SN Computer Science manuscript No.
(will be inserted by the editor)

dPipe: An IoT-based Water Leak Detection Framework for Smart Factory and Smart Homes

Aryan Kishore Saran · Harsh Barsaiyan · Jay Padia · Debanjan Das ⋆

the date of receipt and acceptance should be inserted later

Abstract Water is the most precious resource for all the living creatures in this world. Proper usage and conservation of water is very important now-a-days. A lot of water is wasted around the globe due to the leakages in the water pipelines. This paper proposes dPipe, an intelligent water pipeline monitoring and leakage detection system for smart factory and smart homes to detect the leakage in the pipelines by using the sound recordings of the surroundings in which the pipeline is present. Machine Learning techniques are used to differentiate between the characteristic sound and other sounds near the pipelines. The kit includes microphone which takes recording regularly after certain interval of time. Moreover, a model has been deployed over Raspberry Pi which predicts about the leakage in the pipelines and sends messages in Telegram about the same. This work do not need human intervention as it automatically gives notifications about the leakages. The proposed setup is able to detect leaks with an accuracy of 89 percent.

Keywords Leakage Detection; Pipeline Monitoring; Internet of Things; Machine Learning; Smart factory; Sound recordings.

Aryan Kishore Saran
Electronics Communication Engg.,IIIT-Naya Raipur, India
E-mail: aryan21101@iiitnr.edu.in

Jay PareshKumar Padia
Electronics Communication Engg.,IIIT-Naya Raipur, India
E-mail: jay21101@iiitnr.edu.in

Harsh Barsaiyan
Electronics Communication Engg.,IIIT-Naya Raipur, India
E-mail: harsh21101@iiitnr.edu.in

Dr. Debanjan Das (Corresponding Author)
Dept. of Electronics and Communication Engg., IIIT–Naya Raipur, India
E-mail: debanjan@iiitnr.edu.in

1 INTRODUCTION

Water and the source of its transportation i.e. pipelines is a very important part of any factory or a home. Water is the basic need of every living creature on this planet. People are facing water shortages all over the world and its wastage through the leak in water pipelines looks ugly. Humans are the most superior creature on this planet. We have the responsibility to find the solution to this type of water wastage.

Pipelines are responsible for transportation of water in the factories, industries and also in our houses. They are made up of those materials which are durable so that it cannot be damaged in the long run. Leakage in these pipelines not only leads to the wastage of water but also needs human intervention to resolve the issue. During such cases, monitoring of the underground pipelines becomes difficult.Hence, in order to minimise water wastage, monitoring and analysis of these pipelines becomes very important and crucial. Talking about the industries, poor maintenance of pipelines which plays a very important role in water transportation will result in financial losses and water wastage. To cop up with the problem, an effective technique is required to minimise such accidents in the industries and factories.

From a long time, researchers are trying to find ways to solve this issue of leakage detection in the water pipelines.Many models have been designed , created and implemented and the result is still going on to find the best solution to the problem so that water wastage can be stopped. In modern times, IoT is used in various fields including medical, banks, educational institutions, offices, smart cities and many more. IoT can help us to notify the officials or people present in the factories and homes as soon as any leak is detected in the pipelines by the model deployed with the Raspberry Pi.

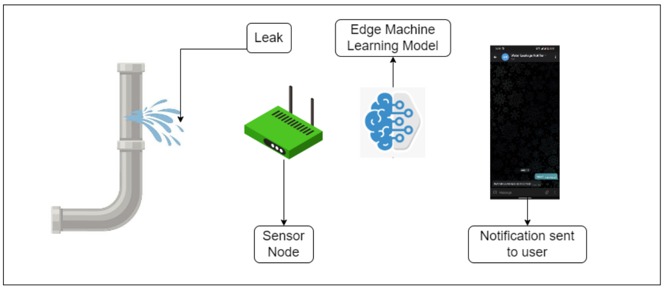

Fig. 1 Working of the proposed model

This paper presents dPipe, a method to continuously monitor the water pipelines for the leakages. It is one of the best methods is the predicting the leakage on the basis of the sound. Fig.1 shows the conceptual view of the dPipe. This paper can be very useful in different industries for pipeline maintenance as it allows the workers and officials to monitor the pipelines without any difficulty. The major contributions of the work include:

– Designing a signal processing and ML based approach for detecting leaks based on sound recordings.
– Reducing the need of human intervention by automatically notifying about the leakage in the pipelines.

The rest of the paper is organised as follows: Section II presents an overview of the existing literature on leak detection research, many hardware and software methods are described briefly. A comprehensive view of the proposed pipeline monitoring and leakage detection system is presented in III. Implementation details, algorithms of the proposed system and simulation results are given in IV. The last section contains the conclusion along with the future scope of this work.

2 RELATED WORKS

We can employ a variety of techniques to find water leaking from pipelines. Some of them are less successful than others at accurately predicting the outcomes. Chemical analysis and acoustic techniques are examples of hardware methods. Utilizing thermal imaging equipment, thermal imaging monitoring technique examines changes in thermal radiation along the pipeline.

A vapour monitoring device is used in a method based on chemical composition analysis. The test tube is a partially absorbable test tube. Gas travels through the tube's membrane and is pushed to a station for testing vapours for analysis. An acoustic emission-based method for the detection of tiny leaks was proposed by Bui at el. Compared to the direct AE-based technique, this one is more noise-resistant and uses a Hanning window. Gas leak detection pipelines were discussed by Adnan et al. The key variables in this strategy are wave propagation and pressure brought on the pipe wall friction. According to this study, Hilbert-Huang transformation and acoustic signals are more reliable ways to find gas pipeline leaks.

Typically, machine learning and signal processing are used in software-based approaches. This paper is also based on it. The pressure signal and time differential change as a result of pipeline leakage. Based on the aforementioned method, the sensors at each end of a leaking point determine the direction of the leak. Data on water flow is processed using the Random Forest technique by Coelho et al. The system has a 75 percent accuracy rate. After a pipeline leak has been discovered, we must select the optimal solution with the highest operating efficiency and the lowest cost of pipeline maintenance.

3 DPIPE: PROPOSED METHODOLOGY FOR PIPELINE LEAKAGE DETECTION

The step by step process used in this paper is explained in Fig.2 in a sequential manner.As machine learning models have two sub-parts i.e. training and testing, this project is also based on the same pattern. Training part of this model includes collection of data and its processing,feature extraction, feature selection and machine learning training.The testing part includes sound recordings collection and predicting the results.If any type of leakage is detected,the message will be sent to telegram.The entire description of the process is briefly described below.

First,the model has to be trained by necessary sound recordings.For this purpose,200 samples have been collected which are then categorised into normal and abnormal signals.Features like Mel Frequency Cepstral Coefficients (MFCC) are extracted from the sound signals.Overall,13 features have been extracted using MFCC. PCA helps in the dimensionality reduction to retain some meaningful properties of the original data.After training with several models, the best model has been selected which gives the maximum training accuracy.The model is then exported for testing with the new

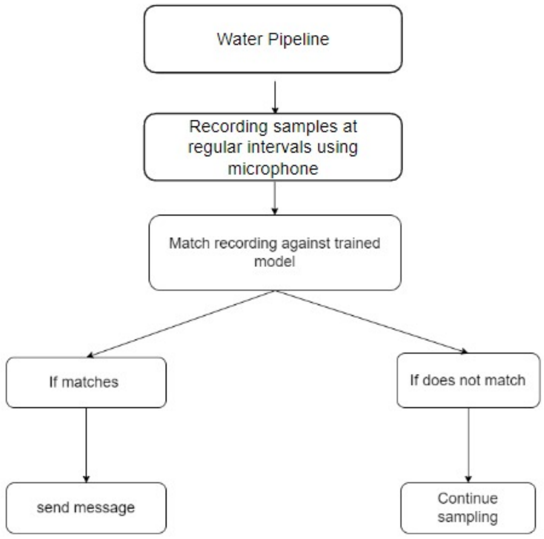

Fig. 2 Basic flow chart for the proposed pipeline monitoring system

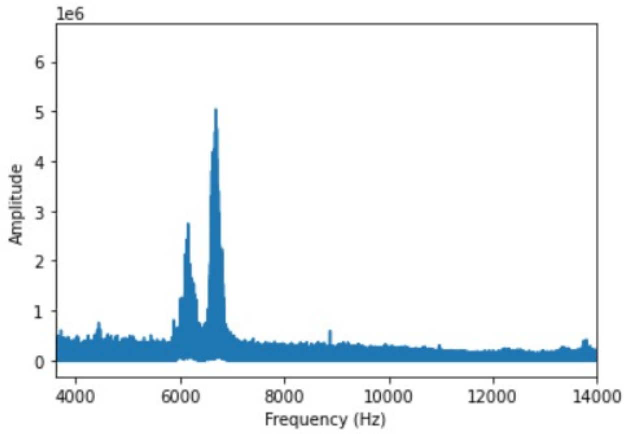

Fig. 3 FFT of the sound signal

data.The signals which will now be coming can be tested with the developed model.This will help to take a continuous look at the pipelines.In case of any abnormality in the signals,a message will be sent in Telegram so that the necessary steps can be taken at the earliest.

3.1 Signal Processing

The first stage of machine learning is collecting and processing the data.At first,a total of 200 samples have been collected , with a duration of 1 minute each.To extract the useful information from this signal,MFCC has been used for sound recognition and PCA for dimensionality reduction.An extra column has been added along with th 13 features extracted using MFCC. If there is leak in that particular recording, then it is assigned "0" otherwise "1" is assigned.This creates our whole data-set for further usage.

3.2 Feature Extraction

The next step is to extract necessary and useful features from the data-set for proper training of the model.The machine learning models are unable to understand the time domain data.Feature extraction converts them into a format that models can understand.

3.2.1 Mel-Frequency Cepstral Coefficients(MFCC)

Cepstrum indicates the rate of change of information in spectral bands. The Fourier transform is used to first transform the signal, and then then the logarithm of the magnitude of the Fourier transform is used to first transform the signal, and then then the logarithm of the magnitude of The periodic components of the initial time domain signal can then be observed as peaks. Cepstrum is what is in this. Mel scale provides a relationship between the perceived frequency of

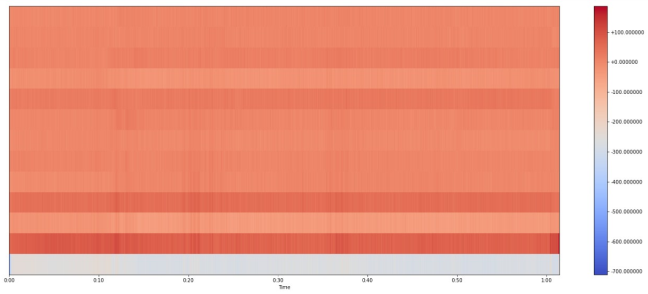

Fig. 4 Diagram showing Heatmap

a tone.As a result, the frequency is scaled to correspond to the hearing frequency of the human ear.(1).

$$f_{mel} = 2595 \log_{10}(1 + \frac{f}{700}), \tag{1}$$

where f_{mel} is the subjective pitch in Mels for a particular frequency. Thus MFCC can be defined as a feature set for audio signal which is implemented in speech and speaker identification methodologies.

MFCC coefficients include a set of Discrete Cosine Transform (DCT) decorrelated frameworks, which are calculated through a conversion of the logarithmically compressed flter-output energies, resultant of a conceptually spaced triangular flter bank that compiles the Discrete Fourier Transformed (DFT) speech signal. An M-point DFT of the discrete input signal y(n) is shown in Eq.

$$Y(k) = \sum_{n=1}^{M} y(n) e^{\frac{-j2\pi nk}{M}} = 1 \tag{2}$$

where, $1 \leq$ k $\leq$ M. Next, the filter bank which has linearly distributed filters in the Mel scale, are extracted on the spectrum. The filter output Ψ_i(k)) of the ith filter in the bank is defined by equation (3).

$$\begin{aligned} \Psi_i(k) &= 0, for k \leq b_{i-1} \\ &= \frac{k - k_{b_{i-1}}}{k_{b_i} - k_{b_{i-1}}}, for k_{b_{i-1}} \leq_{b_b} \\ &= \frac{k_{b_{i+1}} - k}{k_{b_{i+1}} - k_{b_i}}, for k_{b_{i-1}} \leq \leq b_{i+1} \\ &= 0, for k \leq b_{i+1} \end{aligned} \tag{3}$$

If Q denotes the number of filters in the filter bank, then $\{K_{b_i}\}_{i=0}^{Q+1}$ are the borderline points of the filters and k shows us the coefficient parameters in the M-point DFT. The borderline items for each filter i (i=1,2,...,Q) are computed as uniformly distributed points in the Mel scale utilizing equation (4).

$$K_{b_i} = (\frac{M}{f_s}).f_{mel}{}^{-1}[f_{mel}(f_{low}) + \frac{i\{f_{mel}(f_{high}) - f_{mel}(f_{low})\}}{Q+1}] \tag{4}$$

where, f_s is the sampling frequency in Hz and f_{low} and f_{high} are the low and high frequency border of the filter bank, respectively. f_{mel}^{-1} is the inverse of the conversion displayed in equation(1) and is shown in equation (5).

$$f_{mel}^{-1}(f_{mel}) = 700 \cdot [10^{\frac{f_{mel}}{2595}} - 1] \tag{5}$$

In the next step, the output energies e(i)(i=1,2,.....,Q) of the Mel-scaled band-pass filters are computed as an aggregate of the signal energies $|y(k)|^2$ falling into a given Mel frequency band subjected by the particular frequency output of $\Psi_i(k)$ as shown in equation(6).

$$e(i) = \sum_{K}^{M} |y(k)|^2 \Psi_i(k) \tag{6}$$

Finally, an implementation of DCT is computed over the log filter bank energies $\{\log[e(i)]\}_{i=1}^{Q}$ to de-correlate the energies and the final MFCC coefficients C_m are shown in equation (7).

$$C_m = \sqrt{\frac{2}{N}} \sum_{l=0}^{M-1} \log[e(l+1)].\cos[m(\frac{2l+1}{2}).\frac{\Pi}{Q}] \tag{7}$$

where, m=0,1,2,...., R-1 and R is the required number of MFCCs. The importance of MFCC can be understood using [**?**]. Apart from MFCC, pitch of the audio signals is also used for successful feature extraction.

3.2.2 Principle Component Analysis(PCA)

PCA is an unsupervised statistical technique which helps in the reduction of the dimension of the data-set.When used with a larger input data-set, ML models with a high number of input variables or higher dimensions frequently fail. PCA aids in discovering connections between several variables and then coupling them.

3.3 Training Classifier

We trained and compared the classification results of various machine learning models: Logistic regression, SVM, KNN, Decision Tree, Random Forest and Naive Bayes.Random Forest gives the best result in terms of accuracy and precision.

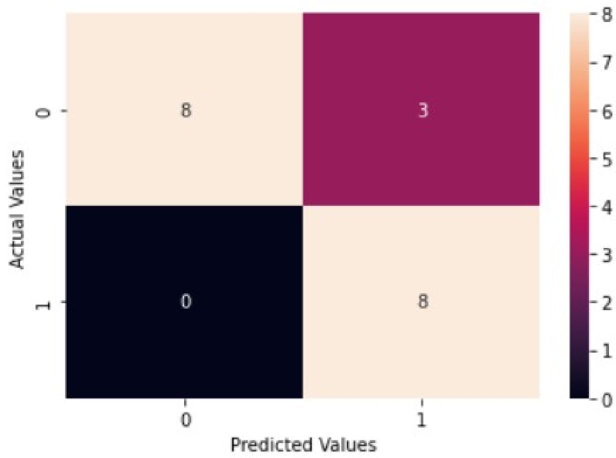

Fig. 5 Diagram showing Confusion Matrix

3.4 Real Time Data Collection

Data collection is performed using sensor nodes which consist of microphone sensors and Raspberry Pi. These are placed along the length of the pipe. The sensor nodes gather the data where The Raspberry Pi processes it and analyses the data accordingly.

3.5 Instantaneous Data Classification

In the Raspberry Pi, the data is fed into the python script and our trained model processes in the real time whether there is any type of leakage in the pipelines.The below written algorithm gives a synopsis of the leakage detection technique used here.

4 IMPLEMENTATION OF REAL-TIME LEAKAGE DETECTION

4.1 Creating Dataset

We have gathered a variety of used and old pipes, many of which already had holes on them and some of which we made ourselves. The sound of water leaking from these holes was recorded in order to get the data for training the model.We have trained and tested the dataset to find the accuracy whether our trained model can understand the outside data or not.If it can, then it will warn us about the leak in real time.

4.2 Features

The recorded sound signals is now converted into 13 features for further analysis of the sound signals.Features like MFCC have been used as shown in the Fig. 4.

4.3 Machine Learning Training

Training is done based on the different sounds produced by the pipe leak.Table III shows the comparison between different machine learning models and we came to the conclusion that random forest is the best classification algorithm with 89 percent accuracy.That is why random forest is used for the identification between a leak sound and a non-leak sound.

4.4 Real Time Testing

For the testing purpose, sensor nodes have been placed along the length of the pipe.It collects the data and after which the Raspberry Pi comes into the picture.The python script along with the RF Model processes the sent data and analysis of the data takes place accordingly.The real time testing means the new sound recordings are fed into the python script.Feature extraction takes place of the new data after that it is analysed by the already trained Random Forest model.It helps in the real time testing as whenever the new data has been sent it checks for the leakage or a non-leakage signal.During the testing interval, we have gathered some pipes and made leaks at two random points in the pipe.The testing was carried out till 10 minutes which generated sufficient number of samples for further analysis and prediction of the leakages in the pipelines.

5 RESULTS AND DISCUSSIONS

To test the working, accuracy and predictions of the proposed methods, an environment has been created that is quite similar to what we want to achieve with this process.Sensor nodes have been located near the pipelines. Python script has already been fed into the Raspberry Pi.A continuous notification alert is set and whenever any leak is detected in the pipelines, a message is sent to the telegram user interface. The output results are really pleasing, and we were able to successfully track the status of the leakage in the pipelines.

6 CONCLUSION AND FUTURE WORKS

We have demonstrated dPipe, an Internet of Things (IoT) based system that can monitor and detect leaks in water distribution systems by gathering data in real time using microphone sensors.To find the optimum algorithm, a number of machine learning models were also tried. We need to increase the accuracy of the model in case of minor leaks.A large number of tests can be done for a quick result and preventing false alarms.We can also attach a system that tells us the exact location of the leak so that the maintenance of the pipelines can be done at the earliest.

References

X.-J. Wang, A. Simpson, M. Lambert, and J. V´ıtkovsky,´ “Leak detection in pipeline systems using hydraulic methods: A review,” Jan. 2001

[2] T. Bui Quy, M. Sohaib, and J. Kim, “A reliable acoustic emission based technique for the detection of a small leak in a pipeline system,” Energies, vol. 12, p. 1472, Apr. 2019.

[3] F. Adnan, M. F. Ghazali, M. Amin, and A. Hamat, “Leak detection in gas pipeline by acoustic and signal processing - a review,” IOP Conference Series: Materials Science and Engineering, vol. 100, p. 012 013, Dec. 2015.