Water Resources Management (2025) 39:2551–2566
https://doi.org/10.1007/s11269-024-04077-3

AIoT-Driven Leak Detection in Real Water Networks Using Hydrophones

Beenish Bakhtawar[1] · **Ali Fares**[1] · **Tarek Zayed**[1]

Received: 10 October 2022 / Accepted: 14 December 2024 / Published online: 26 December 2024

Abstract
Acoustic sensing technology is a familiar approach to detect leakage in urban water networks. Critical issues like false alarms, difficult leak locations, missed leaks, unknown site conditions, and high repair costs are still prevalent. The situation warrants developing a more sophisticated and efficient leak detection approach in real water networks. Hydrophone based acoustic technology has a strong promise for high precision detection of leaks. However, AIoT approach using hydroacoustic data for real water leak detection are rarely reported. The current study, therefore, proposes an integrated signal analysis and machine learning-based ensemble model for leak detection using a hydrophone-based smart IoT system. The results show that the most significant features are peak frequency and maximum amplitude. Random forest is the most robust classifier for cost effective long-term monitoring, and the proposed voting ensemble classifies leaks and no leaks with high accuracy on both unseen data and new sites. Specifically, proposed models have very few alarms and missed leaks are reported, a significant problem in models developed using accelerometers and noise loggers. The study shows a significant contribution to the domain of leak detection for real urban water networks.

Keywords Acoustic Leak Detection · Urban Water Network · AIoT · Hydrophones

1 Introduction

Leaks in dense urban water systems are a vexing problem amidst the ongoing global water crises. Water leakage forms a major part of yearly water losses estimated at 126 billion cubic meters (Bakhtawar and Zayed 2023). False alarms, missed leaks, and inaccurate leak localization create further woes by significantly escalating pipeline repair costs (Islam et al. 2022). The situation warrants proactive identification of leaks for proactive maintenance and prevention of pipeline bursts maintaining continuous water supply. Acoustic sensing

✉ Beenish Bakhtawar
beenish1.bakhtawar@connect.polyu.hk

[1] Department of Building and Real Estate, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

technologies can be used for both leak detection, localization and monitoring (Wu et al. 2024). Hydrophones are a non-invasive acoustic technology that can effectively study the in-pipe acoustic signature to indicate presence of leaks (Wang et al. 2021). They considerably differ from out-of-pipe sensing technologies including accelerometers and noise loggers that are significantly prone to background noise (Huang et al. 2025; Leonzio et al. 2024). They are attached to the fire hydrant because of which it is less sensitive to ambient noises (Bakhtawar and Zayed 2021). Past studies suggest that hydrophones have a similar or higher accuracy results than accelerometers, pressure sensors, or ground microphones (Leonzio et al. 2024). Remarkably, Wang et al. (2021) suggests that hydrophones can capture data of high resolution enabling detection at even low frequencies and under high attenuation conditions. Such studies provide empirical evidence to use the hydrophone sensors for identifying real unstimulated leaks in urban water distribution networks (Xu et al. 2019). However, most relevant research in the domain focuses on lab and test bed scope with simplified network conditions (Iwanaga et al. 2022). Recently, IoT-based acoustic monitoring setups are proposed as an innovative solution for diagnostics of invisible leaks in real water networks (Gong et al. 2020).

Thus, experiments for accurate water leak detection are crucial for practical implementation. For acoustic data modeling, signal analysis-based AI solutions have promising potential without historical failure records and utility drawings. For example, Wu et al. (2024) used deep learning classifiers to study the effect of data augmentation on acoustic leak detection. Zhang et al. (2023) used a CNN-based crack and leak detection model in a smart water network. Kammoun et al. (2023) used an unsupervised LSTM to detect and localize leaks using acoustic data. As an alternative, tree-based models show similar or improved detection performance at a lower cost (Fan et al. 2022b). Although some studies employ tree-based ensemble approaches for noise loggers (Tijani et al. 2022), models using hydrophone data are seldom reported. Consequently, the current study proposes a hydrophone-based leak detection experiment in real water networks. The main study objectives are: (1) Demonstrate an intelligent IoT setup for leak detection using hydrophones for real water networks; (2) Develop a signal analysis and AI based ensemble model for identifying leaks from hydroacoustic data; (4) Test the developed machine learning ensemble model under different conditions and (5) Identify the most significant features for predicting leaks.

2 Research Methodology

The study focuses on water leak detection in real urban water networks in Hong Kong using hydrophone-based data. An AIoT setup for hydroacoustic leak detection is proposed comprising a hydrophone sensor and acoustic logger setup with cellular enabled real-time data recording and transmission system. A hydroacoustic signal based machine learning ensemble is proposed to classify leaks and non-leaks signals. The adopted experimental design is further detailed out in Sect. 2.4. Data is collected over a period of seven months and preliminary acoustic signal analysis is used to study the leak and no-leak signature. Feature extraction in both time and frequency domains is then used for building a machine learning based ensemble for accurate leak classification (Fan et al. 2022a). For the AI modeling, several classifiers are tested for their comparative performance based on related studies (Bakhtawar and Zayed 2023). Selected models are then further validated using unseen data from new sites from Hong Kong using a proposed voting ensemble comprising best performance classifiers. Research methodology adopted in the study is shown in Fig. 1a.

2.1 Data Collection

The WSD ensures effective communication between research team and local repair, providing timely leak alarm information and facilitating on-site data collection. Upon receiving any leak alarm, a hydrophone is deployed on any nearby hydrant on the location with a certified expert from the local Water Services Department (WSD) to collect the desired data. Once the hydrophone is deployed, data recording is initiated by the research team. Data was collected from different districts of Hong Kong over seven months. In total, 13 different sites were studies for both leak and no-leak conditions. The pipe diameter ranged from 100 to 300 mm, both metal and non-metal types, with most pipes being made of ductile iron and polyethylene.

An HWM PermaNET+TM based IoT setup consisting of a single hydrophone and an acoustic logger assembly is proposed for data collection, as shown in Fig. 1b (i). A standard London round thread adaptor, shown as the golden metal ring in Fig. 1b (i), is used

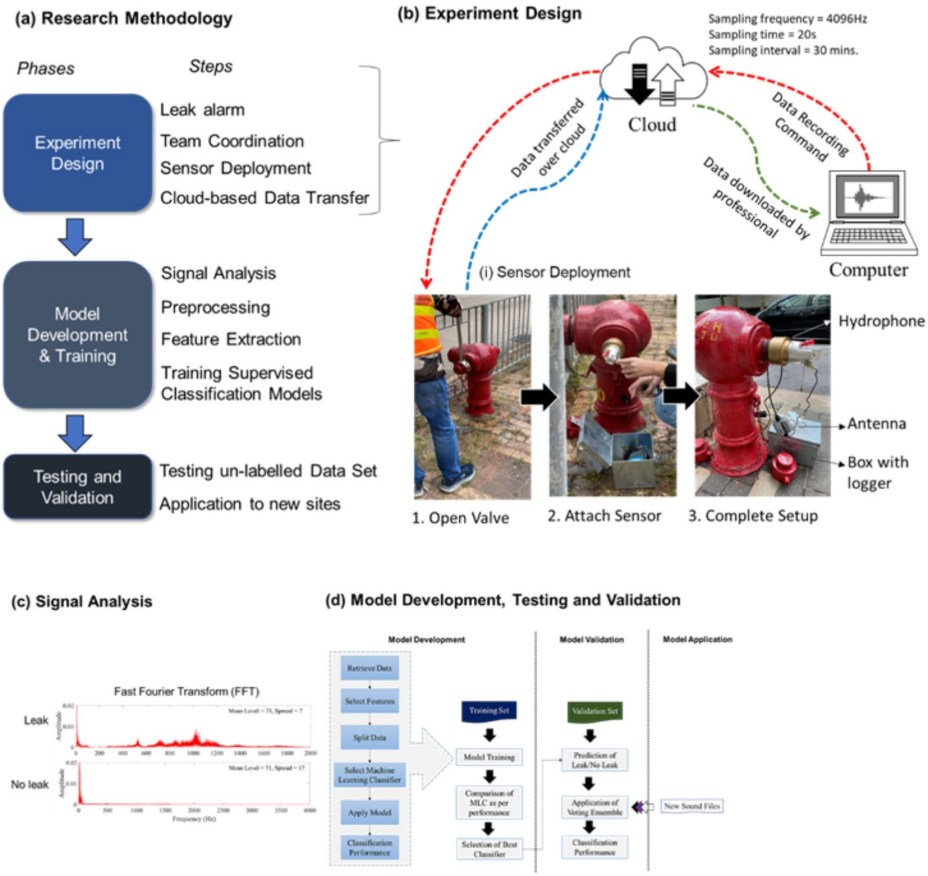

Fig. 1 Research methodology

for connecting the hydrophone with the hydrant. The acoustic logger assembly consists of an integrated antenna enabled with real-time data collection and transmission via low-cost telemetry and an online cloud platform for preliminary data storage and analysis. The data can be downloaded on any internet enabled device for further modeling and analysis. The overall data collection protocol is also explained in Fig. 1b (ii). Signals from the pipelines were recorded at regular 30 min intervals between the quiet times between 3:30 am and 4:30 am every night for 5–7 continuous nights, at a sampling frequency of 4096 Hz.

2.2 Signal Analysis

Acoustic signal analysis based on Fast Fourier transform (FFT) is a powerful preliminary tool for detecting anomalies when continuous data is available (Fares et al. 2023). A representative leak and no-leak acoustic signal recorded from the Hong Kong water network is presented in Fig. 1c. In presence of a water leak, the high-frequency region in FFT shows a strong peak observed consistently for many days. Conversely, in case of NoLeak conditions, the low-frequency shows a peak which probably represents the self-noise of the system. A significant disparity exists between the amplitude of leak and no-leak signals, suggesting the presence of a constant noise source in the system. However, to judge whether the noise source is truly a leak, a more in-depth analysis using an AI-aided classification is desired (Nimri et al. 2023). Mathematically, the FFT algorithm using a discrete transform is expressed as per Eq. 1.

$$X_k = \sum{}_{n=0}^{N-1} x_n e^{-2\pi\, ikn/N} \tag{1}$$

where X_k is the output expressed as a complex number with both amplitude and phase knowledge for time increments at equal spaces, t = 0, 1, ……., N-1. Overall, FFT has several advantages for water leak detection like high sensitivity for early prognosis and resource allocation for further investigation. As compared to time domain analysis, it is easier to distinguish leak from no-leak signals for distinguishing ambient noise frequency ranges for denoising (Sitaropoulos et al. 2023).

2.3 Model Development for Water Leak Detection

After preliminary analysis using signal analysis, AI-based voting ensemble is proposed using the steps covered in this section. The methodology adopted for the model development is explained in Fig. 1d. The recorded signals are pre-processed using linear prediction and split into individual files of one second each to increase signal population. The dataset is split using the 80:20 ratio. 20% of the data was separated for testing, which contained unlabeled and unseen data from the sites. Selected machine learning classifiers were compared for their performance to detect water leaks. Finally, acoustic recordings from new sites are collected to validate the voting ensemble and predict the new sound files' Leak/NoLeak status.

2.3.1 Feature Extraction Using Linear Prediction

Acoustic data is complex to analyze and needs further transformation to retrieve the most relevant information to detect presence of water leakage in the systems. Feature extraction aids in signal characterization by selecting relevant features containing important information on leaks. Hu et al. (2021) and Fan et al. (2022a) studied various feature extraction methods exist for acoustic signals. Linear prediction (LP) for feature extraction from hydroacoustic data has shown promise for leak detection in real urban water networks (Cody and Narasimhan 2020; Tijani et al. 2022). A comparable LP-based feature extraction has been adopted in the current study for real water distribution networks. The LP algorithm uses an autoregressive algorithm to computes predictor coefficients, as shown in Eq. 2. Resulting reconstructed signals can help identify even minor deviations and anomalies in the acoustic signature (Hu et al. 2021).

$$s\left[n\right] = \sum {}_{k=1}^{p} a_k s\left[n-k\right] + e\ \left[n\right] \tag{2}$$

where a_k are the linear predictor coefficients of p^{th} order and the residual prediction error is represented as $e\ \left[n\right]$. Post LP transformation, spectral features belonging from both the time and frequency domains are extracted from signals, as presented in Table 1.

2.3.2 Algorithm Selection and Parameters

To evaluate the effectiveness for leak detection, several supervised learning approaches were evaluated. A total of eleven different AI classifiers, all suitable for classification problems, were selected for comparison. These methods are classified as Lazy learners, Bayesian methods, artificial neural networks, decision trees, ensemble models, and support vector machines (Ahmed et al. 2024). The model parameters used in the study are shown in Table 2. Many of these supervised learning classifiers have not been tested in past studies to predict leaks using hydroacoustic data.

Table 1 Explanation of extracted features

Features	Type	References
Average Level	Time Domain	(Dhifaoui, 2019; Fan et al., 2022a; Tariq et al., 2022; Tijani et al. 2022)
Spread	Time Domain	
Root Mean Square	Time Domain	
Maximal Lyapunov exponent (MLE)	Time Domain	
TD Average Amplitude	Time Domain	
Autocorrelation Kurtosis	Time Domain	
Skewness	Frequency Domain	
Peak Amplitude	Time Domain	
Crest Factor	Time Domain	
Total Energy	Time Domain	
Autocorrelation MLE	Time Domain	
FD Average Amplitude	Frequency Domain	
Maximum Amplitude	Frequency Domain	
Frequency Spread	Frequency Domain	
Peak Frequency	Frequency Domain	
Kurtosis	Frequency Domain	

2.4 Classification Performance Metrics

Various classification performance metrics used are presented in the current section as follows:

2.4.1 Confusion Matrix for Leak Detection

Common classification performance indicators that can be used for model evaluation are accuracy, precision, and recall as depicted in Eq. 3 to Eq. 5. These metrics indicate the correct predictions resulting from the model. For the leak detection problem, however, it is also beneficial to also investigate the true leak rate (TLR), true noleak rate (TNR), false alarm rate (FAR), and missed leak rate (MLR), calculated using Eqs. 6 to 9.

$$Classification\ Accuracy = \frac{True\ Leak + True\ NoLeak}{Total\ Sample} \tag{3}$$

$$Precision = \frac{TL}{TL + FA} \tag{4}$$

Table 2 Parameters of supervised machine learning models

Type	Classifier	Parameters
Lazy	k-nearest neighbor (KNN)	K=5. Distance measure=Euclidean
Bayesian	Naïve-Bayes (Kernal)	Estimation Mode=Greedy; Maximum Bandwidth=0.1 Number of kernels=10
Artificial Neural Network	AutoMLP	Training Cycles=10; Number of generations=10
	Deep Learning	Activation=Rectifier; Loss Function=Automatic
	Neural Net	Training Cycles=200; Learning Rate=0.01
Trees	Decision Tree	Criterion=gain ratio; Minimal size for split=4. Minimal leaf size=2; Minimal gain=0.01 Maximal depth=10; Confidence=0.1
Ensembles	Gradient Boosted Trees	Number of trees=50; Maximal depth=5; Min. rows=10 Number of bins=20; Learning Rate=0.01
	Random Forest	Criterion=gain ratio; Minimal size for split=4 Minimal leaf size=2; Minimal gain=0.1 Maximal depth=20; Confidence=0.25
	Adaboost	Iterations=10 Sub-Operator=Decision Tree
Support Vector Machines	SVM	Kernel type=dot; Convergence epsilon=0.001
Regression	Logistic Regression	Solver=Auto

$$Recall = \frac{TL}{TL + ML} \tag{5}$$

$$True\ Leak\ Rate = TLR = \frac{True\ Leak}{(Missed\ Leak + True\ Leak)} = \frac{TL}{(ML + TL)} \tag{6}$$

$$True\ NoLeak\ Rate = TNR = \frac{True\ No\ Leak}{(False\ Alarm + True\ No\ Leak)} = \frac{TN}{(FA + TN)} \tag{7}$$

$$False\ Alarm\ Rate = FAR = \frac{False\ Alarm}{(False\ Alarm + True\ No\ Leak)} = \frac{FA}{(FA + TN)} \tag{8}$$

$$Missed\ Leak\ Rate = MLR = \frac{Missed\ Leak}{(Missed\ Leak + True\ Leak)} = \frac{ML}{(ML + TL)} \tag{9}$$

2.4.2 F1 Score Comparison

The F1 score is a combined measure of classification accuracy that considers both precision and recall metrics for a balanced evaluation. Mathematically, F1 can be expressed as the harmonic mean of precision and recall as per Eq. 10.

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{10}$$

It is important to note that the F1 score concentrates only on the positive prediction results. Therefore, the results depend on which category is designated as positive in the binary classification. Reversing the positive and negative categories will render the F1 score meaningless. Hence, selecting the positive category carefully is crucial for accuracy.

2.4.3 Matthews's Correlation Coefficient (MCC)

The Matthews Correlation Coefficient (MCC) is an effective measure of binary classification, especially in case of imbalanced datasets. Equation 11 shows that a specific advantage of using the MCC is that combines multiple classification characteristics into a single metric. The metric of MCC is basically a correlation coefficient between observed and predicted classes, as shown by Eq. 11 of the confusion matrix.

$$MCC = \frac{(TPxTN) - (FPxFN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{11}$$

For the data collected in the current study, acoustic signatures from NoLeak cases are much higher in number than the leak cases. Therefore, MCC is a comparatively robust measure to demonstrate the level of agreement between observed and predicted class values. For interpretation, the values of MCC range from -1 and $+1$. A value closer to -1 indicates total divergence, but a value closer to $+1$ indicates perfect conformity between predicted and observed values. Additionally, a value of 0 means that the prediction is random, just like a coin flip.

2.4.4 ROC Curves

Receiver operating characteristic (ROC) curves are a graphical plot used to evaluate the performance of a binary classification operator. The ROC curves show the trade-off between true leak rate (sensitivity) and the false alarm rate (specificity) for the selected classifier against different settings. A classifier with the ROC curve bending closer towards the top left corner is considered to have better performance in terms of higher sensitivity and lower specificity. The area under the ROC curve (AUC) can be used to quantify the ability of the model to discriminate the Leak and No Leak classes. An AUC of 1 denotes a perfect model and an AUC of 0.5 or below denotes a random model.

2.5 Voting Ensemble Based Validation

A voting ensemble is a machine learning method that integrates the prediction results of multiple learners and presents a final prediction based on hard voting for classification problems. The resulting class is the result of most votes from the individual models improving prediction reliability, stability, avoiding risk of overfitting. In the study, a voting ensemble is proposed for the best performing algorithms. The voting ensemble was validated using both (i) unseen data and (ii) new data received from the water and services department to test the AIoT setup proposed in the study. For the testing on unseen data, 20% of unlabeled data was saved for testing the model on unseen data. Furthermore, Fares et al. (2023) proposed to evaluate leak detection models on data collected from new sites not considered during model developed. Consequently, new sound files were received from the contractor from four different test sites around Hong Kong. In total, 120 new data recordings from four new sites were used to further validate the model.

2.6 Feature Selection

Feature selection is an optimization method used to improve model performance by optimum selection of feature sub-set from the feature space (Fan et al. 2022a). It is essential to select the most significantly contributing features to model. The features contributing most to the model prediction capabilities were explored using two feature selection mechanisms: (1) Attribute weights and (2) Brute force. For the attribute weight's function, there are two options to select the features: (a) Forward Selection and (b) Backward Selection operators (Fan et al. 2022a; Tijani et al. 2022). Both the attribute weight-based methods are stepwise regression methods but they approach the feature selection task in converse to each other. The forward selection starts the iterations with an empty set, adding the features one by one, evaluating their impact on the model performance. In comparison, the backward selection begins with the full set of features, eliminating the features one by one through the iterations. The brute force method carries out an exhaustive iteration based on all possible combinations of features selecting the feature set giving the optimized model performance.

3 Results and Discussion

3.1 Training and Testing Model Performance

3.1.1 Confusion Matrix Results

Confusion matrix results are illustrated in Table 3 for both training and testing data. Overall, all the learners show high classification accuracy. However, accuracy is a limited metric of performance. Using only the accuracy can be misleading because it is calculated only based on correct classifications. In the case of leak diagnostics, insights into false alarms, and missed leaks are crucial for the timely detection of leaks. False alarms in leak detection systems have significant negative consequences like additional repair costs, unnecessary digging, and extra workforce. Similarly, missed leaks can remain in the system as undetected leaks eventually causing large bursts in the pipeline. Hence, false alarm rate (FAR) and missed leak rate (MLR) are crucial but undermined metrics for water leak detection.

Based on these criteria, the best-performing classifiers for the data are Adaboost and Random Forest considering the least level of MLR and FAR, as well. Adaboost has a 100% true prediction rate for the training set, followed by Random Forest. For the testing set, both learning algorithms have the lowest FAR/MLR levels (Random Forest = 0.54/1.61 and Adaboost = 0.73/1.61). Naïve Bayes (Kernal), ANN and AutoMLP also show promising classification performance with low false alarms but have a slightly high rate of missed leaks.

Table 3 Performance of different predictive operators

Predictive Operator	Data Type	Accuracy	Precision	Recall	TLR	FA *R*	TNR	MLR
Deep Learning	Training	96.49	98.20	66.40	66.40	0.14	99.86	33.60
	Testing	96.74	97.73	69.35	69.35	0.18	99.82	30.65
Logistic Regression	Training	99.55	98.36	97.17	97.17	0.18	99.82	2.83
	Testing	98.53	93.44	91.94	91.94	0.73	99.27	8.06
Naïve-Bayes (Kernal)	Training	99.96	100.00	99.60	99.60	0.00	100.00	0.40
	Testing	99.35	98.33	95.16	95.16	0.18	99.82	4.84
Random Forest	Training	99.96	100.00	99.60	99.60	0.00	100.00	0.40
	Testing	99.35	95.31	98.39	98.39	0.54	99.46	1.61
SVM	Training	96.82	99.42	68.83	68.83	0.05	99.95	31.17
	Testing	96.9	97.78	70.97	70.97	1.96	98.04	29.03
ANN	Training	99.76	98.79	98.79	98.79	0.14	99.86	1.21
	Testing	98.86	92.31	96.77	96.77	0.91	99.09	3.23
k-NN	Training	99.59	98.37	97.57	97.57	0.18	99.82	2.43
	Testing	98.68	92.31	96.77	96.77	0.91	99.09	3.23
Gradient Boosted Trees	Training	99.35	96.02	97.57	97.57	0.45	99.55	2.43
	Testing	98.86	91.04	98.39	98.39	1.09	98.91	1.61
Decision Tree	Training	99.76	98.79	98.79	98.79	0.14	99.86	1.21
	Testing	99.02	92.42	98.39	98.39	0.91	99.09	1.61
Auto MLP	Training	99.71	99.18	97.98	97.98	0.09	99.91	2.02
	Testing	99.18	93.85	98.39	98.39	0.73	99.27	1.61
AdaBoost	Training	100	100.00	100.00	100.00	0.00	100.00	0.00
	Testing	99.18	93.85	98.39	98.39	0.73	99.27	1.61

True Leak Rate (TLR), False Alarm Rate (FAR), True No-Leak Rate (TNR), Missed Leak Rate (MLR)

Thus, it can be concluded that the ensemble-based supervised learning classifiers are more accurate and robust when classifying leak/no leak acoustic data.

3.1.2 F1 Score and MCC Coefficient

To further study the model performance, F1 and MCC scores for the training and testing data sets are shown in Fig. 2a and b, respectively. Overall, F1 scores for the developed models ranged between 0.81 and 0.97 when tested in the training data set. For the testing set, F1 scores also show a good agreement lying in the range of 0.79–1.00.

As per the results, Random Forest, Auto MLP, Naïve-Bayes (Kernel), Neural Net and Adaboost have the highest values of F1 Scores during both training and testing. Furthermore, the MCC results for these algorithms are close to 1 showing the flexibility of the models to predict both leaks and no leak situations, efficiently. On the contrary, SVM and deep learning were among the worst performing algorithms as per both the F1 and MCC scores further validating the confusion matrix results.

3.2 Validation of Proposed Voting Ensemble

A voting ensemble is set up using the best-performing five classifiers. The infrastructure of the voting ensemble is shown in Fig. 3a. The performance of the developed ensemble is tested both using (i) unseen data and (ii) new data received from the water and services department to test the AIoT setup proposed in the study. The voting ensemble runs all five classifiers simultaneously. Then, it takes a vote on the data classification as a leak or no leak, as explained in Fig. 3a.

3.2.1 Testing for Unseen Data

For the current study, the developed voting ensemble was most effective in detecting leaks with no misclassifications and consequent 100% testing accuracy on the unlabeled data set as shown in Fig. 3b. Figure 3b shows that all the developed models gave very promising accuracy levels individually as well. However, for accuracy, the voting ensemble is considered the preferred method for leak detection in real networks.

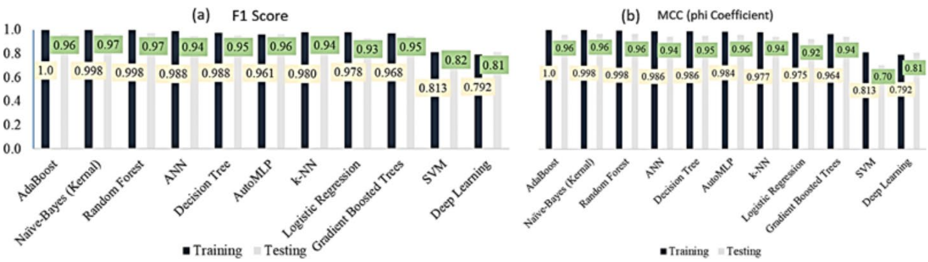

Fig. 2 (**a**) F1 Score of different classifiers (**b**) MCC score of different classifiers

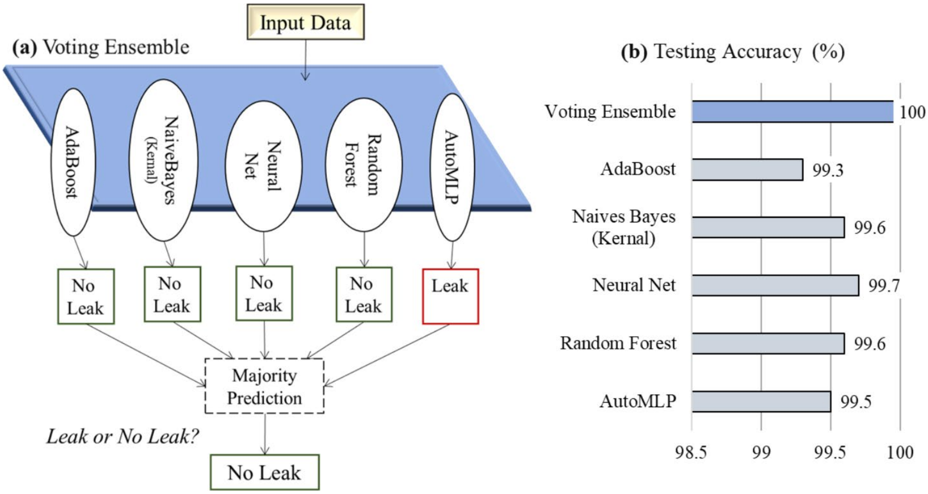

Fig. 3 Voting model architecture and testing accuracy

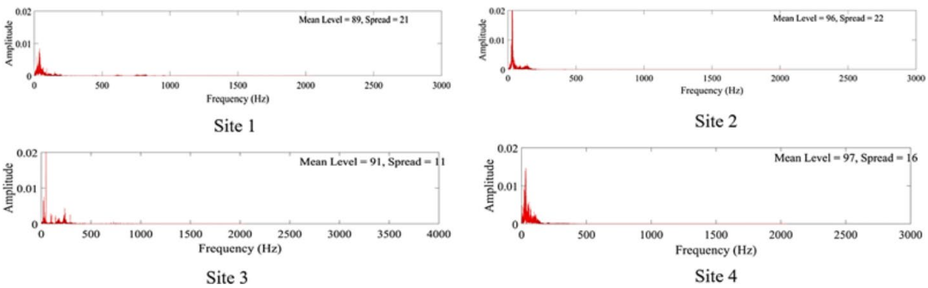

Fig. 4 Signal samples from four new sites

3.2.2 Testing for New Sites

Some frequency spectrums of signals from the new sites are shown in Fig. 4. The signals from the new sites show peaks in the very low-frequency zones akin to other no-leak signals, as shown in Fig. 1c. Previous literature also highlights that very low-frequency acoustic signals may be the system's internal noise (Gupta and Kulat 2018). Thus, these signals were considered to be possible false alarms.

To confirm the inference from the FFT, the developed ensemble model was used to predict the leak/no-leak status of the examined sites. The model results match the signal analysis showing a no-leak status for all the sites. A field inspection was then conducted which confirmed the results of the model to be correct. More data from other sites with real leaks can further help validate the real network applicability and practicality of the model implementation.

3.3 Feature Selection

The original hydrophone model was developed using 17 features enlisted in Table 1. Results of the feature selection differ for forward and backward selection. According to the forward selection, peak frequency and maximum amplitude are the most significant features for water leak detection models. In contrast to the forward selection, only the maximum amplitude is eliminated by the backward selection. Rest all the features are considered significant. As per the brute force method, MLE, energy, crest factor, autocorrelation kurtosis, autocorrelation MLE, and peak frequency are among the significant features showing the highest model performance. Both the forward selection and brute force method successfully reduce the number of features showing the viability to be used for cost-effective model application. Results show peak frequency to be a top feature by all three feature selection methods, as shown in Fig. 5, which shows a clear difference between a leak and no leak signals. So, the model performance was also observed based only on the peak frequency. Performance comparison with the original model and a model developed only based on peak frequency was carried out.

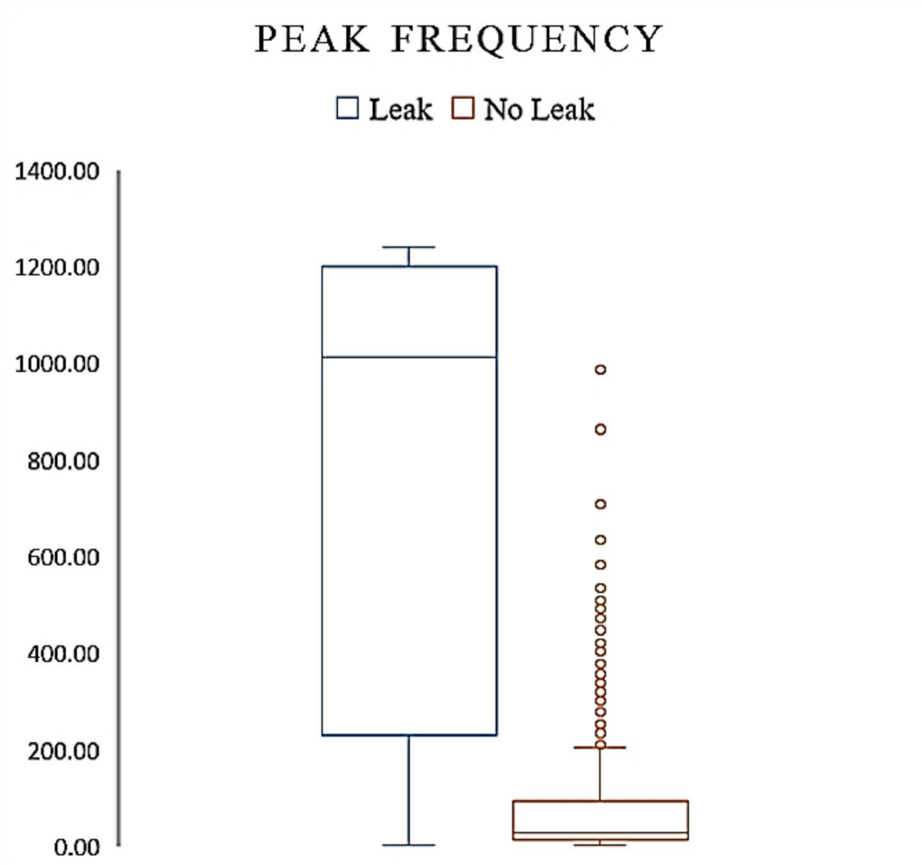

Fig. 5 Peak Frequency Histogram for Leak and No Leak Signals

3.3.1 Comparison of Original and Optimized Models

The comparative performance of feature selection and the original model for two of the top performing algorithms, Adaboost and Random Forest is observed as shown in Table 4. From Table 4, in the Adaboost based model, MLR and FAR are increasing with the decreasing number of features. In comparison, however, the Random Forest model shows better performance, as its MLR and FAR decrease with the decrease in the number of features selected. When trained with just the peak frequency, the Random Forest and Adaboost both show an increase in the MLR and FAR. Thus, leak and no-leaks cannot be identified using a single feature. Overall, Random Forest show less sensitivity and more robust performance with the features selected using Brute Force.

3.3.2 ROC Curves

For the leak and no leak classification, plotting the ROC curve can give a deeper insight into the diagnostic ability of any classifier to classify leaks and no leaks. For the four best performing models, the ROC curves are plotted to illustrate their classification prediction and the confidence level against different settings, as shown in Fig. 6. Three different settings are used to test the performance: (i) optimistic, (ii) neutral, and ii) pessimistic.

First, the ROC curves show that Adaboost and Random Forest show similar performance under optimistic bias conditions. Secondly, Random Forest shows less sensitivity towards the change in features than the Adaboost over the three settings. Overall, Adaboost with seventeen features, and Random Forest with two features are observed to have the highest performance under different conditions. However, Random Forest has a greater flexibility and robustness with seldom incorrect predictions. Using only two features, the model is a good option for cost-effective real water network leak monitoring.

Table 4 Performance metrics for original and feature selection-based models

Predictive Operator	Feature Selection Method	Data Type	TLR	FAR	TNR	MLR	F1 Score	MCC
AdaBoost	Original Model	Training	100.00	0.00	100.00	0.00	1.000	1.00
		Testing	98.39	0.73	99.27	1.61	0.961	0.96
AdaBoost _BF	Brute Force	Training	97.98	0.05	99.95	2.02	0.988	0.99
		Testing	98.39	0.73	99.27	1.61	0.961	0.96
AdaBoost _FS	Forward Selection	Training	94.33	0.18	99.82	5.67	0.963	0.96
		Testing	95.16	0.91	99.09	4.84	0.937	0.93
AdaBoost _PF	Peak Frequency	Training	88.26	0.14	99.86	11.74	0.932	0.93
		Testing	91.94	0.54	99.46	8.06	0.934	0.93
Random Forest	Original Model	Training	99.60	0.00	100.00	0.40	0.998	1.00
		Testing	98.39	0.54	99.46	1.61	0.968	0.96
Random Forest _BF	Brute Force	Training	99.60	0.00	100.00	0.40	0.998	1.00
		Testing	100.00	0.36	99.64	0.00	0.984	0.98
Random Forest _ FS	Forward Selection	Training	97.98	0.00	100.00	2.02	0.990	0.99
		Testing	98.39	0.36	99.64	1.61	0.976	0.97
Random Forest _PF	Peak Frequency	Training	91.90	0.00	100.00	8.10	0.958	0.95
		Testing	95.16	0.54	99.46	4.84	0.952	0.95

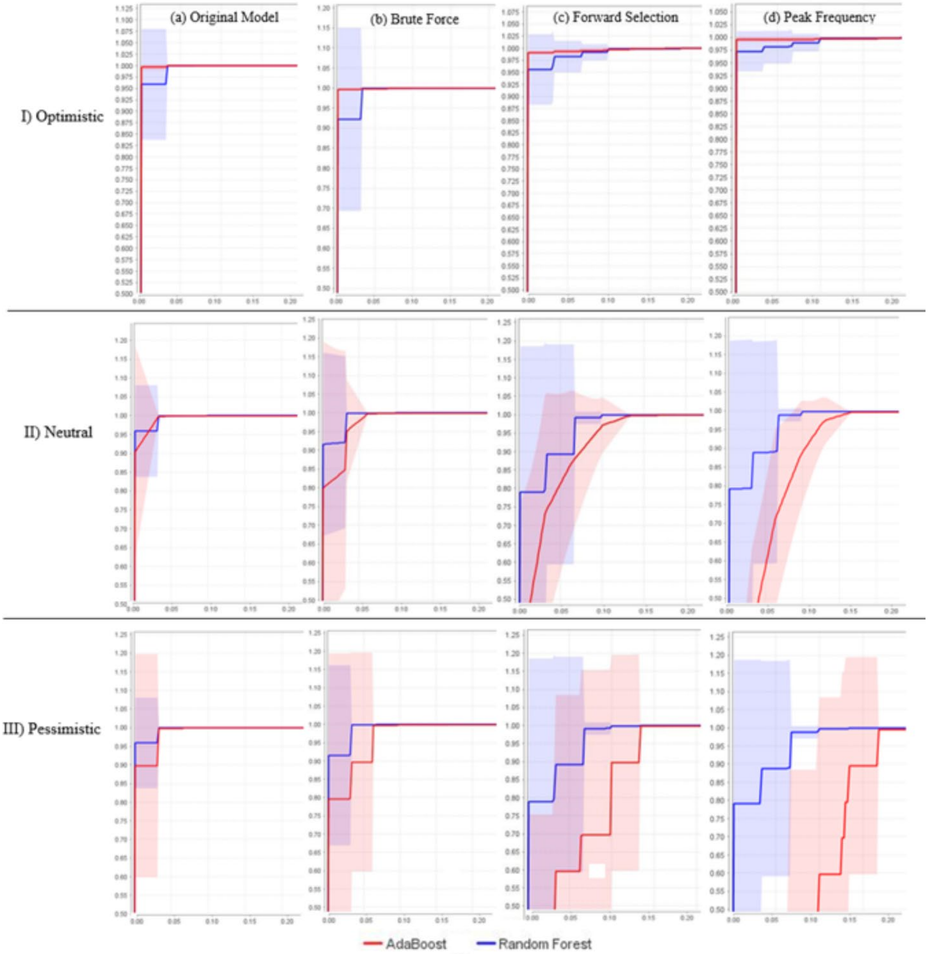

Fig. 6 ROC curves for Adaboost and Random Forest

4 Conclusion

The study focuses on the detection and identification of real leaks in dense urban environments based on hydroacoustic data using an IoT experiment design. An integrated signal analysis and machine learning ensemble is proposed to classify leaks and no-leak data. Rich data from different locations in Hong Kong is collected under the supervision of local contractors and the government water and sanitation department. A comparison of machine learning classifiers reveals that Adaboost and Random Forest are the best-performing classifiers with meager false positive and false negative rates. Further feature selection based on forward, backward, and brute force concludes that the random forest-based model using peak frequency and maximum amplitude performs best.

The main highlights of the study include (i) An innovative IoT-based intelligent system to identify real leaks in real water networks. (ii) Only the leak signature collected from the site is used for model development. So, the proposed ensemble model does not need site information to predict leaks. (iii) A cost-effective model based on only two significant features is also proposed for long-term monitoring. (iv) The testing results show the model is robust against missed leaks and false alarms, which is a primary concern in models based on accelerometers and noise logger data. These results represent the effectiveness and innovation of the study findings.

One activation limitation of the hydrophones, however, is related to difficulties in data collection. The on-site deployment of hydrophones is more complicated than out-of-pipe sensors. To open the hydrant valve for hydrophone deployment, official permissions and coordination from the government departments are necessary, making it a time consuming effort especially for high-urgency leak cases. Future research using image classification of the leak signature can increase detection efficiency and aid in leak localization. A multi-sensor setup using hydrophone and noise logger data integration can also help identify difficult leaks.

Author Contributions All authors contributed to the study. Prof. Tarek lead the conception and design of the study supported by Beenish Bakhtawar. Prof. Tarek is also the principal investigator of the project and supervisor of the study. Data analysis was performed by Beenish Bakhtawar. The first draft of the manuscript was written by Beenish Bakhtawar supported by Ali Fares. Ali Fares also helped with the interpretation of results. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by The Hong Kong Polytechnic University. The authors gratefully acknowledge the support from the Innovation and Technology Fund (Innovation and Technology Support Programme (ITSP)) under grant number ITS/067/19FP and the Water Supplies Department of the Government of Hong Kong.

Data Availability All data supporting the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethical Approval Not applicable.

Consent to Participate Not applicable.

Consent to Publish Not applicable.

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

References

Ahmed AA, Sayed S, Abdoulhalik A, Moutari S, Oyedele L (2024) Applications of machine learning to water resources management: a review of present status and future opportunities. J Clean Prod 441:140715. https://doi.org/10.1016/j.jclepro.2024.140715

Bakhtawar B, Zayed T (2021) Review of Water Leak detection and localization methods through Hydrophone Technology. J Pipeline Syst 12(4). https://doi.org/10.1061/(ASCE)PS.1949-1204.0000574

Bakhtawar B, Zayed T (2023) State-of-the-art review of leak diagnostic experiments: toward a smart water network. Wiley Interdisciplinary Reviews: Water 10(5):e1667

Cody RA, Narasimhan S (2020) A field implementation of linear prediction for leak-monitoring in water distribution networks. Adv Eng Inform 45:101103

Dhifaoui Z (2019) Scale-wise Lyapunov exponent for bounded variation mappings. SN Appl Sci 1:1081. https://doi.org/10.1007/s42452-019-1122-z

Fan H, Tariq S, Zayed T (2022a) Acoustic leak detection approaches for water pipelines. Autom Constr 138:104226. https://doi.org/10.1016/j.autcon.2022.104226

Fan X, Wang X, Zhang X, Asce X, Yu PEF (2022b) Machine learning based water pipe failure prediction: the effects of engineering, geology, climate and socio-economic factors. Reliab Eng Syst Saf 219. https://doi.org/10.1016/j.ress.2021.108185

Fares A, Tijani I, Rui Z, Zayed T (2023) Leak detection in real water distribution networks based on acoustic emission and machine learning. Environ Technol 44(25):3850–3866. https://doi.org/10.1080/09593330.2022.2074320

Gong J, Lambert MF, Stephens ML, Cazzolato BS, Zhang C (2020) Detection of emerging through-wall cracks for pipe break early warning in water distribution systems using permanent acoustic monitoring and acoustic wave analysis. Water Resour Manage 34:2419–2432

Gupta A, Kulat K (2018) A selective literature review on leak management techniques for water distribution system. Water Resour Manage 32:3247–3269

Hu Z, Chen B, Chen W, Tan D, Shen D (2021) Review of model-based and data-driven approaches for leak detection and location in water distribution systems. Water Supply 21(7):3282–3306

Huang Y, Liu Z, Li S (2025) Leak detection and localization of fluid-filled pipeline using accelerometer pairs and mode separation method. Measurement 239:115491

Islam MR, Azam S, Shanmugam B, Mathur D (2022) A review on current technologies and future direction of water leakage detection in water distribution network. IEEE Access 10:107177–107201

Iwanaga M, Brennan M, Almeida F, Scussel O, Cezar S (2022) A laboratory-based leak noise simulator for buried water pipes. Appl Acoust 185:108346

Kammoun M, Kammoun A, Abid M (2023) LSTM-AE-WLDL: unsupervised LSTM auto-encoders for leak detection and location in water distribution networks. Water Resour Manage 37(2):731–746

Leonzio DU, Bestagini P, Marcon M, Quarta GP, Tubaro S (2024) *Water Leak Detection and Classification Using Multiple Sensors.* Paper presented at the 2024 IFIP Networking Conference (IFIP Networking)

Nimri W, Wang Y, Zhang Z, Deng C, Sellstrom K (2023) Data-driven approaches and model-based methods for detecting and locating leaks in water distribution systems: a literature review. Neural Comput Appl 35(16):11611–11623

Sitaropoulos K, Salamone S, Sela L (2023) Frequency-based leak signature investigation using acoustic sensors in urban water distribution networks. Adv Eng Inform 55:101905

Tariq S, Bakhtawar B, Zayed T (2022) Data-driven application of MEMS-based accelerometers for leak detection in water distribution networks. Sci Total Environ 809:151110–151110. https://doi.org/10.1016/j.scitotenv.2021.151110

Tijani IA, Abdelmageed S, Fares A, Fan KH, Hu ZY, Zayed T (2022) Improving the leak detection efficiency in water distribution networks using noise loggers. Sci Total Environ 821. https://doi.org/10.1016/j.scitotenv.2022.153530

Wang W, Sun H, Guo J, Lao L, Wu S, Zhang J (2021) Experimental study on water pipeline leak using In-Pipe acoustic signal analysis and artificial neural network prediction. Measurement: J Int Meas Confederation 186. https://doi.org/10.1016/j.measurement.2021.110094

Wu Y, Ma X, Guo G, Jia T, Huang Y, Liu S, Wu X (2024) Advancing deep learning-based acoustic leak detection methods towards application for water distribution systems from a data-centric perspective. Water Res 261:121999

Xu JH, Chai KTC, Wu GQ, Han BB, Wai ELC, Li W, Gu YD (2019) Low-Cost, tiny-sized MEMS Hydrophone Sensor for Water Pipeline Leak Detection. IEEE Trans Industr Electron 66(8):6374–6382. https://doi.org/10.1109/tie.2018.2874583

Zhang C, Alexander BJ, Stephens ML, Lambert MF, Gong J (2023) A convolutional neural network for pipe crack and leak detection in smart water network. Struct Health Monit 22(1):232–244