machine Learning* **24** (2), 123–140.
- Cantos, W. P., Juran, I. & 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), 1–10.
- Chuang, W. Y., Tsai, Y. L. & Wang, L. H. 2019 Leak detection in water distribution pipes based on CNN with MEL frequency cepstral coefficients. In: *ACM International Conference Proceeding Series*, March 2019, Suzhou, China, pp. 83–86.
- Cody, R., Harmouche, J. & Narasimhan, S. 2018 Leak detection in water distribution pipes using singular spectrum analysis. *Urban Water Journal* **15** (7), 636–644.
- Ebtehaj, I., Sattar Ahmed, M. A., Bonakdari, H. & Zaji, A. H. 2017 Prediction of scour depth around bridge piers using self-adaptive extreme learning machine. *Journal of Hydroinformatics* **19** (2), 207–224.
- El-Zahab, S. & Zayed, T. 2019 Leak detection in water distribution networks: an introductory overview. *Smart Water* **4** (1), 1–23.
- Eyuboglu, S., Mahdi, H. & Al-shukri, H. 2003 Detection of water leaks using ground penetrating radar. In: *Proceedings of the Third International Conference on Applied Geophysics*, December 2003, Orlando, FL.
- Fahmy, M. & Moselhi, O. 2010 Automated detection and location of leaks in water mains using infrared photography. *Journal of Performance of Constructed Facilities* **24** (3), 242–248.
- Folkman, S. 2018 Water main break rates in the USA and Canada: a comprehensive study. In: *Mechanical and Aerospace Engineering Faculty Publications*, March 2018, Logan, UT, pp. 1–49.
- Freund, Y. & Schapire, R. E. 1996 Experiments with a new boosting algorithm. In: *Machine Learning: Proceedings of the Thirteenth International Conference*, January 1996, Murray Hill, NJ.
- Fuchs, H. V. & Riehle, R. 1991 Ten years of experience with leak detection by acoustic signal analysis. *Applied Acoustics* **33** (1), 1–19.
- Gao, Y., Brennan, M. J., Joseph, P. F., Muggleton, J. M. & 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.
- Gao, Y., Brennan, M. J., Liu, Y., Almeida, F. C. L. & Joseph, P. F. 2017 Improving the shape of the cross-correlation function for leak detection in a plastic water distribution pipe using acoustic signals. *Applied Acoustics* **127**, 24–33.
- Gavahi, K., Mousavi, S. J. & Ponnambalam, K. 2018 Comparison of two data-driven streamflow forecast approaches in an adaptive optimal reservoir operation model. *EPiC Series in Engineering* **3** (1), 755–763.
- Gavahi, K., Mousavi, S. J. & Ponnambalam, K. 2019 Adaptive forecast-based real-time optimal reservoir operations: application to Lake Urmia. *Journal of Hydroinformatics* **21** (5), 908–924.
- Hamilton, S. 2014 *Water Management and Water Loss*. Water Intelligence Online. IWA Publishing, London.
- Hamilton, S. & Charalambous, B. 2013 *Leak Detection: Technology and Implementation*. IWA Publishing, London.
- Han, D., Chan, L. & Zhu, N. 2007 Flood forecasting using support vector machines. *Journal of Hydroinformatics* **9** (4), 267–276.
- Hastie, T., Tibshirani, R. & Friedman, J. 2009 *The Elements of Statistical Learning*. Springer Series in Statistics.
- Hunaidi, O. & Chu, W. T. 1999 Acoustical characteristics of leak signals in plastic water distribution pipes. *Applied Acoustics* **58** (3), 235–254.
- Jafarzadegan, K., Merwade, V. & Moradkhani, H. 2020 Combining clustering and classification for the regionalization of

- environmental model parameters: application to floodplain mapping in data-scarce regions. *Environmental Modelling and Software* **125**, 104613.
- Jung, N. C., Popescu, I., Kelderman, P., Solomatine, D. P. & Price, R. K. 2010 Application of model trees and other machine learning techniques for algal growth prediction in Yongdam reservoir, Republic of Korea. *Journal of Hydroinformatics* **12** (3), 262–274.
- Khulief, Y. A., Khalifa, A., Mansour, R. B. & Habib, M. A. 2012 Acoustic detection of leaks in water pipelines using measurements inside pipe. *Journal of Pipeline Systems Engineering and Practice* **3** (2), 47–54.
- Liggett, J. A. & Chen, L. 1994 Inverse transient analysis in pipe networks. *Journal of Hydraulic Engineering, ASCE* **120** (8), 934–955.
- Maier, H. R. & Dandy, G. C. 2000 Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling and Software* **15** (1), 101–124.
- Mamo, T., Juran, I. & Shahrour, I. 2014 Virtual DMA municipal water supply pipeline leak detection and classification using advance pattern recognizer multi-class SVM. *Journal of Pattern Recognition Research* **9** (1), 25–42.
- Mashford, J., De Silva, D., Burn, S. & Marney, D. 2012 Leak detection in simulated water pipe networks using SVM. *Applied Artificial Intelligence* **26** (5), 429–444.
- Mostafapour, A. & Davoudi, S. 2013 Analysis of leakage in high pressure pipe using acoustic emission method. *Applied Acoustics* **74** (3), 335–342.
- Mounce, S. R. & Machell, J. 2006 Burst detection using hydraulic data from water distribution systems with artificial neural networks. *Urban Water Journal* **3** (1), 21–31.
- Mounce, S. R., Day, A. J., Wood, A. S., Khan, A., Widdop, P. D. & Machell, J. 2002 A neural network approach to burst detection. *Water Science and Technology* **45**, 237–246.
- Mounce, S. R., Khan, A., Wood, A. S., Day, A. J., Widdop, P. D. & Machell, J. 2003 Sensor-fusion of hydraulic data for burst detection and location in a treated water distribution system. *Information Fusion* **4** (3), 217–229.
- Mounce, S. R., Boxall, J. B. & Machell, J. 2010 Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows. *Journal of Water Resources Planning and Management* **136** (3), 309–318.
- Muggleton, J. M., Brennan, M. J. & Pinnington, R. J. 2003 Wavenumber prediction of waves in buried pipes for water leak detection. *Journal of Sound and Vibration* **249** (5), 939–954.
- Perrier, S., Moslemi, M., Bruce, R., Valentin, B. & Ian, M. 2019 Characterization of soil attenuation for acoustic condition assessment of water-filled metal pipes. In: *International Congress on Sound and Vibration*, Montreal, Canada.
- Pudar, R. S. & Liggett, J. A. 1992 Leaks in pipe networks. *Journal of Hydraulic Engineering* **118** (7), 1031–1046.
- Romano, M., Kapelan, Z. & Savić, D. A. 2012 Real-time leak detection in water distribution systems. In: *Proceedings of the 12th International Conference, WDSA*, September 2010, Tucson, AZ, pp. 1074–1082.
- Romano, M., Kapelan, Z. & Savić, D. A. 2014 Automated detection of pipe bursts and other events in water distribution systems. *Journal of Water Resources Planning and Management* **140** (4), 457–467.
- Shen, C. 2018 A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research* **54** (11), 8558–8593.
- Snider, B. & McBean, E. A. 2018 Improving time-to-failure predictions for water distribution systems using gradient boosting algorithm. In: *1st International WDSA/CCWI 2018 Joint Conference*, July 2018, Kingston, Ontario, Canada.
- Solomatine, D. P. & Ostfeld, A. 2008 Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics* **10** (1), 3–22.
- Ting, L. L., Tey, J. Y., Tan, A. C., King, Y. J. & Faidz, A. R. 2019 Improvement of acoustic water leak detection based on dual tree complex wavelet transform-correlation method. *IOP Conference Series: Earth and Environmental Science* **268** (1), 012025.
- Tyralis, H., Papacharalampous, G. & Langousis, A. 2019 A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water (Switzerland)* **11** (5), 910.
- Webb, G. I. & Zheng, Z. 2004 Multistrategy ensemble learning: reducing error by combining ensemble learning techniques. *IEEE Transactions on Knowledge and Data Engineering* **16** (8), 980–991.
- Wolpert, D. H. 1992 Stacked generalization. *Neural Networks* **5** (2), 241–259.
- Wu, H., Chen, Y., Lin, W. & Wang, F. 2018 Novel signal denoising approach for acoustic leak detection. *Journal of Pipeline Systems Engineering and Practice* **9** (4), 1–9.
- Wylie, E. B. & Streeter, V. L. 1978 *Fluid Transients*. McGraw-Hill International Book Co., New York.
- Yang, J., Wen, Y., Li, P. & Wang, X. 2013 Study on an improved acoustic leak detection method for water distribution systems. *Urban Water Journal* **10** (2), 71–84.
- Zhou, M., Pan, Z., Liu, Y., Zhang, Q., Cai, Y. & Pan, H. 2019 Leak detection and location based on ISLMD and CNN in a pipeline. *IEEE Access* **7**, 30457–30464.

First received 20 May 2020; accepted in revised form 25 December 2020. Available online 25 January 2021